ISSN 2644-3236 (online)

Artificial Intelligence in *Gastroenterology*

Quarterly Volume 5 Number 1 April 30, 2024

Published by Baishideng Publishing Group Inc

A I G

Artificial Intelligence in Gastroenterology

Contents

Quarterly Volume 5 Number 1 April 30, 2024

EDITORIAL

ORIGINAL ARTICLE

Observational Study

Artificial Intelligence in Gastroenterology

Quarterly Volume 5 Number 1 April 30, 2024

Contents

ABOUT COVER

Editorial Board Member of *Artificial Intelligence in Gastroenterology*, Rogerio Serafim Parra, MD, PhD, Assistant Professor, Department of Surgery and Anatomy, Ribeirao Preto Medical School, University of Sao Paulo, Ribeirão Preto 14048-900, São Paulo, Brazil. rsparra@hcrp.usp.br

AIMS AND SCOPE

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

AIG mainly publishes articles reporting research results obtained in the field of artificial intelligence in gastroenterology and covering a wide range of topics, including artificial intelligence in gastrointestinal cancer, liver cancer, pancreatic cancer, hepatitis B, hepatitis C, nonalcoholic fatty liver disease, inflammatory bowel disease, irritable bowel syndrome, and *Helicobacter pylori* infection.

INDEXING/ABSTRACTING

The *AIG* 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 Gastroenterology

ISSN
ISSN 2644-3236 (online)

LAUNCH DATE
July 28, 2020

FREQUENCY
Quarterly

EDITORS-IN-CHIEF
Rajvinder Singh, Ferruccio Bonino

EDITORIAL BOARD MEMBERS
https://www.wjgnet.com/2644-3236/editorialboard.htm

PUBLICATION DATE
April 30, 2024

COPYRIGHT
© 2024 Baishideng Publishing Group Inc

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 Gastroenterology

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

Artif Intell Gastroenterol 2024 April 30; 5(1): 90096

DOI: 10.35712/aig.v5.i1.90096

ISSN 2644-3236 (online)

EDITORIAL

Application of artificial intelligence in the prediction of immunotherapy efficacy in hepatocellular carcinoma: Current status and prospects

Wei Zhang, Li-Ning Song, Yun-Fei You, Feng-Nan Qi, Xiao-Hong Cui, Ming-Xun Yi, Guang Zhu, Ren-An Chang, Hai-Jian Zhang

Specialty type: Computer science, artificial intelligence

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): 0
Grade E (Poor): 0

P-Reviewer: Liakina V, Lithuania

Received: November 23, 2023
Peer-review started: November 23, 2023
First decision: January 12, 2024
Revised: January 28, 2024
Accepted: March 12, 2024
Article in press: March 12, 2024
Published online: April 30, 2024

Wei Zhang, Li-Ning Song, Yun-Fei You, Feng-Nan Qi, Ming-Xun Yi, Ren-An Chang, Research Center of Clinical Medicine and Department of General Surgery, The Affiliated Hospital of Nantong University, Nantong 226001, Jiangsu Province, China

Xiao-Hong Cui, Department of General Surgery, Shanghai Electric Power Hospital, Shanghai 200050, China

Guang Zhu, Hai-Jian Zhang, Division of Life Science, The Hong Kong University of Science and Technology, Hong Kong, China

Hai-Jian Zhang, Research Center of Clinical Medicine, The Affiliated Hospital of Nantong University, Nantong 226001, Jiangsu Province, China

Corresponding author: Hai-Jian Zhang, MD, PhD, Professor, Research Scientist, Research Center of Clinical Medicine, The Affiliated Hospital of Nantong University, No. 20 West Temple Road, Nantong 226001, Jiangsu Province, China. hjzhang@ntu.edu.cn

Abstract

Artificial Intelligence (AI) has increased as a potent tool in medicine, with promising oncology applications. The emergence of immunotherapy has transformed the treatment terrain for hepatocellular carcinoma (HCC), offering new hope to patients with this challenging malignancy. This article examines the role and future of AI in forecasting the effectiveness of immunotherapy in HCC. We highlight the potential of AI to revolutionize the prediction of therapy response, thus improving patient selection and clinical outcomes. The article further outlines the challenges and future research directions in this emerging field.

Key Words: Artificial intelligence; Hepatocellular carcinoma; Immunotherapy; Predictive modeling

Core Tip: Recently, there has been a lot of progress in predicting the effect of immunotherapy for hepatocellular carcinoma using artificial intelligence, but it also faces serious challenges. Therefore, in this article we summarize and discuss these issues.

Citation: Zhang W, Song LN, You YF, Qi FN, Cui XH, Yi MX, Zhu G, Chang RA, Zhang HJ. Application of artificial intelligence in the prediction of immunotherapy efficacy in hepatocellular carcinoma: Current status and prospects. *Artif Intell Gastroenterol* 2024; 5(1): 90096
URL: https://www.wjgnet.com/2644-3236/full/v5/i1/90096.htm
DOI: https://dx.doi.org/10.35712/aig.v5.i1.90096

INTRODUCTION

Hepatocellular carcinoma (HCC), the most prevalent primary liver cancer and a top contributor to global cancer mortality, is frequently detected at an advanced stage, offering few treatment choices and a bleak prognosis[1,2]. Traditional treatment modalities, such as surgery, radiotherapy, and chemotherapy, have been the mainstay of HCC management, but their efficacy is frequently limited by tumor recurrence and adverse side effects[3-5]. With the advent of immunotherapy, the treatment landscape for HCC has been changed, offering new hope to patients with this aggressive form of liver cancer[6,7]. Immunotherapy has shown promising results in improving survival rates. It is being evaluated in various stages of HCC treatment, from its role in the adjuvant setting to its use in advanced stages of the disease[4,6,8].

Despite progress, the heterogeneity of HCC and the complexity of the tumor microenvironment (TME) hinder consistent, durable responses across all patient groups[9]. Predicting which patients will benefit from these treatments remains a significant challenge. Currently, ongoing clinical trials and research efforts are focused on understanding the mechanisms of resistance and identifying biomarkers to predict immunotherapy response[10,11]. This is where artificial intelligence (AI) comes into play, offering a new dimension to these efforts. AI has the potential to analyze complex biomedical data, identifying patterns that could predict treatment outcomes. Integrating AI into clinical practice may lead to more personalized, effective treatment strategies, optimizing patient care and resource use[12].

AI models have achieved remarkable success in various medical applications, such as diagnostic imaging, genomics, and drug discovery. In the context of HCC, AI applications extend to predicting patient prognosis, and treatment response, and even suggesting potential therapeutic targets[13,14]. However, the application of AI in predicting the efficacy of immunotherapy for HCC remains nascent, with numerous challenges yet to be overcome.

The evolution of immunotherapy for HCC

Historically, the initial attempts to harness the immune system to combat HCC centered around cytokine-based therapies, such as interferon-alpha and interleukin-2. Later, it was realized that immune checkpoints, such as CTLA-4 and PD-(L)1, could be manipulated to improve anti-tumor immunity[15].

The first breakthrough in the immunotherapy of HCC came with the approval of nivolumab, a PD-1 inhibitor, for use in patients with advanced HCC who had previously received sorafenib[7]. More recently, researchers have explored combination therapies, such as PD-L1 inhibitors with CTLA-4 inhibitors or with other therapeutic modalities like targeted therapies and locoregional treatments, which are predicated on the potential to synergize different mechanisms of action to enhance anti-tumor responses[16]. For example, the combination of atezolizumab and bevacizumab demonstrated improved survival outcomes relative to sorafenib in the IMbrave150 trial[5].

Despite these advances, the response to immunotherapy in HCC remains variable, with a significant proportion of patients not experiencing benefit. Consequently, this variability has spurred ongoing research into biomarkers that can predict response to immunotherapy.

The potential of AI in oncology

AI, encompassing a wide area of computer science, works toward building systems able to accomplish functions commonly needing human cognition. Among these subfields are machine learning (ML), deep learning (DL), natural language processing, and robotics[17,18]. In the field of oncology, the potential role of AI is to enhance diagnosis, and treatment, and predict treatment outcomes or disease progression.

For diagnosis, DL algorithms that analyze low-dose computed tomography (CT) scans can detect early-stage lung cancer with precision comparable to expert radiologists, potentially leading to earlier and more effective interventions [19]. Similarly, applying DL algorithms to whole-slide pathology images can aid pathologists in identifying cancerous tissues, thus significantly expediting the diagnostic process and enhancing diagnostic accuracy and efficiency[20]. Using AI integrated with machines and DL in radiomics can help to more accurately define tissue characteristics[21].

Regarding treatment, AI's integration and analysis of genomic data alongside clinical histories enable the creation of personalized treatment plans that predict patient benefits from specific therapies. Zhang *et al*[22] utilized Garson's algorithm, Lek's profile, local interpretable model-agnostic explanations, and partial dependence plots to aid clinicians and medical policymakers in understanding artificial neural networks, powerful tools for effectively predicting outcome variable relationships.

Additionally, AI also aids in anticipating treatment outcomes, helping to optimize treatment regimens by forecasting patient responses and potential side effects based on historical data. For instance, DL algorithms, in particular convolutional networks, utilize imaging data to track tumor advancement or decline, offering impartial markers for modifying the treatment process[23]. AI models integrate various patient data, such as genetic information, clinical histories, and imaging findings, to predict disease progression. This comprehensive approach leads to more accurate prognostic assessments than traditional methods, which often consider fewer data points[24]. Survival convolutional neural networks integrate histology images and genomic biomarkers using DL to predict patient outcomes, surpassing current clinical methods for predicting overall survival (OS) in glioma patients[25]. Huang *et al*[26] identified 10 signature genes from a pool of 166 stem cell-related genes using the least absolute shrinkage operator (LASSO) and multivariate Cox regression analysis. They found that this signature effectively predicted the response to chemotherapy in lung adenocarcinoma patients. Furthermore, in Ding *et al*'s experiment, the mirlncRNA signature, comprising five notable lncRNAs, not only differentiates molecular typing and identifies the related tumor immune subtypes and their chromatin accessibility, but also underscores the immune efficacy and drug sensitivity of tumor immune subtypes[27].

The role of AI in predicting immunotherapy efficacy in HCC

ML, a subset of AI, may become a powerful tool for predicting the response of HCC patients to immunotherapy. ML models, like support vector machines (SVMs), have been utilized for predicting HCC recurrence, screening drugs, identifying potential targets, and determining which patients are more likely to experience recurrence with specific treatments[24,28-30]. Shi *et al*[31] examined peripheral blood mononuclear cells from various cohorts, creating an AdaBoost-SVM logistic model that can identify early-stage HCC *via* immune markers, surpassing alpha-fetoprotein in accuracy.

DL, another subset of AI, uses algorithms to model and understand complex patterns in data. Zeng *et al*[32] explored three DL approaches (patch-based, classic MIL, and CLAM) to create and verify AI-based pathology models for predicting immune and inflammatory gene signatures. Their findings suggest that these signatures could be associated with heightened sensitivity to immunotherapy in patients with advanced HCC.

Additionally, genomic data also play a crucial role in predicting treatment responses. AI models have been trained to identify genetic mutations and expression profiles that correlate with better immunotherapy outcomes. Gong *et al*[33] utilized ML to develop a risk scoring system known as 'neutrophil-derived signature' (NDS), comprising 10 crucial genes. The RiskScore of NDS showed higher accuracy compared to clinical variables and was associated with increased malignancy levels. Consequently, the predictive prowess of DL models can guide clinicians in identifying patients who stand to gain the most from immunotherapy, fostering personalized and efficacious treatment approaches. Xie *et al*[34] found that the m6A- and ferroptosis-lncRNA signature, which has significant prognostic value, provides new perspectives in distinguishing 'cold' and 'hot' tumors and could have important implications for personalized therapy to improve the survival rate of HCC patients. Feng *et al*[35] utilized the LASSO and CoxBoost algorithms to combine and create a signature from 11 natural killer cell-related genes. This provided a new method for evaluating the prognosis and immunotherapeutic response of HCC patients. Dai *et al*[36] employed the LASSO regression model to create an immune-related gene-based prognostic index. This index can predict immune cell infiltration in the HCC TME, as well as the response to immunotherapy. Shen *et al*[37] used genes related to aging to create a predictive model. Through Spearman correlation analysis, they found that the model's risk score was closely related to Mismatch Repair and expression of immune checkpoints.

AI models, particularly those based on ML and DL, have several advantages over traditional statistical methods in predicting outcomes and treatment responses in HCC. Traditional methods often rely on predefined clinical and pathological criteria, which may not capture the complex biological interactions underlying HCC progression and response to treatment. In contrast, AI models can integrate a wide range of data types and identify non-linear relationships within the data[30,38]. Comparative studies have shown that AI models can outperform traditional scoring systems and clinical judgment in prognostication and treatment prediction. For instance, the random survival forests model showed greater accuracy in predicting early recurrence of HCC after surgery compared to COX proportional hazard models[39]. Moreover, AI-driven tools can continuously learn and improve as they are exposed to new data, a feature that static traditional models lack.

Moreover, AI has been instrumental in discovering novel biomarkers for HCC. Through the analysis of large datasets, DL can uncover subtle correlations between biomarkers and treatment responses that may not be apparent to human investigators. Liang *et al*[40] introduced an interpretable human-centric DL-guided framework, Pathological-biomarker-finder, to aid pathologists in identifying new tissue biomarkers using effective DL models.

