**Name of Journal:** *Artificial Intelligence in Gastroenterology*

**Manuscript NO:** 65772

**Manuscript Type:** MINIREVIEWS

**Application of artificial intelligence in microbiome study promotes precision medicine for gastric cancer**

Li ZM *et al*. Application of AI in microbiome study in GC

Zhi-Ming Li, Xuan Zhuang

**Zhi-Ming Li,** Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, Hubei Province, China

**Zhi-Ming Li, Xuan Zhuang,** Department of Urology, The First Affiliated Hospital of Xiamen University, Xiamen 361003, Fujian Province, China

**Xuan Zhuang,** Department of Clinical Medicine, Fujian Medical University, Fuzhou 350122, Fujian Province, China

**Author contributions:** Li ZM conceptualized the paper; Li ZM and Zhuang X wrote the paper; all authors read and approved the final manuscript.

**Supported by** Health Commission of Hubei Province Scientific Research Project, No. WJ2021Q023.

**Corresponding author: Zhi-Ming Li, PhD, Academic Fellow,** Tongji Medical College, Huazhong University of Science and Technology, Baofeng Street, Qiaokou District, Wuhan 430030, Hubei Province, China. lzmleo@xmu.edu.cn

**Received:** March 16, 2021

**Revised:** April 22, 2021

**Accepted:** July 9, 2021

**Published online:**

**Abstract**

The microbiome has been identified as a causing factor for many cancers. *Helicobacter pylori* contributes to the development of gastric cancer (GC) and impacts disease treatments. The rapid development of sequencing technology is increasingly producing large-scale and complex big data. However, there are many obstacles in the analysis of these data by humans, which limit clinicians from making rapid decisions. Recently, the emergence of artificial intelligence (AI), including machine learning and deep learning, has greatly assisted clinicians in processing and interpreting large microbiome data. This paper reviews the application of AI in the study of the microbiome and discusses its potential in the diagnosis and therapy of GC. We also exemplify strategies for implementing microbiome-based precision medicines for patients with GC.

**Key Words:** Artificial intelligence; Sequencing; Microbiome; Gastric cancer

Li ZM, Zhuang X. Application of artificial intelligence in microbiome study promotes precision medicine for gastric cancer. *Artif Intell Gastroenterol* 2021; In press

**Core Tip:** Artificial intelligence (AI) helps us understand the role of the microbiome in gastric cancer (GC) and further promote the development precision medicine. AI can be applied in the following three aspects: (1) AI improves the diagnostic accuracy for GC based on big data and gastric microbiome; (2) AI aids pathologists to diagnose gastric biopsies rapidly by sensitively detecting low abundance microbes; and (3) AI regulates individual’s dietary intake by giving new insight into host-microbiome interactions.

**INTRODUCTION**

Gastric cancer (GC, also known as stomach cancer) is the second leading cause of cancer-related mortality globally, with over 70000 new cases diagnosed every year[1]. The 5-year survival rate of GC is lower than 15%, even in the United States[2]. According to Lauren's criteria, GC can be classified into two main types: Diffuse and intestinal. The diffuse type usually appears in younger patients and tends to be more aggressive, whereas the intestinal type is usually found in older patients and is caused by chronic infection with *Helicobacter pylori* (*H. pylori*)[3]. The microbiota in the stomach is extremely rich and complex[4]. DNA sequencing and computational methods are making astounding advances in the identification of conserved ribosomal RNA (rRNA) genes for pathogenic microorganisms. More than 100 phylotypes have been uncovered in humans, and the majority of gastric microbiota falls within five phyla, including *Bacteroidetes, Firmicutes, Proteobacteria, Actinobacteria*, and *Fusobacteria*. *H. pylori* belongs to *Proteobacteria*. *H. pylori* infection triggers multistep progression from chronic gastritis, atrophic gastritis, and intestinal metaplasia to carcinoma finally. However, the issue of how the gastric microbiota interplays with *H. pylori* (namely, does the gastric microbiota lead to a more virulent *H. pylori* or, *vice versa*, does *H. pylori* facilitate the carcinogenesis of the microbiota?) is still not clear. This might have implications for clinical management.

Artificial intelligence (AI) is the simulation of human intelligence processes by computers and has been applied in various fields, such as image processing and natural language processing. AI is playing an increasingly important role in healthcare. It has been demonstrated that AI algorithms can support humans in simplifying the multidimensional, complex metagenomic data of gene profiling and elucidating the peculiar signatures of beneficial microbes in the gastrointestinal tract[5]. As a core branch of AI, machine learning (ML) focuses on building mathematical models that help machines make predictions or decisions without being explicitly programmed. In the field of ML, deep learning (DL) has become the dominant approach for ongoing work with big data. DL, a subset of ML, is inspired by the information processing system discovered in the human brain. DL uses numerous layers of algorithms (artificial neural networks) to extract higher-level features from raw input. Briefly, ML is a core branch of AI, and DL is performed to implement ML. ML and DL have been successfully used to predict the risk of GC[6].