Given the promising prospects of AI in predicting the immunotherapy efficacy of HCC, an increasing array of predictive variables is being incorporated into clinical practice. These variables include risk-scoring systems, gene phenotypes, and other types of biomarkers. For instance, Hatanaka *et al*[41] conducted a multicenter retrospective analysis that employed the modified Gustave Roussy Immune (GRIm) score as a new prognostic tool for HCC patients treated with atezolizumab and bevacizumab. Their findings indicated that a high GRIm score is a significant adverse factor for both progression-free survival (PFS) and OS. The retrospective analysis by Sangro *et al*[42] revealed an inflammatory gene signature consisting of four genes: CD274, CD8A, LAG3, and STAT1, which correlated with improved response rates and OS in advanced HCC patients treated with nivolumab. Similarly, the atezolizumab-bevacizumab response signature (ABRS), associated with PFS after starting treatment with atezolizumab-bevacizumab, includes genes like CXCR2P1, ICOS, and TIMD4[43]. Building on this, Zeng *et al*[44] developed a prediction model (ABRS-P) and found that patients with ABRS-P-high tumors had a significantly longer median PFS than those with ABRS-P-low tumors. In addition, Sun *et al*[45] used specific patient cohorts with advanced solid tumors to develop and validate a radiomic signature capable of predicting immunotherapy responses by assessing CD8 T cell infiltration in tumors. In another

clinical trial, researchers identified pre-existing CD8 T cells as a promising biomarker for forecasting responses to combined lenvatinib and PD-1 inhibitors in unresectable HCC[46].

Challenges in AI application to HCC immunotherapy

Despite these advancements, several challenges still impede the broader application of AI in predicting HCC immunotherapy efficacy. A significant challenge is the availability of high-quality, annotated datasets. AI models require large amounts of data to learn effectively; however, the scarcity of such datasets can limit the performance of these models[47]. Furthermore, AI models predicting immunotherapy outcomes need to be updated regularly to incorporate the latest clinical knowledge and patient data. Additionally, the heterogeneity of HCC presents another challenge, as it can vary greatly in its genetic makeup and clinical presentation. This variability can hinder AI models from generalizing their predictions effectively across different patient populations[11]. Furthermore, the interpretability of AI models, particularly DL models, remains a significant concern. The unclear and puzzling nature of these models can make it tough for healthcare professionals to comprehend the logic behind the predictions, which is essential for making well-informed clinical decisions[48].

The use of AI in healthcare also raises important ethical and regulatory considerations. Issues including patient privacy, data security, and informed consent must be addressed to ensure the ethical application of AI in predicting immunotherapy response[49].

Prospects and future directions

The integration of AI into clinical practice for predicting HCC immunotherapy response is an ongoing endeavor. A crucial part of this integration process is the thorough validation of AI models across various patient groups to ensure that the predictions are strong and dependable[50]. Additionally, the development of user-friendly AI platforms that healthcare professionals can easily access and utilize is another important aspect. To facilitate their adoption in routine practice, these platforms must be explicitly designed with a focus on clinical workflow integration[51].

As biomedical data continuously accumulates and AI technology advances, we can expect the predictive capabilities of AI models to improve correspondingly. One promising area of future development involves integrating AI with other emerging technologies, such as liquid biopsy and single-cell sequencing. Analyses of circulating nucleic acids, often called 'liquid biopsies', can monitor treatment response, evaluate drug resistance emergence, and measure minimal residual disease[52]. Compared to traditional bulk sequencing, single-cell sequencing can analyze HCC at single-cell resolution, accurately identify different cell types, and uncover the heterogeneity of HCC cells[53]. Lu *et al*[54] previously developed a new diagnostic model for HCC using single-cell RNA sequencing data and discovered that patients with high-risk scores were less likely to benefit from immunotherapy. Another area of exploration is AI's role in dynamic prediction models, which can monitor patient responses in real time and adjust predictions accordingly. This approach has the potential to result in more predictable outcomes and treatment strategies that are personalized and adaptable[55].

In the future, AI applications are expected to benefit from the development of more advanced algorithms capable of processing complex biological data, which includes genomics, proteomics, and metabolomics. The goal of these algorithms will be to identify novel biomarkers and molecular signatures that can predict immunotherapy response in HCC patients[56]. The effectiveness of immunotherapy often relies on the interaction of immunomodulation in the TME[57]. By integrating multi-omics data, we will gain a more comprehensive understanding of the TME and its interactions with the immune system[58,59]. In addition, AI is poised to play a critical role in the design and implementation of clinical trials for HCC immunotherapy by identifying patient subgroups that are more likely to benefit from specific treatments. This will help categorize participants and improve the results of trials (Figure 1).

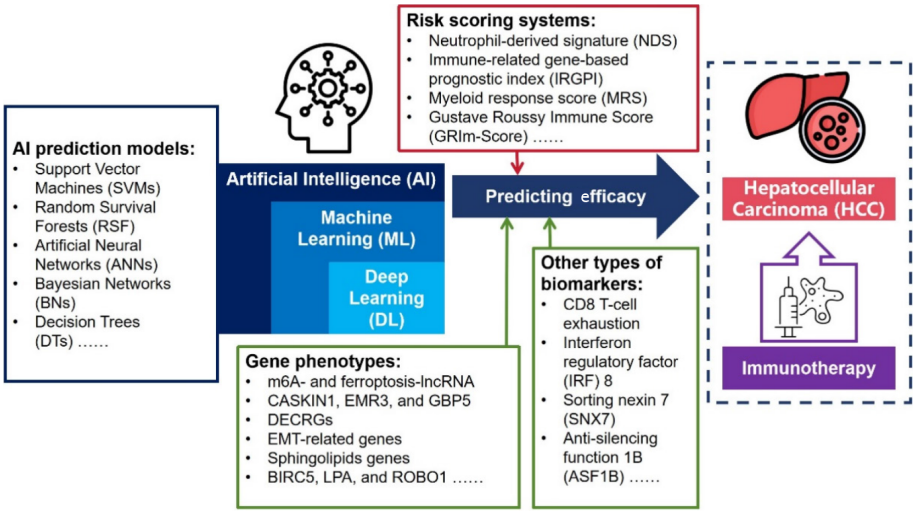

Figure 1 Schematic diagram of artificial intelligence in the prediction of immunotherapy efficacy in hepatocellular carcinoma. This

schematic diagram displays the application of artificial intelligence in the prediction of immunotherapy efficacy in hepatocellular carcinoma. It mentions several artificial intelligence prediction models closely connected to this article, along with risk scoring systems, gene phenotypes, and other types of biomarkers that have surfaced in recent research. AI: Artificial intelligence.

CONCLUSION

The current status of AI in predicting the efficacy of immunotherapy for HCC is marked by significant advancements and potential, tempered by challenges and considerations that must be addressed. The schematic diagram in Figure 1 displays the application of AI in predicting the effectiveness of immunotherapy for HCC. As AI technology evolves and becomes more integrated into healthcare, it possesses the potential to transform HCC prognosis and treatment through personalized and precise predictions for immunotherapy.

FOOTNOTES

Author contributions: Zhang W and Zhang HJ have access to all the data in this study and take responsibility for the integrity and accuracy of the data analyses. Zhang HJ, Zhu G, and Chang RA study concept and design; Zhang W drafted the manuscript; Song LN revised the manuscript; You YF, Qi FN, Cui XH, and Yi MX supervised the study and provided modification suggestions; All authors have read and approved the final manuscript; Zhang W and Song LN contributed equally to this work.

Supported by the National Natural Science Foundation of China, No. 81401988; China Postdoctoral Science Foundation, No. 2019M661907; Jiangsu Postdoctoral Science Foundation, No. 2019K159, and No. 2019Z153; General Project of Jiangsu Provincial Health Committee, No. H2023136; General Project of Nantong Municipal Health Committee, No. MS2023013; Jiangsu Provincial Research Hospital, No. YJXYY202204-YSB28; and College Student Innovation Program, No. 202210304128Y, and No. 2023103041055.

Conflict-of-interest statement: All authors certify that they have no conflicts of interest related to this work.

Country/Territory of origin: China

ORCID number: Xiao-Hong Cui 0000-0002-7435-4317; Hai-Jian Zhang 0000-0003-2516-4215.

S-Editor: Liu JH
L-Editor: Webster JR
P-Editor: Zhao YQ

REFERENCES

1 **Global Burden of Disease 2019 Cancer Collaboration**, Kocarnik JM, Compton K, Dean FE, Fu W, Gaw BL, Harvey JD, Henrikson HJ, Lu D, Pennini A, Xu R, Ababneh E, Abbasi-Kangevari M, Abbastabar H, Abd-Elsalam SM, Abdoli A, Abedi A, Abidi H, Abolhassani H, Adedeji IA, Adnani QES, Advani SM, Afzal MS, Aghaali M, Ahinkorah BO, Ahmad S, Ahmad T, Ahmadi A, Ahmadi S, Ahmed Rashid T, Ahmed Salih Y, Akalu GT, Aklilu A, Akram T, Akunna CJ, Al Hamad H, Alahdab F, Al-Aly Z, Ali S, Alimohamadi Y, Alipour V, Aljunid SM, Alkhayyat M, Almasi-Hashiani A, Almasri NA, Al-Maweri SAA, Almustanyir S, Alonso N, Alvis-Guzman N, Amu H, Anbesu EW, Ancuceanu R, Ansari F, Ansari-Moghaddam A, Antwi MH, Anvari D, Anyasodor AE, Aqeel M, Arabloo J, Arab-Zozani M, Aremu O, Ariffin H, Aripov T, Arshad M, Artaman A, Arulappan J, Asemi Z, Asghari Jafarabadi M, Ashraf T, Atorkey P, Aujayeb A, Ausloos M, Awedew AF, Ayala Quintanilla BP, Ayenew T, Azab MA, Azadnajafabad S, Azari Jafari A, Azarian G, Azzam AY, Badiye AD, Bahadory S, Baig AA, Baker JL, Balakrishnan S, Banach M, Bärnighausen TW, Barone-Adesi F, Barra F, Barrow A, Behzadifar M, Belgaumi UI, Bezabhe WMM, Bezabih YM, Bhagat DS, Bhagavathula AS, Bhardwaj N, Bhardwaj P, Bhaskar S, Bhattacharyya K, Bhojaraja VS, Bibi S, Bijani A, Biondi A, Bisignano C, Bjørge T, Bleyer A, Blyuss O, Bolarinwa OA, Bolla SR, Braithwaite D, Brar A, Brenner H, Bustamante-Teixeira MT, Butt NS, Butt ZA, Caetano Dos Santos FL, Cao Y, Carreras G, Catalá-López F, Cembranel F, Cerin E, Cernigliaro A, Chakinala RC, Chattu SK, Chattu VK, Chaturvedi P, Chimed-Ochir O, Cho DY, Christopher DJ, Chu DT, Chung MT, Conde J, Cortés S, Cortesi PA, Costa VM, Cunha AR, Dadras O, Dagnew AB, Dahlawi SMA, Dai X, Dandona L, Dandona R, Darwesh AM, das Neves J, De la Hoz FP, Demis AB, Denova-Gutiérrez E, Dhamnetiya D, Dhimal ML, Dhimal M, Dianatinasab M, Diaz D, Djalalinia S, Do HP, Doaei S, Dorostkar F, Dos Santos Figueiredo FW, Driscoll TR, Ebrahimi H, Eftekharzadeh S, El Tantawi M, El-Abid H, Elbarazi I, Elhabashy HR, Elhadi M, El-Jaafary SI, Eshrati B, Eskandarieh S, Esmaeilzadeh F, Etemadi A, Ezzikouri S, Faisaluddin M, Faraon EJA, Fares J, Farzadfar F, Feroze AH, Ferrero S, Ferro Desideri L, Filip I, Fischer F, Fisher JL, Foroutan M, Fukumoto T, Gaal PA, Gad MM, Gadanya MA, Gallus S, Gaspar Fonseca M, Getachew Obsa A, Ghafourifard M, Ghashghaee A, Ghith N, Gholamalizadeh M, Gilani SA, Ginindza TG, Gizaw ATT, Glasbey JC, Golechha M, Goleij P, Gomez RS, Gopalani SV, Gorini G, Goudarzi H, Grosso G, Gubari MIM, Guerra MR, Guha A, Gunasekera DS, Gupta B, Gupta VB, Gupta VK, Gutiérrez RA, Hafezi-Nejad N, Haider MR, Haj-Mirzaian A, Halwani R, Hamadeh RR, Hameed S, Hamidi S, Hanif A, Haque

S, Harlianto NI, Haro JM, Hasaballah AI, Hassanipour S, Hay RJ, Hay SI, Hayat K, Heidari G, Heidari M, Herrera-Serna BY, Herteliu C, Hezam K, Holla R, Hossain MM, Hossain MBH, Hosseini MS, Hosseini M, Hosseinzadeh M, Hostiuc M, Hostiuc S, Househ M, Hsairi M, Huang J, Hugo FN, Hussain R, Hussein NR, Hwang BF, Iavicoli I, Ibitoye SE, Ida F, Ikuta KS, Ilesanmi OS, Ilic IM, Ilic MD, Irham LM, Islam JY, Islam RM, Islam SMS, Ismail NE, Isola G, Iwagami M, Jacob L, Jain V, Jakovljevic MB, Javaheri T, Jayaram S, Jazayeri SB, Jha RP, Jonas JB, Joo T, Joseph N, Joukar F, Jürisson M, Kabir A, Kahrizi D, Kalankesh LR, Kalhor R, Kaliyadan F, Kalkonde Y, Kamath A, Kameran Al-Salihi N, Kandel H, Kapoor N, Karch A, Kasa AS, Katikireddi SV, Kauppila JH, Kavetskyy T, Kebede SA, Keshavarz P, Keykhaei M, Khader YS, Khalilov R, Khan G, Khan M, Khan MN, Khan MAB, Khang YH, Khater AM, Khayamzadeh M, Kim GR, Kim YJ, Kisa A, Kisa S, Kissimova-Skarbek K, Kopec JA, Koteeswaran R, Koul PA, Koulmane Laxminarayana SL, Koyanagi A, Kucuk Bicer B, Kugbey N, Kumar GA, Kumar N, Kurmi OP, Kutluk T, La Vecchia C, Lami FH, Landires I, Lauriola P, Lee SW, Lee SWH, Lee WC, Lee YH, Leigh J, Leong E, Li J, Li MC, Liu X, Loureiro JA, Lunevicius R, Magdy Abd El Razek M, Majeed A, Makki A, Male S, Malik AA, Mansournia MA, Martini S, Masoumi SZ, Mathur P, McKee M, Mehrotra R, Mendoza W, Menezes RG, Mengesha EW, Mesregah MK, Mestrovic T, Miao Jonasson J, Miazgowski B, Miazgowski T, Michalek IM, Miller TR, Mirzaei H, Mirzaei HR, Misra S, Mithra P, Moghadaszadeh M, Mohammad KA, Mohammad Y, Mohammadi M, Mohammadi SM, Mohammadian-Hafshejani A, Mohammed S, Moka N, Mokdad AH, Molokhia M, Monasta L, Moni MA, Moosavi MA, Moradi Y, Moraga P, Morgado-da-Costa J, Morrison SD, Mosapour A, Mubarik S, Mwanri L, Nagarajan AJ, Nagaraju SP, Nagata C, Naimzada MD, Nangia V, Naqvi AA, Narasimha Swamy S, Ndejjo R, Nduaguba SO, Negoi I, Negru SM, Neupane Kandel S, Nguyen CT, Nguyen HLT, Niazi RK, Nnaji CA, Noor NM, Nuñez-Samudio V, Nzoputam CI, Oancea B, Ochir C, Odukoya OO, Ogbo FA, Olagunju AT, Olakunde BO, Omar E, Omar Bali A, Omonisi AEE, Ong S, Onwujekwe OE, Orru H, Ortega-Altamirano DV, Otstavnov N, Otstavnov SS, Owolabi MO, P A M, Padubidri JR, Pakshir K, Pana A, Panagiotakos D, Panda-Jonas S, Pardhan S, Park EC, Park EK, Pashazadeh Kan F, Patel HK, Patel JR, Pati S, Pattanshetty SM, Paudel U, Pereira DM, Pereira RB, Perianayagam A, Pillay JD, Pirouzpanah S, Pishgar F, Podder I, Postma MJ, Pourjafar H, Prashant A, Preotescu L, Rabiee M, Rabiee N, Radfar A, Radhakrishnan RA, Radhakrishnan V, Rafiee A, Rahim F, Rahimzadeh S, Rahman M, Rahman MA, Rahmani AM, Rajai N, Rajesh A, Rakovac I, Ram P, Ramezanzadeh K, Ranabhat K, Ranasinghe P, Rao CR, Rao SJ, Rawassizadeh R, Razeghinia MS, Renzaho AMN, Rezaei N, Rezapour A, Roberts TJ, Rodriguez JAB, Rohloff P, Romoli M, Ronfani L, Roshandel G, Rwegerera GM, S M, Sabour S, Saddik B, Saeed U, Sahebkar A, Sahoo H, Salehi S, Salem MR, Salimzadeh H, Samaei M, Samy AM, Sanabria J, Sankararaman S, Santric-Milicevic MM, Sardiwalla Y, Sarveazad A, Sathian B, Sawhney M, Saylan M, Schneider IJC, Sekerija M, Seylani A, Shafaat O, Shaghaghi Z, Shaikh MA, Shamsoddin E, Shannawaz M, Sharma R, Sheikh A, Sheikhbahaei S, Shetty A, Shetty JK, Shetty PH, Shibuya K, Shirkoohi R, Shivakumar KM, Shivarov V, Siabani S, Siddappa Malleshappa SK, Silva DAS, Singh JA, Sintayehu Y, Skryabin VY, Skryabina AA, Soeberg MJ, Sofi-Mahmudi A, Sotoudeh H, Steiropoulos P, Straif K, Subedi R, Sufiyan MB, Sultan I, Sultana S, Sur D, Szerencsés V, Szócska M, Tabarés-Seisdedos R, Tabuchi T, Tadbiri H, Taherkhani A, Takahashi K, Talaat IM, Tan KK, Tat VY, Tedla BAA, Tefera YG, Tehrani-Banihashemi A, Temsah MH, Tesfay FH, Tessema GA, Thapar R, Thavamani A, Thoguluva Chandrasekar V, Thomas N, Tohidinik HR, Touvier M, Tovani-Palone MR, Traini E, Tran BX, Tran KB, Tran MTN, Tripathy JP, Tusa BS, Ullah I, Ullah S, Umapathi KK, Unnikrishnan B, Upadhyay E, Vacante M, Vaezi M, Valadan Tahbaz S, Velazquez DZ, Veroux M, Violante FS, Vlassov V, Vo B, Volovici V, Vu GT, Waheed Y, Wamai RG, Ward P, Wen YF, Westerman R, Winkler AS, Yadav L, Yahyazadeh Jabbari SH, Yang L, Yaya S, Yazie TSY, Yeshaw Y, Yonemoto N, Younis MZ, Yousefi Z, Yu C, Yuce D, Yunusa I, Zadnik V, Zare F, Zastrozhin MS, Zastrozhina A, Zhang J, Zhong C, Zhou L, Zhu C, Ziapour A, Zimmermann IR, Fitzmaurice C, Murray CJL, Force LM. Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life Years for 29 Cancer Groups From 2010 to 2019: A Systematic Analysis for the Global Burden of Disease Study 2019. *JAMA Oncol* 2022; **8**: 420-444 [PMID: 34967848 DOI: 10.1001/jamaoncol.2021.6987]

2 **Forner A**, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet* 2018; **391**: 1301-1314 [PMID: 29307467 DOI: 10.1016/S0140-6736(18)30010-2]

3 **Vogel A**, Meyer T, Sapisochin G, Salem R, Saborowski A. Hepatocellular carcinoma. *Lancet* 2022; **400**: 1345-1362 [PMID: 36084663 DOI: 10.1016/S0140-6736(22)01200-4]

4 **Llovet JM**, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, Lencioni R, Koike K, Zucman-Rossi J, Finn RS. Hepatocellular carcinoma. *Nat Rev Dis Primers* 2021; **7**: 6 [PMID: 33479224 DOI: 10.1038/s41572-020-00240-3]

5 **Cheng AL**, Hsu C, Chan SL, Choo SP, Kudo M. Challenges of combination therapy with immune checkpoint inhibitors for hepatocellular carcinoma. *J Hepatol* 2020; **72**: 307-319 [PMID: 31954494 DOI: 10.1016/j.jhep.2019.09.025]

6 **Ziogas IA**, Evangeliou AP, Giannis D, Hayat MH, Mylonas KS, Tohme S, Geller DA, Elias N, Goyal L, Tsoulfas G. The Role of Immunotherapy in Hepatocellular Carcinoma: A Systematic Review and Pooled Analysis of 2,402 Patients. *Oncologist* 2021; **26**: e1036-e1049 [PMID: 33314549 DOI: 10.1002/onco.13638]

7 **El-Khoueiry AB**, Sangro B, Yau T, Crocenzi TS, Kudo M, Hsu C, Kim TY, Choo SP, Trojan J, Welling TH Rd, Meyer T, Kang YK, Yeo W, Chopra A, Anderson J, Dela Cruz C, Lang L, Neely J, Tang H, Dastani HB, Melero I. Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *Lancet* 2017; **389**: 2492-2502 [PMID: 28434648 DOI: 10.1016/S0140-6736(17)31046-2]

8 **Yau T**, Kang YK, Kim TY, El-Khoueiry AB, Santoro A, Sangro B, Melero I, Kudo M, Hou MM, Matilla A, Tovoli F, Knox JJ, Ruth He A, El-Rayes BF, Acosta-Rivera M, Lim HY, Neely J, Shen Y, Wisniewski T, Anderson J, Hsu C. Efficacy and Safety of Nivolumab Plus Ipilimumab in Patients With Advanced Hepatocellular Carcinoma Previously Treated With Sorafenib: The CheckMate 040 Randomized Clinical Trial. *JAMA Oncol* 2020; **6**: e204564 [PMID: 33001135 DOI: 10.1001/jamaoncol.2020.4564]

9 **Greten TF**, Lai CW, Li G, Staveley-O'Carroll KF. Targeted and Immune-Based Therapies for Hepatocellular Carcinoma. *Gastroenterology* 2019; **156**: 510-524 [PMID: 30287171 DOI: 10.1053/j.gastro.2018.09.051]

10 **Harding JJ**, Nandakumar S, Armenia J, Khalil DN, Albano M, Ly M, Shia J, Hechtman JF, Kundra R, El Dika I, Do RK, Sun Y, Kingham TP, D'Angelica MI, Berger MF, Hyman DM, Jarnagin W, Klimstra DS, Janjigian YY, Solit DB, Schultz N, Abou-Alfa GK. Prospective Genotyping of Hepatocellular Carcinoma: Clinical Implications of Next-Generation Sequencing for Matching Patients to Targeted and Immune Therapies. *Clin Cancer Res* 2019; **25**: 2116-2126 [PMID: 30373752 DOI: 10.1158/1078-0432.CCR-18-2293]

11 **Sia D**, Jiao Y, Martinez-Quetglas I, Kuchuk O, Villacorta-Martin C, Castro de Moura M, Putra J, Camprecios G, Bassaganyas L, Akers N, Losic B, Waxman S, Thung SN, Mazzaferro V, Esteller M, Friedman SL, Schwartz M, Villanueva A, Llovet JM. Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features. *Gastroenterology* 2017; **153**: 812-826 [PMID: 28624577 DOI: 10.1053/j.gastro.2017.06.007]

12 **Jiang P**, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, Liu J, Freeman GJ, Brown MA, Wucherpfennig KW, Liu XS. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* 2018; **24**: 1550-1558 [PMID: 30127393 DOI: 10.1038/s41591-018-0136-1]

13 **Sabottke CF**, Spieler BM, Moawad AW, Elsayes KM. Artificial Intelligence in Imaging of Chronic Liver Diseases: Current Update and Future Perspectives. *Magn Reson Imaging Clin N Am* 2021; **29**: 451-463 [PMID: 34243929 DOI: 10.1016/j.mric.2021.05.011]

14 **Chen D**, Liu J, Zang L, Xiao T, Zhang X, Li Z, Zhu H, Gao W, Yu X. Integrated Machine Learning and Bioinformatic Analyses Constructed a Novel Stemness-Related Classifier to Predict Prognosis and Immunotherapy Responses for Hepatocellular Carcinoma Patients. *Int J Biol Sci* 2022; **18**: 360-373 [PMID: 34975338 DOI: 10.7150/ijbs.66913]

15 **Pardoll DM**. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer* 2012; **12**: 252-264 [PMID: 22437870 DOI: 10.1038/nrc3239]

16 **Finn RS**, Qin S, Ikeda M, Galle PR, Ducreux M, Kim TY, Kudo M, Breder V, Merle P, Kaseb AO, Li D, Verret W, Xu DZ, Hernandez S, Liu J, Huang C, Mulla S, Wang Y, Lim HY, Zhu AX, Cheng AL; IMbrave150 Investigators. Atezolizumab plus Bevacizumab in Unresectable Hepatocellular Carcinoma. *N Engl J Med* 2020; **382**: 1894-1905 [PMID: 32402160 DOI: 10.1056/NEJMoa1915745]

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

18 **Froń A**, Semianiuk A, Lazuk U, Ptaszkowski K, Siennicka A, Lemiński A, Krajewski W, Szydełko T, Małkiewicz B. Artificial Intelligence in Urooncology: What We Have and What We Expect. *Cancers (Basel)* 2023; **15** [PMID: 37686558 DOI: 10.3390/cancers15174282]

19 **Ardila D**, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, Tse D, Etemadi M, Ye W, Corrado G, Naidich DP, Shetty S. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019; **25**: 954-961 [PMID: 31110349 DOI: 10.1038/s41591-019-0447-x]

20 **Ehteshami Bejnordi B**, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, van der Laak JAWM; the CAMELYON16 Consortium, Hermsen M, Manson QF, Balkenhol M, Geessink O, Stathonikos N, van Dijk MC, Bult P, Beca F, Beck AH, Wang D, Khosla A, Gargeya R, Irshad H, Zhong A, Dou Q, Li Q, Chen H, Lin HJ, Heng PA, Haß C, Bruni E, Wong Q, Halici U, Öner MÜ, Cetin-Atalay R, Berseth M, Khvatkov V, Vylegzhanin A, Kraus O, Shaban M, Rajpoot N, Awan R, Sirinukunwattana K, Qaiser T, Tsang YW, Tellez D, Annuscheit J, Hufnagl P, Valkonen M, Kartasalo K, Latonen L, Ruusuvuori P, Liimatainen K, Albarqouni S, Mungal B, George A, Demirci S, Navab N, Watanabe S, Seno S, Takenaka Y, Matsuda H, Ahmady Phoulady H, Kovalev V, Kalinovsky A, Liauchuk V, Bueno G, Fernandez-Carrobles MM, Serrano I, Deniz O, Racoceanu D, Venâncio R. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA* 2017; **318**: 2199-2210 [PMID: 29234806 DOI: 10.1001/jama.2017.14585]

21 **Parekh VS**, Jacobs MA. Deep learning and radiomics in precision medicine. *Expert Rev Precis Med Drug Dev* 2019; **4**: 59-72 [PMID: 31080889 DOI: 10.1080/23808993.2019.1585805]

22 **Zhang Z**, Beck MW, Winkler DA, Huang B, Sibanda W, Goyal H; written on behalf of AME Big-Data Clinical Trial Collaborative Group. Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. *Ann Transl Med* 2018; **6**: 216 [PMID: 30023379 DOI: 10.21037/atm.2018.05.32]

23 **Litjens G**, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. *Med Image Anal* 2017; **42**: 60-88 [PMID: 28778026 DOI: 10.1016/j.media.2017.07.005]

24 **Kourou K**, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J* 2015; **13**: 8-17 [PMID: 25750696 DOI: 10.1016/j.csbj.2014.11.005]

25 **Mobadersany P**, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, Brat DJ, Cooper LAD. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci U S A* 2018; **115**: E2970-E2979 [PMID: 29531073 DOI: 10.1073/pnas.1717139115]

26 **Huang Z**, Shi M, Zhou H, Wang J, Zhang HJ, Shi J-. Prognostic signature of lung adenocarcinoma based on stem cell-related genes. *Sci Rep* 2021; **11**: 1687 [PMID: 33462260 DOI: 10.1038/s41598-020-80453-4]