**AI makes accurate predictions with big data and the gastric microbiome**

Gastroenterology is a ﬁeld where AI can make a signiﬁcant difference. Traditional diagnostic methods have insufficient resolution ability to estimate the invasion depth of early GC in the clinic. Thus, over one-third of advanced GC cases with lesions around the cardia are not easily detected by image-based methods[7]. However, AI-assisted image analysis using endoscopic detection can make more accurate assessments and provide more details than conventional analysis[8]. There are still two main limitations in AI-assisted image analysis. First, there are relatively few data serving as learning and testing materials for building DL models. Second, the diagnostic accuracy is greatly affected when low-resolution images, which endoscopists usually encounter in clinical practice, are input. The above two points may cause certain defects in medical decisions based on image analysis. Remarkably, the combination of AI and the microbiome shows great potential in precision medicine for GC.

High-throughput sequencing is becoming a common technology for typing microbial isolates, especially in clinical samples. Many gene mutations, transcriptional differences, translational differences, epigenetic variations, and metabolic changes have been identified as being associated with the heterogeneity and stage of GC. High-throughput sequencing generates massive microbial data. A deep understanding of microbial data is helpful to explain the relationship between microbes and diseases[9]. Virulence among *H. pylori* strains and host genetic polymorphisms contribute to GC susceptibility*.* AI algorithms effectively improve our understanding of the gastric microbiota due to two major advantages. First, AI methods can be applied to extract microbial genomic DNA from sequencing samples. Second, AI methods can simultaneously examine all genes in all organisms contained in a sample. Combined with other parameters, such as food habits, duration of infection, and physical activity, AI algorithms can provide better health advice to GC patients. A recent study has started to explore the ability of DL to treat diseases related to gut dysbiosis based on the individual’s microbiome pattern[10]. In the future, researchers can develop AI algorithms to regulate the individual’s dietary intake and plan their meals when we fully understand the microbiome differences between people with and without disease (Figure 1).

**AI identifies low abundance microbes using sequencing data**

Studying the microbiome composition of primary samples provides a chance to understand the role of pathogenic microorganisms in disease development. In the late 2000s, two large-scale international human microbiome projects (HMPs), Metagenomics of the Human Intestinal Tract[11] and the HMP[12], were initiated to study microorganisms in the human body and to develop computational methods that analyze sequenced metagenomes. However, it seems challenging due to the low number of microbial DNA relative to the host DNA. Accurate identification of the microbiome requires the removal of all possible sequencing reads that originate from human DNA. Bacterial identification was commonly completed by characterization of uniform genomic coverage[13]. For example, the sequence identity of 16S rRNA gene fragments greater than 97% can be classified into separate operational taxonomic units (OTUs), which means the phylogenetic boundaries of different bacterial species[14]. Bacterial identification can also be completed based on coverage along a narrow region of their genomes. For example, analysis of amplicon sequence variants improves the sensitivity and specificity and decreases the problem of inflated microbiota datasets due to falsely identified OTUs originating from misclustered sequences[15]. Recently, Lupolova *et al*[16] found that ML algorithms made a good attribution of the host sources of *S. enterica* serovar Typhimurium isolates[16]. The combination of 16S rRNA gene sequencing data and AI algorithms may reveal the essential role of low-abundance bacteria in the alteration of the gut microbiota composition.

It is challenging to quantify and characterize microbiome profiling in samples where the bacterial content is relatively low. The microbial community in the stomach is typically restricted by the lower luminal pH, which selects for acid-resistant bacterial populations and usually limits the colonization densities to < 1000 colony-forming units per gram (CFU/g)[17]. The current approach for detecting the bacteria of fecal or environmental samples cannot be directly used to analyze the microbiome from the upper gastrointestinal tract, such as the stomach. This is partly because the high amount of human DNA in the samples confounds microbial identification. Klein *et al*[18] designed a DL algorithm that can be used to detect *H. pylori* on regular whole slide images of gastric biopsies, achieving a sensitivity of 100%[18]. Detecting the low abundance bacteria without sample processing facilitates the establishment of a rapid diagnostic method. Recently, we designed magnetic nanoparticles with a broad range of capture potentials *via* electrostatic attractions[19]. This system can rapidly and efficiently capture bacteria at a low concentration of 10 CFU/mL within 1 h. The capture efficiency was more than 90%. It can be used to evaluate the microbiome profile of gastric biopsies in future studies.