27 **Ding Z**, Liu Y, Huang Q, Cheng C, Song L, Zhang C, Cui X, Yan W, Han Y, Zhang H. m6A- and immune-related lncRNA signature confers robust predictive power for immune efficacy in lung squamous cell carcinoma. *VIEW* 2023; **4** [DOI: 10.1002/VIW.20220083]

28 **Yang WL**, Lee YE, Chen MH, Chao KM, Huang CY. In-silico drug screening and potential target identification for hepatocellular carcinoma using Support Vector Machines based on drug screening result. *Gene* 2013; **518**: 201-208 [PMID: 23220021 DOI: 10.1016/j.gene.2012.11.030]

29 **Liang JD**, Ping XO, Tseng YJ, Huang GT, Lai F, Yang PM. Recurrence predictive models for patients with hepatocellular carcinoma after radiofrequency ablation using support vector machines with feature selection methods. *Comput Methods Programs Biomed* 2014; **117**: 425-434 [PMID: 25278224 DOI: 10.1016/j.cmpb.2014.09.001]

30 **Nayarisseri A**, Khandelwal R, Tanwar P, Madhavi M, Sharma D, Thakur G, Speck-Planche A, Singh SK. Artificial Intelligence, Big Data and Machine Learning Approaches in Precision Medicine & Drug Discovery. *Curr Drug Targets* 2021; **22**: 631-655 [PMID: 33397265 DOI: 10.2174/1389450122999210104205732]

31 **Shi J**, Liu J, Tu X, Li B, Tong Z, Wang T, Zheng Y, Shi H, Zeng X, Chen W, Yin W, Fang W. Single-cell immune signature for detecting early-stage HCC and early assessing anti-PD-1 immunotherapy efficacy. *J Immunother Cancer* 2022; **10** [PMID: 35101942 DOI: 10.1136/jitc-2021-003133]

32 **Zeng Q**, Klein C, Caruso S, Maille P, Laleh NG, Sommacale D, Laurent A, Amaddeo G, Gentien D, Rapinat A, Regnault H, Charpy C, Nguyen CT, Tournigand C, Brustia R, Pawlotsky JM, Kather JN, Maiuri MC, Loménie N, Calderaro J. Artificial intelligence predicts immune and inflammatory gene signatures directly from hepatocellular carcinoma histology. *J Hepatol* 2022; **77**: 116-127 [PMID: 35143898 DOI: 10.1016/j.jhep.2022.01.018]

33 **Gong Q**, Chen X, Liu F, Cao Y. Machine learning-based integration develops a neutrophil-derived signature for improving outcomes in hepatocellular carcinoma. *Front Immunol* 2023; **14**: 1216585 [PMID: 37575244 DOI: 10.3389/fimmu.2023.1216585]

34 **Xie H**, Shi M, Liu Y, Cheng C, Song L, Ding Z, Jin H, Cui X, Wang Y, Yao D, Wang P, Yao M, Zhang H. Identification of m6A- and ferroptosis-related lncRNA signature for predicting immune efficacy in hepatocellular carcinoma. *Front Immunol* 2022; **13**: 914977 [PMID: 36032107 DOI: 10.3389/fimmu.2022.914977]

35 **Feng Q**, Huang Z, Song L, Wang L, Lu H, Wu L. Combining bulk and single-cell RNA-sequencing data to develop an NK cell-related prognostic signature for hepatocellular carcinoma based on an integrated machine learning framework. *Eur J Med Res* 2023; **28**: 306 [PMID: 37649103 DOI: 10.1186/s40001-023-01300-6]

36 **Dai Y**, Qiang W, Lin K, Gui Y, Lan X, Wang D. An immune-related gene signature for predicting survival and immunotherapy efficacy in hepatocellular carcinoma. *Cancer Immunol Immunother* 2021; **70**: 967-979 [PMID: 33089373 DOI: 10.1007/s00262-020-02743-0]

37 **Shen J**, Gao H, Li B, Huang Y, Shi Y. The integration of machine learning and multi-omics analysis provides a powerful approach to screen aging-related genes and predict prognosis and immunotherapy efficacy in hepatocellular carcinoma. *Aging (Albany NY)* 2023; **15**: 6848-6864 [PMID: 37517087 DOI: 10.18632/aging.204876]

38 **Konerman MA**, Beste LA, Van T, Liu B, Zhang X, Zhu J, Saini SD, Su GL, Nallamothu BK, Ioannou GN, Waljee AK. Machine learning models to predict disease progression among veterans with hepatitis C virus. *PLoS One* 2019; **14**: e0208141 [PMID: 30608929 DOI: 10.1371/journal.pone.0208141]

39 **Zeng J**, Zeng J, Lin K, Lin H, Wu Q, Guo P, Zhou W, Liu J. Development of a machine learning model to predict early recurrence for hepatocellular carcinoma after curative resection. *Hepatobiliary Surg Nutr* 2022; **11**: 176-187 [PMID: 35464276 DOI: 10.21037/hbsn-20-466]

40 **Liang J**, Zhang W, Yang J, Wu M, Dai Q, Yin H, Xiao Y, Kong L. Deep learning supported discovery of biomarkers for clinical prognosis of liver cancer. *Nat Mach Intell* 2023; **5**: 408-420 [DOI: 10.1038/s42256-023-00635-3]

41 **Hatanaka T**, Naganuma A, Hiraoka A, Tada T, Hirooka M, Kariyama K, Tani J, Atsukawa M, Takaguchi K, Itobayashi E, Fukunishi S, Tsuji K, Ishikawa T, Tajiri K, Ochi H, Yasuda S, Toyoda H, Ogawa C, Nishimura T, Shimada N, Kawata K, Kosaka H, Kakizaki S, Tanaka T, Ohama H, Nouso K, Morishita A, Tsutsui A, Nagano T, Itokawa N, Okubo T, Arai T, Imai M, Koizumi Y, Nakamura S, Kaibori M, Iijima H, Hiasa Y, Kumada T; Real-life Practice Experts for HCC (RELPEC) Study Group, and HCC 48 Group (hepatocellular carcinoma experts from 48 clinics in Japan). The hepatocellular carcinoma modified Gustave Roussy Immune score (HCC-GRIm score) as a novel prognostic score for patients treated with atezolizumab and bevacizumab: A multicenter retrospective analysis. *Cancer Med* 2023; **12**: 4259-4269 [PMID: 36156452 DOI: 10.1002/cam4.5294]

42 **Sangro B**, Melero I, Wadhawan S, Finn RS, Abou-Alfa GK, Cheng AL, Yau T, Furuse J, Park JW, Boyd Z, Tang HT, Shen Y, Tschaika M, Neely J, El-Khoueiry A. Association of inflammatory biomarkers with clinical outcomes in nivolumab-treated patients with advanced hepatocellular carcinoma. *J Hepatol* 2020; **73**: 1460-1469 [PMID: 32710922 DOI: 10.1016/j.jhep.2020.07.026]

43 **Zhu AX**, Abbas AR, de Galarreta MR, Guan Y, Lu S, Koeppen H, Zhang W, Hsu CH, He AR, Ryoo BY, Yau T, Kaseb AO, Burgoyne AM, Dayyani F, Spahn J, Verret W, Finn RS, Toh HC, Lujambio A, Wang Y. Molecular correlates of clinical response and resistance to atezolizumab in combination with bevacizumab in advanced hepatocellular carcinoma. *Nat Med* 2022; **28**: 1599-1611 [PMID: 35739268 DOI: 10.1038/s41591-022-01868-2]

44 **Zeng Q**, Klein C, Caruso S, Maille P, Allende DS, Mínguez B, Iavarone M, Ningarhari M, Casadei-Gardini A, Pedica F, Rimini M, Perbellini R, Boulagnon-Rombi C, Heurgué A, Maggioni M, Rela M, Vij M, Baulande S, Legoix P, Lameiras S; HCC-AI study group, Bruges L, Gnemmi V, Nault JC, Campani C, Rhee H, Park YN, Iñarrairaegui M, Garcia-Porrero G, Argemi J, Sangro B, D'Alessio A, Scheiner B, Pinato DJ, Pinter M, Paradis V, Beaufrère A, Peter S, Rimassa L, Di Tommaso L, Vogel A, Michalak S, Boursier J, Loménie N, Ziol M, Calderaro J. Artificial intelligence-based pathology as a biomarker of sensitivity to atezolizumab-bevacizumab in patients with hepatocellular carcinoma: a multicentre retrospective study. *Lancet Oncol* 2023; **24**: 1411-1422 [PMID: 37951222 DOI: 10.1016/S1470-2045(23)00468-0]

45 **Sun R**, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, Verlingue L, Brandao D, Lancia A, Ammari S, Hollebecque A, Scoazec JY, Marabelle A, Massard C, Soria JC, Robert C, Paragios N, Deutsch E, Ferté C. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol* 2018; **19**: 1180-1191 [PMID: 30120041 DOI: 10.1016/S1470-2045(18)30413-3]

46 **Zhang W**, Tong S, Hu B, Wan T, Tang H, Zhao F, Jiao T, Li J, Zhang Z, Cai J, Ye H, Wang Z, Chen S, Wang Y, Li X, Wang F, Cao J, Tian L, Zhao X, Chen M, Wang H, Cai S, Hu M, Bai Y, Lu S. Lenvatinib plus anti-PD-1 antibodies as conversion therapy for patients with unresectable intermediate-advanced hepatocellular carcinoma: a single-arm, phase II trial. *J Immunother Cancer* 2023; **11** [PMID: 37730273 DOI: 10.1136/jitc-2023-007366]

47 **Topol EJ**. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019; **25**: 44-56 [PMID: 30617339 DOI: 10.1038/s41591-018-0300-7]

48 **Holzinger A**, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov* 2019; **9**: e1312 [PMID: 32089788 DOI: 10.1002/widm.1312]

49 **Char DS**, Shah NH, Magnus D. Implementing Machine Learning in Health Care - Addressing Ethical Challenges. *N Engl J Med* 2018; **378**: 981-983 [PMID: 29539284 DOI: 10.1056/NEJMp1714229]

50 **Komura D**, Ishikawa S. Machine Learning Methods for Histopathological Image Analysis. *Comput Struct Biotechnol J* 2018; **16**: 34-42 [PMID: 30275936 DOI: 10.1016/j.csbj.2018.01.001]

51 **Esteva A**, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J. A guide to deep learning in healthcare. *Nat Med* 2019; **25**: 24-29 [PMID: 30617335 DOI: 10.1038/s41591-018-0316-z]

52 **Siravegna G**, Marsoni S, Siena S, Bardelli A. Integrating liquid biopsies into the management of cancer. *Nat Rev Clin Oncol* 2017; **14**: 531-548 [PMID: 28252003 DOI: 10.1038/nrclinonc.2017.14]

53 **Qin R**, Zhao H, He Q, Li F, Li Y. Advances in single-cell sequencing technology in the field of hepatocellular carcinoma. *Front Genet* 2022; **13**: 996890 [PMID: 36303541 DOI: 10.3389/fgene.2022.996890]

54 **Lu J**, Chen Y, Zhang X, Guo J, Xu K, Li L. A novel prognostic model based on single-cell RNA sequencing data for hepatocellular carcinoma. *Cancer Cell Int* 2022; **22**: 38 [PMID: 35078458 DOI: 10.1186/s12935-022-02469-2]

55 **Haibe-Kains B**, Adam GA, Hosny A, Khodakarami F; Massive Analysis Quality Control (MAQC) Society Board of Directors, Waldron L, Wang B, McIntosh C, Goldenberg A, Kundaje A, Greene CS, Broderick T, Hoffman MM, Leek JT, Korthauer K, Huber W, Brazma A, Pineau J, Tibshirani R, Hastie T, Ioannidis JPA, Quackenbush J, Aerts HJWL. Transparency and reproducibility in artificial intelligence. *Nature* 2020; **586**: E14-E16 [PMID: 33057217 DOI: 10.1038/s41586-020-2766-y]

56 **Wang T**, Dai L, Shen S, Yang Y, Yang M, Yang X, Qiu Y, Wang W. Comprehensive Molecular Analyses of a Macrophage-Related Gene Signature With Regard to Prognosis, Immune Features, and Biomarkers for Immunotherapy in Hepatocellular Carcinoma Based on WGCNA and the LASSO Algorithm. *Front Immunol* 2022; **13**: 843408 [PMID: 35693827 DOI: 10.3389/fimmu.2022.843408]

57 **Xiang S**, Li J, Shen J, Zhao Y, Wu X, Li M, Yang X, Kaboli PJ, Du F, Zheng Y, Wen Q, Cho CH, Yi T, Xiao Z. Identification of Prognostic Genes in the Tumor Microenvironment of Hepatocellular Carcinoma. *Front Immunol* 2021; **12**: 653836 [PMID: 33897701 DOI: 10.3389/fimmu.2021.653836]

58 **Murai H**, Kodama T, Maesaka K, Tange S, Motooka D, Suzuki Y, Shigematsu Y, Inamura K, Mise Y, Saiura A, Ono Y, Takahashi Y, Kawasaki Y, Iino S, Kobayashi S, Idogawa M, Tokino T, Hashidate-Yoshida T, Shindou H, Miyazaki M, Imai Y, Tanaka S, Mita E, Ohkawa K, Hikita H, Sakamori R, Tatsumi T, Eguchi H, Morii E, Takehara T. Multiomics identifies the link between intratumor steatosis and the exhausted tumor immune microenvironment in hepatocellular carcinoma. *Hepatology* 2023; **77**: 77-91 [PMID: 35567547 DOI: 10.1002/hep.32573]

59 **Zhang FP**, Huang YP, Luo WX, Deng WY, Liu CQ, Xu LB, Liu C. Construction of a risk score prognosis model based on hepatocellular carcinoma microenvironment. *World J Gastroenterol* 2020; **26**: 134-153 [PMID: 31969776 DOI: 10.3748/wjg.v26.i2.134]

Artificial Intelligence in Gastroenterology

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

Artif Intell Gastroenterol 2024 April 30; 5(1): 91607

DOI: 10.35712/aig.v5.i1.91607

ISSN 2644-3236 (online)

EDITORIAL

Scope and caveats: Artificial intelligence in gastroenterology

Gumpeny Ramachandra Sridhar, Atmakuri V Siva Prasad, Gumpeny Lakshmi

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

P-Reviewer: Soreq L, United Kingdom

Received: January 23, 2024
Peer-review started: January 23, 2024
First decision: February 6, 2024
Revised: February 18, 2024
Accepted: March 29, 2024
Article in press: March 29, 2024
Published online: April 30, 2024

Gumpeny Ramachandra Sridhar, Department of Endocrinology and Diabetes, Endocrine and Diabetes Centre, Visakhapatnam 530002, India

Atmakuri V Siva Prasad, Department of Gastroenterology, Institute of Gastroenterology, Visakhapatnam 530003, India

Gumpeny Lakshmi, Department of Internal Medicine, Gayatri Vidya Parishad Institute of Healthcare & Medical Technology, Visakhapatnam 530048, India

Corresponding author: Gumpeny Ramachandra Sridhar, FRCP (Hon), MD, Consultant Physician-Scientist, Department of Endocrinology and Diabetes, Endocrine and Diabetes Centre, 15-12-15 Krishnanagar, Visakhapatnam 530002, India. grsridhar@hotmail.com

Abstract

The use of Artificial intelligence (AI) has evolved from its mid-20th century origins to playing a pivotal tool in modern medicine. It leverages digital data and computational hardware for diverse applications, including diagnosis, prognosis, and treatment responses in gastrointestinal and hepatic conditions. AI has had an impact in diagnostic techniques, particularly endoscopy, ultrasound, and histopathology. AI encompasses machine learning, natural language processing, and robotics, with machine learning being central. This involves sophisticated algorithms capable of managing complex datasets, far surpassing traditional statistical methods. These algorithms, both supervised and unsupervised, are integral for interpreting large datasets. In liver diseases, AI's non-invasive diagnostic applications, particularly in non-alcoholic fatty liver disease, and its role in characterizing hepatic lesions is promising. AI aids in distinguishing between normal and cirrhotic livers and improves the accuracy of lesion characterization and prognostication of hepatocellular carcinoma. AI enhances lesion identification during endoscopy, showing potential in the diagnosis and management of early-stage esophageal carcinoma. In peptic ulcer disease, AI technologies influence patient management strategies. AI is useful in colonoscopy, particularly in detecting smaller colonic polyps. However, its applicability in non-academic settings requires further validation. Addressing these issues is vital for harnessing the potential of AI. In conclusion, while AI offers transformative possibilities in gastroenterology, careful integration and balancing of technical possibilities with ethical and practical application, is essential for optimal use.

Key Words: Machine learning; Neural networks; Diagnosis; Work-flow; Ethics; Image; Polyps; Hepatoma

Core Tip: Artificial intelligence helps in the early identification, management and prognostication of gastrointestinal diseases through applications in endoscopy and histopathological interpretation. Proof of concept studies exist for all of these, but need validation by randomized clinical trials before they can be incorporated into clinical work flow.

Citation: Sridhar GR, Siva Prasad AV, Lakshmi G. Scope and caveats: Artificial intelligence in gastroenterology. *Artif Intell Gastroenterol* 2024; 5(1): 91607
URL: https://www.wjgnet.com/2644-3236/full/v5/i1/91607.htm
DOI: https://dx.doi.org/10.35712/aig.v5.i1.91607

INTRODUCTION

Artificial intelligence (AI) has origins dating back to the mid-20th century[1]. Its application, particularly in gastroenterology, has been more recent, being dependent on the availability of digital data and powerful computational hardware. AI is now used in diagnosing and predicting treatment responses for a spectrum of gastrointestinal and hepatic disorders [2-5].

AI encompasses multiple fields and is often defined as a computer's ability to perform tasks requiring human-like cognition[5], or as a machine that simulates complex human thinking[2].

In contrast to traditional statistical methods, AI processes data using a large number of variables and sophisticated formulas, making it possible to perform otherwise impractical analyses[2].

NOMENCLATURE AND CLASSIFICATION OF AI

AI broadly includes three areas: (1) Machine learning, which involves artificial neural networks, deep learning, and convolutional neural networks; (2) Natural language processing; and (3) Robotics. Machine learning techniques are principally used in gastroenterology[6].

Machine learning can be considered an advanced statistical approach to uncover relationships among various parameters; it utilizes algorithms such as linear regression for predicting relationships between variables, and classification algorithms like support vector machines and random forests for categorizing data.

Neural networks are more complex, utilizing nodes to determine calculated parameters, thereby allowing the use of intricate formulas. Deep learning networks, which are multi-layered, enable more advanced learning, are used in image processing[2].

Machine learning models are categorized into supervised and unsupervised. Supervised models label each sample, while unsupervised models aim to discover data structures without labels.

APPLICATIONS OF AI IN GASTROENTEROLOGY

AI's applications in gastroenterology promise to enhance patient care by reducing diagnostic errors[6]. They are employed in various conditions including gastritis, gastrointestinal bleeding, gastric malignancy, non-alcoholic fatty liver disease (NAFLD), cirrhosis, inflammatory bowel disease, colorectal polyps and cancer, and computer-aided endoscopy. Other potential applications include Helicobacter pylori infection, celiac disease, and pancreatic lesions[5]. A MEDLINE database search indicates that China and the USA are leading in AI research in this field[3].

LIVER DISEASES

In liver diseases, AI applications span from hepatocellular carcinoma to NAFLD, benign tumours, and viral hepatitis.

Non-invasive methods like ultrasound or transient elastography are used to identify NAFLD, now classified as metabolic dysfunction-associated steatohepatitis[7]. Probabilistic neural networks differentiate normal livers from those with cirrhosis in NAFLD patients, with the gold standard being liver biopsy. The area under the curve for this method ranges between 0.857 and 0.901[8]. AI also aids in automating histopathological examination, achieving high accuracy in characterizing alterations found in NAFLD[9]. Predictive models using multiple data sources, including electronic medical records, imaging, and biomarkers, have improved accuracy in identifying at-risk patients[10,11].

Hepatic mass lesions can be interpreted with high accuracy by the use of AI. Deep learning methods achieved receiver operating curves of 0.93 in lesion differentiation and 0.916 in characterization[12]. AI also aids in prognosticating established hepatocellular carcinoma[11] and predicting graft failure following liver transplantation[4].

However, when compared with standard scores, AI did not significantly improve the accuracy of short-term predictions for readmission and mortality risks in patients with cirrhosis[13].

ESOPHAGEAL LESIONS

AI methodologies improved the identification of suspicious lesions during upper GI endoscopy; this was reported in the differentiation of dysplasia and early neoplastic changes in Barrett's esophagus[14]. Other technologies such as white-light endoscopy/narrow band imaging, wide-area transepithelial sampling, and volumetric laser endomicroscopy lend themselves to machine learning[14]. As of 2020, AI algorithms were shown to be effective in diagnosing and thereby improving the outcomes of early-stage esophageal carcinoma.

ACID PEPTIC DISORDERS

In peptic ulcer disease, AI is useful in diagnosis, management, and complications[15]. Helicobacter pylori, a significant pathogenic factor, can be identified using AI. The first application was reported in 2004[16]; recent studies employ convoluted neural network models on large datasets, achieving high sensitivity, specificity, and accuracy[17].

AI also assists in diagnosing and differentiating peptic ulcers with wireless capsule endoscopic images by the use of deep learning[18,19]. It achieved an overall sensitivity of 89.7%. AI can also help identify infections, ulcers, polyps, and submucosal xanthomas[20,21].

For rare cases requiring surgical intervention, AI finds application in robot-assisted minimally invasive surgery, and in predicting complications like bleeding and perforation by the use of data from electronic medical records[15,22].

COLONOSCOPY

AI aids in classifying structures and has notably improved polyp detection rates, particularly in identifying polyps less than 5cm in size[2]; these are often missed by conventional procedures.

AI is also effective in grading remission of ulcerative colitis[23,24] and in assessing video capsule endoscopy for ulcers and bleeding detection, which is a time-consuming task[25].

However, mixed results were reported for polyp detection using computer-aided endoscopy in non-academic community-based practice[26], indicating the need for further studies[27].

PROS AND CONS OF AI IN GASTROENTEROLOGY

The potential of AI to enhance quality of care is significant, but integration into clinical workflows remains a challenge [28]. In specific tasks, AI-based devices match or even surpass expert gastroenterologists in identifying and differentiating neoplasms in the gastrointestinal tract. However, for adoption into routine clinical practice, randomized trial validation is necessary[28].

Other factors such as disease prevalence, physician competence, and human-machine interaction also affect AI's clinical benefit[29]. Despite the availability of commercial AI tools in the USA and Europe, their integration into clinical workflows is still a work in progress before the full potential is realized[30].

ETHICAL AND LEGAL ASPECTS

The growing use of AI in clinical practice brings ethical and legal challenges such as data privacy and security, method reliability and safety, and ensuring fairness, inclusivity, transparency, and accountability[31]. The extent of reliance on AI for decision-making is a key consideration[28], given that the ultimate responsibility rests on the end user, viz the physician.

AI has the potential for bias in clinical problem selection, variable choices, algorithm development, and system use [32]. In gastroenterology, as in other fields, training sets must be inclusive and diverse to avoid bias in diagnosing diseases with varying prevalence rates.

To mitigate potential biases, health equity goals should be incorporated early in algorithm development by involving technically diverse research teams. Regulatory standards should include pre-deployment audits to ensure algorithmic performance equality[32].

CONCLUSION

AI in gastroenterology has primarily been applied in endoscopic image analysis, radiology, and histopathology to aid in early detection of lesions and appropriate treatment. While its role in diagnostic endoscopy is expanding, evidence for improved clinical outcomes in real-life scenarios remains to be established.

Issues surrounding human-machine interaction, AI integration into clinical culture and practice, and the balance between AI-assisted management and practitioner skill maintenance need to be addressed[33]. Generative AI, such as ChatGPT, launched in late November 2022[34], has become pervasive in medicine, including gastroenterology. While beneficial in diagnosis in complex scenarios[35], privacy and legal concerns arise, especially when scientific publications could eventually heavily on AI-generated results[36].

ACKNOWLEDGEMENTS

We thank Mr Venkat Yarabati for assistance in the preparation of this manuscript.

FOOTNOTES

Author contributions: All three authors contributed to this manuscript; Sridhar GR designed the concept and outline; Siva Prasad AV provided inputs to the discussion and design; Lakshmi G contributing to the writing and editing of the manuscript.

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

Country/Territory of origin: India

ORCID number: Gumpeny Ramachandra Sridhar 0000-0002-7446-1251; Gumpeny Lakshmi 0000-0002-1368-745X.

S-Editor: Gong ZM
L-Editor: A
P-Editor: Zhao YQ

REFERENCES

1 **Turing AM**. I.—Computing machinery and intelligence. *Mind* 1950; **49**: 433-460 [DOI: 10.1093/mind/LIX.236.433]
2 **Moldoveanu AC**, Fierbinteanu-Braticevici C. A Primer into the Current State of Artificial Intelligence in Gastroenterology. *J Gastrointestin Liver Dis* 2022; **31**: 244-253 [PMID: 35694986 DOI: 10.15403/jgld-4180]
3 **Nam D**, Chapiro J, Paradis V, Seraphin TP, Kather JN. Artificial intelligence in liver diseases: Improving diagnostics, prognostics and response prediction. *JHEP Rep* 2022; **4**: 100443 [PMID: 35243281 DOI: 10.1016/j.jhepr.2022.100443]
4 **Stan-Ilie M**, Sandru V, Constantinescu G, Plotogea OM, Rinja EM, Tincu IF, Jichitu A, Carasel AE, Butuc AC, Popa B. Artificial Intelligence-The Rising Star in the Field of Gastroenterology and Hepatology. *Diagnostics (Basel)* 2023; **13** [PMID: 36832150 DOI: 10.3390/diagnostics13040662]
5 **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]
6 **Poalelungi DG**, Musat CL, Fulga A, Neagu M, Neagu AI, Piraianu AI, Fulga I. Advancing Patient Care: How Artificial Intelligence Is Transforming Healthcare. *J Pers Med* 2023; **13** [PMID: 37623465 DOI: 10.3390/jpm13081214]
7 **Rinella ME**, Lazarus JV, Ratziu V, Francque SM, Sanyal AJ, Kanwal F, Romero D, Abdelmalek MF, Anstee QM, Arab JP, Arrese M, Bataller R, Beuers U, Boursier J, Bugianesi E, Byrne CD, Castro Narro GE, Chowdhury A, Cortez-Pinto H, Cryer DR, Cusi K, El-Kassas M, Klein S, Eskridge W, Fan J, Gawrieh S, Guy CD, Harrison SA, Kim SU, Koot BG, Korenjak M, Kowdley KV, Lacaille F, Loomba R, Mitchell-Thain R, Morgan TR, Powell EE, Roden M, Romero-Gómez M, Silva M, Singh SP, Sookoian SC, Spearman CW, Tiniakos D, Valenti L, Vos MB, Wong VW, Xanthakos S, Yilmaz Y, Younossi Z, Hobbs A, Villota-Rivas M, Newsome PN; NAFLD Nomenclature consensus group. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. *J Hepatol* 2023; **79**: 1542-1556 [PMID: 37364790 DOI: 10.1016/j.jhep.2023.06.003]
8 **Lee JH**, Joo I, Kang TW, Paik YH, Sinn DH, Ha SY, Kim K, Choi C, Lee G, Yi J, Bang WC. Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network. *Eur Radiol* 2020; **30**: 1264-1273 [PMID: 31478087 DOI: 10.1007/s00330-019-06407-1]
9 **Teramoto T**, Shinohara T, Takiyama A. Computer-aided classification of hepatocellular ballooning in liver biopsies from patients with NASH using persistent homology. *Comput Methods Programs Biomed* 2020; **195**: 105614 [PMID: 32650090 DOI: 10.1016/j.cmpb.2020.105614]
10 **Pournik O**, Dorri S, Zabolinezhad H, Alavian SM, Eslami S. A diagnostic model for cirrhosis in patients with non-alcoholic fatty liver disease: an artificial neural network approach. *Med J Islam Repub Iran* 2014; **28**: 116 [PMID: 25678995]

11 **Calderaro J**, Seraphin TP, Luedde T, Simon TG. Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma. *J Hepatol* 2022; **76**: 1348-1361 [PMID: 35589255 DOI: 10.1016/j.jhep.2022.01.014]

12 **Schmauch B**, Herent P, Jehanno P, Dehaene O, Saillard C, Aubé C, Luciani A, Lassau N, Jégou S. Diagnosis of focal liver lesions from ultrasound using deep learning. *Diagn Interv Imaging* 2019; **100**: 227-233 [PMID: 30926443 DOI: 10.1016/j.diii.2019.02.009]

13 **Hu C**, Anjur V, Saboo K, Reddy KR, O'Leary J, Tandon P, Wong F, Garcia-Tsao G, Kamath PS, Lai JC, Biggins SW, Fallon MB, Thuluvath P, Subramanian RM, Maliakkal B, Vargas H, Thacker LR, Iyer RK, Bajaj JS. Low Predictability of Readmissions and Death Using Machine Learning in Cirrhosis. *Am J Gastroenterol* 2021; **116**: 336-346 [PMID: 33038139 DOI: 10.14309/ajg.0000000000000971]

14 **Lazăr DC**, Avram MF, Faur AC, Goldiş A, Romoşan I, Tăban S, Cornianu M. The Impact of Artificial Intelligence in the Endoscopic Assessment of Premalignant and Malignant Esophageal Lesions: Present and Future. *Medicina (Kaunas)* 2020; **56** [PMID: 32708343 DOI: 10.3390/medicina56070364]

15 **Zhao PY**, Han K, Yao RQ, Ren C, Du XH. Application Status and Prospects of Artificial Intelligence in Peptic Ulcers. *Front Surg* 2022; **9**: 894775 [PMID: 35784921 DOI: 10.3389/fsurg.2022.894775]

16 **Huang CR**, Sheu BS, Chung PC, Yang HB. Computerized diagnosis of Helicobacter pylori infection and associated gastric inflammation from endoscopic images by refined feature selection using a neural network. *Endoscopy* 2004; **36**: 601-608 [PMID: 15243882 DOI: 10.1055/s-2004-814519]

17 **Shichijo S**, Nomura S, Aoyama K, Nishikawa Y, Miura M, Shinagawa T, Takiyama H, Tanimoto T, Ishihara S, Matsuo K, Tada T. Application of Convolutional Neural Networks in the Diagnosis of Helicobacter pylori Infection Based on Endoscopic Images. *EBioMedicine* 2017; **25**: 106-111 [PMID: 29056541 DOI: 10.1016/j.ebiom.2017.10.014]

18 **Al-Kasasbeh R**, Korenevskiy N, Alshamasin M, Ionescou F, Smith A. Prediction of gastric ulcers based on the change in electrical resistance of acupuncture points using fuzzy logic decision-making. *Comput Methods Biomech Biomed Engin* 2013; **16**: 302-313 [PMID: 22292589 DOI: 10.1080/10255842.2011.618926]

19 **Wang S**, Xing Y, Zhang L, Gao H, Zhang H. A systematic evaluation and optimization of automatic detection of ulcers in wireless capsule endoscopy on a large dataset using deep convolutional neural networks. *Phys Med Biol* 2019; **64**: 235014 [PMID: 31645019 DOI: 10.1088/1361-6560/ab5086]

20 **Mohammad F**, Al-Razgan M. Deep Feature Fusion and Optimization-Based Approach for Stomach Disease Classification. *Sensors (Basel)* 2022; **22** [PMID: 35408415 DOI: 10.3390/s22072801]

21 **Xia J**, Xia T, Pan J, Gao F, Wang S, Qian YY, Wang H, Zhao J, Jiang X, Zou WB, Wang YC, Zhou W, Li ZS, Liao Z. Use of artificial intelligence for detection of gastric lesions by magnetically controlled capsule endoscopy. *Gastrointest Endosc* 2021; **93**: 133-139.e4 [PMID: 32470426 DOI: 10.1016/j.gie.2020.05.027]

22 **Gao S**, Ji S, Feng M, Lu X, Tong W. A study on autonomous suturing task assignment in robot-assisted minimally invasive surgery. *Int J Med Robot* 2021; **17**: 1-10 [PMID: 33049099 DOI: 10.1002/rcs.2180]

23 **Stidham RW**, Liu W, Bishu S, Rice MD, Higgins PDR, Zhu J, Nallamothu BK, Waljee AK. Performance of a Deep Learning Model *vs* Human Reviewers in Grading Endoscopic Disease Severity of Patients With Ulcerative Colitis. *JAMA Netw Open* 2019; **2**: e193963 [PMID: 31099869 DOI: 10.1001/jamanetworkopen.2019.3963]

24 **Gottlieb K**, Requa J, Karnes W, Chandra Gudivada R, Shen J, Rael E, Arora V, Dao T, Ninh A, McGill J. Central Reading of Ulcerative Colitis Clinical Trial Videos Using Neural Networks. *Gastroenterology* 2021; **160**: 710-719.e2 [PMID: 33098883 DOI: 10.1053/j.gastro.2020.10.024]

25 **Aoki T**, Yamada A, Aoyama K, Saito H, Tsuboi A, Nakada A, Niikura R, Fujishiro M, Oka S, Ishihara S, Matsuda T, Tanaka S, Koike K, Tada T. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointest Endosc* 2019; **89**: 357-363.e2 [PMID: 30670179 DOI: 10.1016/j.gie.2018.10.027]

26 **Wei MT**, Shankar U, Parvin R, Abbas SH, Chaudhary S, Friedlander Y, Friedland S. Evaluation of Computer-Aided Detection During Colonoscopy in the Community (AI-SEE): A Multicenter Randomized Clinical Trial. *Am J Gastroenterol* 2023; **118**: 1841-1847 [PMID: 36892545 DOI: 10.14309/ajg.0000000000002239]

27 **Berzin TM**, Glissen Brown J. Navigating the "Trough of Disillusionment" for CADe Polyp Detection: What Can We Learn About Negative AI Trials and the Physician-AI Hybrid? *Am J Gastroenterol* 2023; **118**: 1743-1745 [PMID: 37141122 DOI: 10.14309/ajg.0000000000002286]

28 **Hassan C**, Mori Y, Sharma P. The Pros and Cons of Artificial Intelligence in Endoscopy. *Am J Gastroenterol* 2023; **118**: 1720-1722 [PMID: 37052360 DOI: 10.14309/ajg.0000000000002287]

29 **Frazzoni L**, Arribas J, Antonelli G, Libanio D, Ebigbo A, van der Sommen F, de Groof AJ, Fukuda H, Ohmori M, Ishihara R, Wu L, Yu H, Mori Y, Repici A, Bergman JJGHM, Sharma P, Messmann H, Hassan C, Fuccio L, Dinis-Ribeiro M. Endoscopists' diagnostic accuracy in detecting upper gastrointestinal neoplasia in the framework of artificial intelligence studies. *Endoscopy* 2022; **54**: 403-411 [PMID: 33951743 DOI: 10.1055/a-1500-3730]

30 **Pecere S**, Antonelli G, Dinis-Ribeiro M, Mori Y, Hassan C, Fuccio L, Bisschops R, Costamagna G, Jin EH, Lee D, Misawa M, Messmann H, Iacopini F, Petruzziello L, Repici A, Saito Y, Sharma P, Yamada M, Spada C, Frazzoni L. Endoscopists performance in optical diagnosis of colorectal polyps in artificial intelligence studies. *United European Gastroenterol J* 2022; **10**: 817-826 [PMID: 35984903 DOI: 10.1002/ueg2.12285]

31 **Sridhar GR**, Lakshmi G. Ethical Issues of Artificial Intelligence in Diabetes Mellitus. *Med Res Arch* 2023; **11** [DOI: 10.18103/mra.v11i8.4287]

32 **Uche-Anya E**, Anyane-Yeboa A, Berzin TM, Ghassemi M, May FP. Artificial intelligence in gastroenterology and hepatology: how to advance clinical practice while ensuring health equity. *Gut* 2022; **71**: 1909-1915 [PMID: 35688612 DOI: 10.1136/gutjnl-2021-326271]

33 **London AJ**. Artificial intelligence in medicine: Overcoming or recapitulating structural challenges to improving patient care? *Cell Rep Med* 2022; **3**: 100622 [PMID: 35584620 DOI: 10.1016/j.xcrm.2022.100622]

34 **Ghassemi M**, Birhane A, Bilal M, Kankaria S, Malone C, Mollick E, Tustumi F. ChatGPT one year on: who is using it, how and why? *Nature* 2023; **624**: 39-41 [PMID: 38036860 DOI: 10.1038/d41586-023-03798-6]

35 **Eriksen AV**, Möller S, Ryg J. Use of GPT-4 to diagnose complex clinical cases. *NEJM AI* 2023; **1**: AIp2300031 [DOI: 10.1056/aip2300031]

36 **Ashraf H**, Ashfaq H. The Role of ChatGPT in Medical Research: Progress and Limitations. *Ann Biomed Eng* 2024; **52**: 458-461 [PMID: 37452215 DOI: 10.1007/s10439-023-03311-0]

Artificial Intelligence in Gastroenterology

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

Artif Intell Gastroenterol 2024 April 30; 5(1): 90503

DOI: 10.35712/aig.v5.i1.90503

ISSN 2644-3236 (online)

ORIGINAL ARTICLE

Observational Study

Evaluating the accuracy and reproducibility of ChatGPT-4 in answering patient questions related to small intestinal bacterial overgrowth

Lauren Schlussel, Jamil S Samaan, Yin Chan, Bianca Chang, Yee Hui Yeo, Wee Han Ng, Ali Rezaie

Specialty type: Gastroenterology and hepatology

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

Peer-review model: Single blind

Peer-review report's classification

Scientific Quality: Grade C, Grade C, Grade C, Grade D, Grade D

Novelty: Grade A, Grade B, Grade B, Grade B, Grade B

Creativity or Innovation: Grade A, Grade B, Grade B, Grade B, Grade B

Scientific Significance: Grade B, Grade B, Grade B, Grade B, Grade C

P-Reviewer: Caboclo JLF, Brazil; Wu L, China; Yu YB, China; Zhang C, China

Received: December 7, 2023

Revised: March 27, 2024

Accepted: April 16, 2024

Published online: April 30, 2024

Lauren Schlussel, Jamil S Samaan, Yin Chan, Bianca Chang, Yee Hui Yeo, Ali Rezaie, Division of Gastroenterology and Hepatology, Cedars-Sinai Medical Center, Los Angeles, CA 90048, United States

Wee Han Ng, Bristol Medical School, University of Bristol, BS8 1TH, Bristol, United Kingdom

Ali Rezaie, Medically Associated Science and Technology Program, Cedars-Sinai Medical Center, Los Angeles, CA 90048, United States

Corresponding author: Ali Rezaie, MD, MSc, FRCPC, Medical Director, Medically Associated Science and Technology Program, Cedars-Sinai Medical Center, Cedars-Sinai, 8730 Alden Drive, Thalians Bldg, #E240, Los Angeles, CA 90048, United States. Ali.rezaie@cshs.org

Abstract

BACKGROUND

Small intestinal bacterial overgrowth (SIBO) poses diagnostic and treatment challenges due to its complex management and evolving guidelines. Patients often seek online information related to their health, prompting interest in large language models, like GPT-4, as potential sources of patient education.

AIM

To investigate ChatGPT-4's accuracy and reproducibility in responding to patient questions related to SIBO.

METHODS

A total of 27 patient questions related to SIBO were curated from professional societies, Facebook groups, and Reddit threads. Each question was entered into GPT-4 twice on separate days to examine reproducibility of accuracy on separate occasions. GPT-4 generated responses were independently evaluated for accuracy and reproducibility by two motility fellowship-trained gastroenterologists. A third senior fellowship-trained gastroenterologist resolved disagreements. Accuracy of responses were graded using the scale: (1) Comprehensive; (2) Correct but inadequate; (3) Some correct and some incorrect; or (4) Completely incorrect. Two responses were generated for every question to evaluate reproducibility in accuracy.

RESULTS
In evaluating GPT-4's effectiveness at answering SIBO-related questions, it provided responses with correct information to 18/27 (66.7%) of questions, with 16/27 (59.3%) of responses graded as comprehensive and 2/27 (7.4%) responses graded as correct but inadequate. The model provided responses with incorrect information to 9/27 (33.3%) of questions, with 4/27 (14.8%) of responses graded as completely incorrect and 5/27 (18.5%) of responses graded as mixed correct and incorrect data. Accuracy varied by question category, with questions related to "basic knowledge" achieving the highest proportion of comprehensive responses (90%) and no incorrect responses. On the other hand, the "treatment" related questions yielded the lowest proportion of comprehensive responses (33.3%) and highest percent of completely incorrect responses (33.3%). A total of 77.8% of questions yielded reproducible responses.

CONCLUSION
Though GPT-4 shows promise as a supplementary tool for SIBO-related patient education, the model requires further refinement and validation in subsequent iterations prior to its integration into patient care.

Key Words: Small intestinal bacterial overgrowth; Motility; Artificial intelligence; Chat-GPT; Large language models; Patient education

Core Tip: ChatGPT-4 demonstrates promise in enhancing patient understanding of basic concepts related to small intestinal bacterial overgrowth (SIBO). However, it exhibits limitations in accurately addressing questions about the diagnosis and treatment of SIBO, which are areas where up-to-date medical guidance is crucial. As such, artificial intelligence can be beneficial for general patient education but should not replace professional medical advice, especially for conditions with complex care protocols. Continuous refinement and updating of Chat-GPT's knowledge are essential for its safe and effective application in healthcare. Rigorous scrutiny of artificial intelligence-generated content is imperative to prevent the dissemination of potentially harmful misinformation.

Citation: Schlussel L, Samaan JS, Chan Y, Chang B, Yeo YH, Ng WH, Rezaie A. Evaluating the accuracy and reproducibility of ChatGPT-4 in answering patient questions related to small intestinal bacterial overgrowth. *Artif Intell Gastroenterol* 2024; 5(1): 90503
URL: https://www.wjgnet.com/2644-3236/full/v5/i1/90503.htm
DOI: https://dx.doi.org/10.35712/aig.v5.i1.90503

INTRODUCTION

Small intestinal bacterial overgrowth (SIBO) is a medical condition characterized by an excessive amount of bacteria in the small intestine, which can lead to a variety of symptoms, including bloating, abdominal pain, diarrhea, and constipation[1]. The diagnosis and treatment of SIBO varies across institutions and by healthcare provider[2]. Though various tests exist, including glucose and lactulose breath tests and small intestine aspiration and culture, there is a lack of universal approach regarding how and when to utilize these tests, as well as how to interpret the results[3].

Due to the need for specialized tests, lack of dedicated International Classification of Diseases codes, and differences in the diagnostic methods across studies, it is challenging to estimate the prevalence of SIBO with studies showing rates ranging from 4% to 79%[2] and 38% to 84% in patients with IBS[4]. Importantly, SIBO has adverse effects on quality of life and may be associated with significant healthcare costs. Though the impact on quality of life for patients with SIBO has not been independently examined, one study showed that the presence of SIBO among patients with IBS was associated with more severe symptoms and led to a decreased quality of life[5]. Patients with IBS constitute a major proportion of patients who seek consultation in gastroenterology specialist clinic[6] and is associated with considerable healthcare resource use[7]. Given the high prevalence of SIBO among patients with IBS and its association with more severe symptoms, it's very likely that SIBO has a significant impact on patients and our healthcare system. Moreover, patients have limited access to motility specialists or physicians that are capable of managing SIBO, and may even encounter health care providers that question the legitimacy of SIBO as a medical condition[8].

The advent of artificial intelligence (AI) and natural language processing technologies has led to the development of large language models (LLMs), such as ChatGPT, which have the potential to revolutionize healthcare communication and patient education[9]. GPT-4, created by OpenAI, is able to produce easy to understand and conversational responses to inquiries by users based on their inquiries. It functions on the principle of predicting subsequent words in a sentence, much akin to an expert player in a game of 'guess the next word'[9]. There is a growing body of evidence demonstrating ChatGPT's ability to answer patient questions related to medical diseases such as cardiovascular disease, bariatric surgery and cirrhosis[10-12]. In a study comparing chatbot and physician responses, evaluators preferred chatbot answers 78.6% of the time[10]. The chatbot's responses were not only more comprehensive but also of higher quality and more empa-

thetic, with a 3.6 times higher prevalence of good or very good quality answers and a 9.8 times higher prevalence of empathetic or very empathetic responses than physicians[10]. Given the increasing trends of patients seeking healthcare related information from online sources, examining the strengths and limitations of LLMs as sources of information for patients is critical to ensuring safe, effective and responsible use of these models[13].

SIBO is a complex medical condition, with differing diagnostic and treatment approaches across institutions and healthcare providers as well as geographic variations in access to specialists. The gap in patient needs versus accessibility may lead individuals to seek information from alternative sources, such as the internet or ChatGPT. If proven safe and effective, emerging AI technologies like ChatGPT offer potential benefits in this space, providing accessible, easy to understand, and informed responses to patient inquiries, which may supplement or complement patient education provided by licensed healthcare professionals. In light of this, our study aimed to evaluate the accuracy of GPT-4 in providing accurate and reproducible responses to patient questions related to SIBO. This involved assessing the quality of information provided by the AI tool against evidence-based guidelines and expert opinions. Furthermore, our research will identify the limitations and potential risks associated with using GPT-4 as a supplementary tool for patient education and support, in order to inform the development of best practices for its implementation in the healthcare context.

MATERIALS AND METHODS

Question curation

A total of 38 patient questions related to SIBO were collected from professional societies and institutions as well as Facebook support groups ("SIBO lifestyle", "SIBO SOS Community") and the Reddit thread r/SIBO. Each question was screened to ensure it was directly related to SIBO. Questions that were not specific to SIBO or were outside the scope of typical patient concerns were excluded. Duplicate and similar questions were excluded to prevent redundancy and to ensure a broad coverage of topics. One question was removed after it was deemed incorrectly worded and containing incorrect information. The final set of 27 questions included in our study represents a diverse range of patient inquiries, covering aspects of basic knowledge, diagnosis, treatment, and other concerns related to SIBO.

ChatGPT

ChatGPT is an AI LLM developed by OpenAI, based on the GPT (Generative Pre-trained Transformer) architecture. The model was designed to generate human-like text based on input, allowing the model to answer questions, engage in conversation, and perform various tasks. ChatGPT was trained on a large corpus of text from the internet, learning grammar, facts, and some reasoning abilities. It does not have a traditional "database" to retrieve information from; instead, the model generates text based on patterns and knowledge learned from the training data. However, it is essential to note the model's knowledge is limited to data up until September 2021, lacking awareness of more recent information. The latest iteration of the model, GPT-4, was released in March of 2023 and has shown promise across multiple domains of tasks[14].

Response generation

GPT-4 was used on 4/23/23 and 4/24/23 to generate responses. Each question was entered as an individual prompt using the "New Chat" function. Each question was entered into GPT-4 twice on separate days to examine reproducibility of accuracy on separate occasions.

Response grading

Reponses to questions were first independently graded for accuracy and reproducibility by two board certified, motility fellowship-trained, academic gastroenterologist reviewers actively practicing in a tertiary medical center. The following grading scale was used to grade the accuracy of each response similar to previous publications[11,12]: (1) Comprehensive (Grade 1): The response provides a complete and thorough answer as one would expect from a board-certified gastroenterologist. This grade implies that there is no additional relevant information that a specialist would deem necessary to include; (2) Correct but inadequate (Grade 2): The response is accurate but lacks certain critical details or depth that a board-certified gastroenterologist would consider important for a patient's understanding or management of SIBO; (3) Some correct and some incorrect (Grade 3): The response contains both correct and incorrect elements, indicating partial knowledge but with significant gaps or errors that require correction; and (4) Completely incorrect (Grade 4): The response does not provide accurate information related to the question asked and is considered misleading or wrong.

Reproducibility was graded based on the similarity in accuracy of the two responses per question generated by GPT-4. Any disagreement in reproducibility or accuracy grading was resolved by a third senior board-certified, motility fellowship trained gastroenterologist reviewer with greater than 10 years of experience in the field of gastrointestinal motility.

Statistical analysis

Descriptive analysis is presented as counts and percentages. For statistical analysis purposes, questions were categorized into multiple subgroups: Basic knowledge, diagnosis, treatment, and others. All statistical analysis was performed in Excel version 2308.

RESULTS

In total, 27 questions related to SIBO were inputted into GPT-4. The model provided 16/27 (59.3%) comprehensive, 2/27 (7.4%) correct but inadequate, 5/27 (18.5%) mixed with correct and incorrect data, and 4/27 (14.8%) completely incorrect responses. When examined by category, the model provided "comprehensive" responses to 90% of "basic knowledge questions", 60% of "diagnosis" questions, and 33.3% of "treatment" questions (Table 1). The model provided reproducible responses to 21/27 (77.8%) of questions (Table 2).

Most of the "completely incorrect" responses were noted to be in the "treatment" subcategory with 33.3% (3/9) of these responses rated as "completely incorrect". For example, when asked "What probiotic strain is recommended for constipation predominant SIBO?" GPT-4 stated that there is evidence that shows certain strains of probiotics helps constipation, which is not in line with current evidence and guidelines. Importantly, the model did recommend consulting with a health professional before starting new supplements. Questions, responses, and reviewer gradings are shown in Supplementary Table 1.

DISCUSSION

SIBO is a common medical condition with variable approaches to management and diagnosis across institutions. The literature shows patients frequently pursue health-related information in lieu of their healthcare providers, with the internet emerging as a common source. Due to its user-friendly interface as well as its easy to understand and conversational responses, patients may utilize ChatGPT as a source of information regarding SIBO. In light of this, we examined ChatGPT's ability to accurately and reliably answer SIBO related questions. While the model provided comprehensive answers to 59.3% of questions, 14.8% of questions were graded as completely incorrect. Our findings show GPT-4's promising future in serving as an adjunct source of information for patient with SIBO but highlight its current limitations and need for further fine tuning, training and validation prior to incorporation into clinical care.

The model provided completely inaccurate responses to 4 (14.8%) questions and mixed correct and incorrect information to 5 (18.5%) questions, which is not in line with previous data which shows its proficiency in areas such as cirrhosis, congestive heart failure, and bariatric surgery[11,12,15]. For example, GPT-3.5 provided comprehensive responses to 86.8% of questions related to bariatric surgery and 83.2% of questions related to heart failure[11,15]. The reason for the difference in performance seen in our study may be related to the dataset used to train ChatGPT. There are well-established, thoroughly researched, and widely accepted guidelines governing the diagnosis and treatment of these conditions. Such robust guidelines offer a standardized framework, enabling ChatGPT to provide accurate and reliable responses. SIBO, however, presents a unique challenge due to its less definitive guidelines that often diverge across institutions and among physicians. Further compounding this issue is ChatGPT's knowledge constraints to information prior to 2021, restricting its ability to integrate the latest studies or consensus in the rapidly evolving field of SIBO and gut microbiome. This data limitation, paired with the inherent variability in SIBO management, showcases the system's vulnerabilities in areas where medical guidelines are either in flux or less established. Considering our analysis shows a considerable number of responses contained incorrect and potentially harmful information, this underscores the importance of exercising caution when utilizing AI-generated information in the context of patient education, particularly related to complex medical conditions like SIBO. Ongoing refinement and development of LLMs are imperative to mitigate the potential risks and enhance their potential role in patient education.

GPT-4 also showed a relatively low reproducibility, only delivering consistent accuracy of responses for 77.8% of questions. This again is in contrast with previous studies which found LLMs deliver high reproducibility of quality of responses[10-12]. Such reproducibility is critical for a tool intended to educate and inform, as consistent messaging is key in enhancing understanding, mitigating confusion and establishing trust among users.

Examining GPT-4's accuracy across different domains of patient questions allowed for a more granular analysis of its performance. In line with previous studies examining ChatGPT's knowledge in cirrhosis and hepatocellular carcinoma, bariatric surgery, and heart failure[11,12,15], we found GPT-4 provided comprehensive and accurate responses to the vast majority of basic knowledge questions. This suggests that AI has the potential to serve as a reliable resource of information for patients to enhance their basic understanding of their condition. Such an application aligns with a growing body of evidence pointing to the potential of AI in augmenting patient education[16]. However, our findings also underscore key limitations of this technology. Most notably, GPT-4's responses related to the diagnosis and treatment of SIBO contain a significant amount of inaccuracies. This finding is particularly concerning given that these areas often present the greatest challenges for patients in terms of understanding and self-management. Misinformation can lead to suboptimal patient decision-making and potential harm. It underlines the importance of caution when using AI for health-related advice and re-emphasizes the need for these tools to be used in tandem with professional medical guidance[9]. This suggests that while LLMs like GPT-4 in their current form may provide beneficial support for patients looking to enhance their general understanding of a condition, they are not yet equipped to offer reliable advice on more complex aspects of medical care. It is consistent with prior research noting the limitations of AI in understanding complex diseases and suggesting tailored, expert human intervention for such scenarios[17,18].

Beyond accuracy, comprehensiveness, and reproducibility, it's important to ensure LLMs produce materials that are easy to understand by patients of all health literacy levels. There is a growing body of literature showing LLMs are able to adjust the readability of outputs when prompted[19,20]. This ensures that access to information is democratized, and patients of all health literacy levels have personalized education materials. One study showed that GPT-4 was able to improve the readability of bariatric surgery patient education materials from 12th grade-college level to 6th-9th grade[19].

Table 1 Grading of responses generated by ChatGPT-4 to questions related to small intestinal bacterial overgrowth categorized by subgroup and overall

	%
Basic knowledge (n = 10)	
Comprehensive	90
Correct but inadequate	10
Mixed with correct and incorrect data	0
Completely incorrect	0
Diagnosis (n = 5)	
Comprehensive	60
Correct but inadequate	0
Mixed with correct and incorrect data	40
Completely incorrect	0
Treatment (n = 9)	
Comprehensive	33.3
Correct but inadequate	0
Mixed with correct and incorrect data	33.3
Completely incorrect	33.3
Other (n = 3)	
Comprehensive	33.3
Correct but inadequate	33.3
Mixed with correct and incorrect data	0
Completely incorrect	33.3
Overall (n = 27)	
Comprehensive	59.3
Correct but inadequate	7.4
Mixed with correct and incorrect data	18.5
Completely incorrect	14.8

Table 2 Reproducibility of ChatGPT-4 responses overall and categorized by subgroup

	%
Overall (n = 27)	77.8
Basic knowledge (n = 10)	90
Diagnosis (n = 5)	80
Treatment (n = 9)	77.8
Other (n = 3)	33.3

Access to high quality patient education materials can also be impacted by patient language preference. Patients who prefer non-English languages have unique barriers to access to patient education materials. Some studies have shown the ability of LLMs in generating patient education materials in languages other than English with promising results[21-23]. Lastly, it's important to ensure outputs do not perpetuate known stereotypes and biases in medicine. There is a growing body of literature examining the presence of implicit bias in LLM outputs, with some studies showing LLMs may propagate racial and gender biases[24,25]. Future research should thoroughly investigate how LLMs can produce patient education materials that are not only accurate and of high quality but also accessible to patients from diverse backgrounds, with an emphasis on minimizing implicit bias and discrimination.

Limitations specific to the design of this study include the use of only two responses generated by GPT-4 to evaluate its reproducibility. While our findings provide initial insights, expanding the number of responses and questions in future research will be crucial to thoroughly assess consistency and reliability. Such expansions will help to substantiate the AI model's utility in patient education. Another limitation of this study is the use of the paid GPT-4 model over the free GPT-3.5, which was selected for its advanced linguistic capabilities and enhanced accuracy in medical contexts. While this choice aligns with our objective to evaluate the most current and sophisticated AI technology for patient education, it may affect the generalizability and accessibility of our findings. Future research could explore the trade-offs between cost and performance by comparing different AI models, including the cost-free GPT-3.5, to optimize the balance between accessibility and quality of information in AI-assisted patient care. Future studies would also benefit from exploring the differences in accuracy and reproducibility amongst different AI tools such as GPT-3.5, GPT-4, and Google Bard. For example, in a study comparing GPT-4 and Google Bard in their ability to diagnose and triage patients' ophthalmologic complaints, GPT-4 performed significantly better than Bard by generating more accurate triage suggestions, responses that experts were satisfied with for patient use, and lower potential harm rates[26]. Another study comparing GPT-3.5 and Bard in their ability to provide appropriate informational responses to patient questions regarding vascular surgery demonstrated that GPT-3.5 responses were more complete and more appropriate compared with Bard responses[27]. Similarly, GPT-3 exhibited greater accuracy and consistency over Google Bard, as well Google and Bing search engines, when addressing patient questions related to lung cancer[28]. These comparative evaluations underscore the evolving landscape of AI tools in healthcare and the importance of ongoing, meticulous analysis to harness their full potential for patient care.

Finally, we must consider other limitations of ChatGPT that pose a challenge for its future utilization in healthcare. OpenAI has not released specific details about the exact datasets used to train GPT-4. This raises concerns regarding the quality of data the model uses to respond to questions, especially when discussing healthcare related topics. The literature in healthcare is rapidly evolving and requires staying up to date with the literature to ensure good practice of medicine. ChatGPT's lack of continuous updates limits its generalized applicability in patient care. Another constraint of GPT-4 and LLMs in general is the "hallucination effect," where the model produces outputs that seem plausible and believable but are incorrect, misleading, or entirely fabricated[29,30]. This is a significant limitation that should be considered when implementing such AI tools in the healthcare setting. Our study design also has its limitations. Responses from ChatGPT were graded based on expert opinion which is subjective and prone to bias. Notably, this is a limitation across the majority of literature examining the clinical knowledge of ChatGPT, given expert opinion guided by the literature and guidelines is currently the gold standard in the practice of medicine. Our study utilized a sample of 27 patient questions, which is not inclusive of all possible patient questions pertaining to SIBO. We performed a systematic approach when curating questions to reduce the risk of selection bias. Furthermore, questions were not removed after the generation of responses from ChatGPT.

CONCLUSION

Our study underscores the potential future value of large language models, like GPT-4, in patient education related to SIBO, especially in providing basic knowledge. However, we highlight the limitations of GPT-4 in its current form due to a significant number of its responses containing inaccurate or out of date information and low reproducibility in accuracy of its responses. While AI may supplement traditional patient education methods in the future, it is not a substitute for professional medical advice. Continued evaluation and development of these technologies are crucial to harness their potential while minimizing potential harm. This iterative process will be key to the future integration of AI into healthcare systems, with the ultimate aim of improving patient understanding, engagement, and outcomes. Our research underscores the need for rigorous scrutiny and cautious application when relying on AI technologies for diseases with complex and less standardized diagnosis and treatments. Given the prevalence and impact of SIBO on patients and the healthcare system, the need for accurate, accessible patient education remains critical. Our research serves as a valuable step in identifying the challenges and opportunities for integrating AI tools in this capacity.

FOOTNOTES

Author contributions: Rezaie A was the guarantor, participated in the acquisition, analysis, and interpretation of the data, and revised the article for critically important intellectual content; Schlussel L drafted the initial manuscript and participated in the acquisition, analysis, and interpretation of the data; Samaan J designed the study and revised the article for critically important intellectual content; Chan Y and Chang B participated in the acquisition, analysis, and interpretation of the data, and revised the article for critically important intellectual content; Yeo YH revised the article for critically important intellectual content; Ng WH participated in the acquisition, analysis, and interpretation of the data.

Institutional review board statement: Our study did not require IRB approval, given our research does not involve human subjects.

Informed consent statement: Our research does not involve human subjects. Therefore, no signed informed consent or documents were obtained.

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

Data sharing statement: No additional data are available.

STROBE statement: The authors have read the STROBE Statement – checklist of items, and the manuscript was prepared and revised according to the STROBE Statement – checklist of items.

ORCID number: Lauren Schlussel 0009-0000-9008-4460; Yin Chan 0000-0003-2741-8258; Yee Hui Yeo 0000-0002-2703-5954; Ali Rezaie 0000-0002-0106-372X.

Corresponding Author's Membership in Professional Societies: American College of Gastroenterology, No. 68989.

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

REFERENCES

1 **Sachdev AH**, Pimentel M. Gastrointestinal bacterial overgrowth: pathogenesis and clinical significance. *Ther Adv Chronic Dis* 2013; **4**: 223-231 [PMID: 23997926 DOI: 10.1177/2040622313496126]
2 **Rao SSC**, Bhagatwala J. Small Intestinal Bacterial Overgrowth: Clinical Features and Therapeutic Management. *Clin Transl Gastroenterol* 2019; **10**: e00078 [PMID: 31584459 DOI: 10.14309/ctg.0000000000000078]
3 **Rezaie A**, Pimentel M, Rao SS. How to Test and Treat Small Intestinal Bacterial Overgrowth: an Evidence-Based Approach. *Curr Gastroenterol Rep* 2016; **18**: 8 [PMID: 26780631 DOI: 10.1007/s11894-015-0482-9]
4 **Posserud I**, Stotzer PO, Björnsson ES, Abrahamsson H, Simrén M. Small intestinal bacterial overgrowth in patients with irritable bowel syndrome. *Gut* 2007; **56**: 802-808 [PMID: 17148502 DOI: 10.1136/gut.2006.108712]
5 **Chuah KH**, Hian WX, Lim SZ, Beh KH, Mahadeva S. Impact of small intestinal bacterial overgrowth on symptoms and quality of life in irritable bowel syndrome. *J Dig Dis* 2023; **24**: 194-202 [PMID: 37200005 DOI: 10.1111/1751-2980.13189]
6 **Chuah KH**, Cheong SY, Lim SZ, Mahadeva S. Functional dyspepsia leads to more healthcare utilization in secondary care compared with other functional gastrointestinal disorders. *J Dig Dis* 2022; **23**: 111-117 [PMID: 35050547 DOI: 10.1111/1751-2980.13082]
7 **Canavan C**, West J, Card T. Review article: the economic impact of the irritable bowel syndrome. *Aliment Pharmacol Ther* 2014; **40**: 1023-1034 [PMID: 25199904 DOI: 10.1111/apt.12938]
8 **Ruscio M**. Is SIBO A Real Condition? *Altern Ther Health Med* 2019; **25**: 30-38 [PMID: 31550680]
9 **Gilson A**, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, Chartash D. How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Med Educ* 2023; **9**: e45312 [PMID: 36753318 DOI: 10.2196/45312]
10 **Ayers JW**, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, Faix DJ, Goodman AM, Longhurst CA, Hogarth M, Smith DM. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Intern Med* 2023; **183**: 589-596 [PMID: 37115527 DOI: 10.1001/jamainternmed.2023.1838]
11 **Samaan JS**, Yeo YH, Rajeev N, Hawley L, Abel S, Ng WH, Srinivasan N, Park J, Burch M, Watson R, Liran O, Samakar K. Assessing the Accuracy of Responses by the Language Model ChatGPT to Questions Regarding Bariatric Surgery. *Obes Surg* 2023; **33**: 1790-1796 [PMID: 37106269 DOI: 10.1007/s11695-023-06603-5]
12 **Yeo YH**, Samaan JS, Ng WH, Ting PS, Trivedi H, Vipani A, Ayoub W, Yang JD, Liran O, Spiegel B, Kuo A. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. *Clin Mol Hepatol* 2023; **29**: 721-732 [PMID: 36946005 DOI: 10.3350/cmh.2023.0089]
13 **Cima RR**, Anderson KJ, Larson DW, Dozois EJ, Hassan I, Sandborn WJ, Loftus EV, Pemberton JH. Internet use by patients in an inflammatory bowel disease specialty clinic. *Inflamm Bowel Dis* 2007; **13**: 1266-1270 [PMID: 17567877 DOI: 10.1002/ibd.20198]
14 **OpenAI**. GPT-4 Technical Report. 2023 [DOI: 10.48550/arXiv.2303.08774]
15 **King RC**, Samaan JS, Yeo YH, Mody B, Lombardo DM, Ghashghaei R. Appropriateness of ChatGPT in answering heart failure related questions. 2023. Available from: https://www.medrxiv.org/content/10.1101/2023.07.07.23292385v1
16 **Ayre J**, Mac O, McCaffery K, McKay BR, Liu M, Shi Y, Rezwan A, Dunn AG. New Frontiers in Health Literacy: Using ChatGPT to Simplify Health Information for People in the Community. *J Gen Intern Med* 2024; **39**: 573-577 [PMID: 37940756 DOI: 10.1007/s11606-023-08469-w]
17 **Alkaissi H**, McFarlane SI. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus* 2023; **15**: e35179 [PMID: 36811129 DOI: 10.7759/cureus.35179]
18 **Dave T**, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell* 2023; **6**: 1169595 [PMID: 37215063 DOI: 10.3389/frai.2023.1169595]
19 **Srinivasan N**, Samaan JS, Rajeev ND, Kanu MU, Yeo YH, Samakar K. Large language models and bariatric surgery patient education: a comparative readability analysis of GPT-3.5, GPT-4, Bard, and online institutional resources. *Surg Endosc* 2024 [PMID: 38472531 DOI: 10.1007/s00464-024-10720-2]
20 **Rouhi AD**, Ghanem YK, Yolchieva L, Saleh Z, Joshi H, Moccia MC, Suarez-Pierre A, Han JJ. Can Artificial Intelligence Improve the Readability of Patient Education Materials on Aortic Stenosis? A Pilot Study. *Cardiol Ther* 2024; **13**: 137-147 [PMID: 38194058 DOI: 10.1007/s40119-023-00347-0]

21 **Yeo YH**, Samaan JS, Ng WH, Ma X, Ting P, Kwak M, Panduro A, Lizaola-Mayo B, Trivedi H, Vipani A, Ayoub W, Yang JD, Liran O, Spiegel B, Kuo A. GPT-4 outperforms ChatGPT in answering non-English questions related to cirrhosis. 2023. Available from: https://www.medrxiv.org/content/10.1101/2023.05.04.23289482v1

22 **Samaan JS**, Yeo YH, Ng WH, Ting PS, Trivedi H, Vipani A, Yang JD, Liran O, Spiegel B, Kuo A, Ayoub WS. ChatGPT's ability to comprehend and answer cirrhosis related questions in Arabic. *Arab J Gastroenterol* 2023; **24**: 145-148 [PMID: 37673708 DOI: 10.1016/j.ajg.2023.08.001]

23 **Wang H**, Wu W, Dou Z, He L, Yang L. Performance and exploration of ChatGPT in medical examination, records and education in Chinese: Pave the way for medical AI. *Int J Med Inform* 2023; **177**: 105173 [PMID: 37549499 DOI: 10.1016/j.ijmedinf.2023.105173]

24 **Omiye JA**, Lester JC, Spichak S, Rotemberg V, Daneshjou R. Large language models propagate race-based medicine. *NPJ Digit Med* 2023; **6**: 195 [PMID: 37864012 DOI: 10.1038/s41746-023-00939-z]

25 **Kaplan DM**, Palitsky R, Arconada Alvarez SJ, Pozzo NS, Greenleaf MN, Atkinson CA, Lam WA. What's in a Name? Experimental Evidence of Gender Bias in Recommendation Letters Generated by ChatGPT. *J Med Internet Res* 2024; **26**: e51837 [PMID: 38441945 DOI: 10.2196/51837]

26 **Zandi R**, Fahey JD, Drakopoulos M, Bryan JM, Dong S, Bryar PJ, Bidwell AE, Bowen RC, Lavine JA, Mirza RG. Exploring Diagnostic Precision and Triage Proficiency: A Comparative Study of GPT-4 and Bard in Addressing Common Ophthalmic Complaints. *Bioengineering (Basel)* 2024; **11** [PMID: 38391606 DOI: 10.3390/bioengineering11020120]

27 **Chervonski E**, Harish KB, Rockman CB, Sadek M, Teter KA, Jacobowitz GR, Berland TL, Lohr J, Moore C, Maldonado TS. Generative artificial intelligence chatbots may provide appropriate informational responses to common vascular surgery questions by patients. *Vascular* 2024; 17085381241240550 [PMID: 38500300 DOI: 10.1177/17085381241240550]

28 **Rahsepar AA**, Tavakoli N, Kim GHJ, Hassani C, Abtin F, Bedayat A. How AI Responds to Common Lung Cancer Questions: ChatGPT vs Google Bard. *Radiology* 2023; **307**: e230922 [PMID: 37310252 DOI: 10.1148/radiol.230922]

29 **Shen Y**, Heacock L, Elias J, Hentel KD, Reig B, Shih G, Moy L. ChatGPT and Other Large Language Models Are Double-edged Swords. *Radiology* 2023; **307**: e230163 [PMID: 36700838 DOI: 10.1148/radiol.230163]

30 **Xiao Y**, Wang WY. On Hallucination and Predictive Uncertainty in Conditional Language Generation. In: Merlo P, Tiedemann J, Tsarfaty R, eds. Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. Association for Computational Linguistics; 2021: 2734-2744 [DOI: 10.18653/v1/2021.eacl-main.236]

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