**AI uncovers host-microbiome interactions**

A comparative study of GC and chronic gastritis using an approach targeting the 16S rRNA gene of mucosal biopsies showed that bacterial diversity was decreased in GC patients[20]. Patients with GC had a large number of non-*Helicobacter* Proteobacteria. Colonization with bacteria other than *H. pylori* breaks the balance between the resident gastric microbiota and the host, which may increase the risk for *H. pylori*-related cancer. Another study evaluated the microbiota composition in normal, peritumoral, and tumoral tissues by 16S rRNA gene profiling and found that microbial diversity was significantly reduced in peritumoral and tumoral microhabitats[21]. *H. pylori*, *Prevotella copri*, and *Bacteroides uniformis* were relatively less abundant in the tumoral microhabitat, whereas *Prevotella melaninogenica*, *Streptococcus anginosus*, and *Propionibacterium acnes* were more abundant. The authors proposed the hypothesis that chronic atrophic gastritis with atrophy (the acidity of the microenvironment of the stomach is reduced) was attributed to *H. pylori* substitution by a cancer-prone microbiota[22]. Additionally, the same research team found a close relationship between the subtype of immune cells (regulatory T cells and plasmacytoid dendritic cells) and gastric microbiota dysbiosis within the tumor microenvironment. It is already known that *H. pylori* infection functions in the development of precancerous lesions, such as chronic gastritis. Nevertheless, the dramatic changes in the composition of the stomach microbiome play a more direct role in the later stages of cancer. Moreover, the microbiome affects the therapeutic response of GC patients, and the treatment also impacts microbial composition. Distal gastrectomy impacts postoperative gut microbiota composition, leading to higher abundances of *Escherichia, Shigella*, *Veillonella*, and *Clostridium XVIII* and a lower abundance of *Bacteroides*[23]. Immune checkpoint inhibitors targeting programmed cell death 1 (PD-1)/programmed cell death ligand 1 were recently added to the therapeutic arsenal for GC. The microbiome composition interferes with the response to these inhibitors. A recent study reported that nonresponders to PD-1 blockade immunotherapy can be distinguished from responders according to the ratio of putatively favorable to unfavorable bacteria[24]. Thus, the role of the microbiome in cancer-immune interactions is gaining much attention. When we learn more about host-microbiome interactions, nonresponders to checkpoint inhibitors are easier to select and treat by personalized immunotherapy.

Due to the practical limitations of analysis methods, there are still large gaps on how the microbiome mechanically affects host function at the system and community levels. Notably, the past few decades has seen significant work on AI in filling these existing gaps. AI algorithms can co-analyze heterogeneous datasets and capture changes at the microbial and host levels. These methods can be classified into four types: Interfering protein-protein interactions, interfering RNA-mediated interactions, interfering microbe-host metabolic networks, and integrating multiple interspecies and intraspecies networks and omic datasets[25]. The powerful multiomics tools and rapidly developed AI algorithms can greatly enhance or perhaps revolutionize microbiome research. This collaboration provides hopeful expectations to improve our current understanding of GC mechanisms, as well as better detection and treatment.

**CONCLUSION**

We live in a world surrounded by data and microbes. The gastric microbiome occupies an important position in maintaining the individual’s health. A large quantity of complex sequencing data are generated by high-throughput technologies. However, inherent challenges still exist in data processing, including confounding variables from abundant organisms, the integration of different omics data, and the relationships between microbes and their hosts. Currently, big data are easier than ever to analyze due to the assistance of AI technologies. AI is evolving as an important tool for the proposal of new biological hypotheses and the discovery of biomarkers from the available data. In the future, the renewal of the stomach of dysbiosis patients may be achieved by synthetic biology and food engineering based on our understanding of the microbiome and the performance of AI.

**REFERENCES**

1 **Brawner KM**, Morrow CD, Smith PD. Gastric microbiome and gastric cancer. *Cancer J* 2014; **20**: 211-216 [PMID: 24855010 DOI: 10.1097/PPO.0000000000000043]

2 **Polk DB**, Peek RM Jr. Helicobacter pylori: gastric cancer and beyond. *Nat Rev Cancer* 2010; **10**: 403-414 [PMID: 20495574 DOI: 10.1038/nrc2857]

3 **Quadri HS**, Smaglo BG, Morales SJ, Phillips AC, Martin AD, Chalhoub WM, Haddad NG, Unger KR, Levy AD, Al-Refaie WB. Gastric Adenocarcinoma: A Multimodal Approach. *Front Surg* 2017; **4**: 42 [PMID: 28824918 DOI: 10.3389/fsurg.2017.00042]

4 **Knight ZA**, Tan K, Birsoy K, Schmidt S, Garrison JL, Wysocki RW, Emiliano A, Ekstrand MI, Friedman JM. Molecular profiling of activated neurons by phosphorylated ribosome capture. *Cell* 2012; **151**: 1126-1137 [PMID: 23178128 DOI: 10.1016/j.cell.2012.10.039]

5 **Qu K**, Guo F, Liu X, Lin Y, Zou Q. Application of Machine Learning in Microbiology. *Front Microbiol* 2019; **10**: 827 [PMID: 31057526 DOI: 10.3389/fmicb.2019.00827]

6 **Taninaga J**, Nishiyama Y, Fujibayashi K, Gunji T, Sasabe N, Iijima K, Naito T. Prediction of future gastric cancer risk using a machine learning algorithm and comprehensive medical check-up data: A case-control study. *Sci Rep* 2019; **9**: 12384 [PMID: 31455831 DOI: 10.1038/s41598-019-48769-y]

7 **Sung IK**, Kim YC, Yun JW, Seo HI, Park DI, Cho YK, Kim HJ, Park JH, Sohn CI, Jeon WK, Kim BI, Oh SJ, Son BH, Yoo CH, Sohn JH, Lee HY, Won KH. Characteristics of advanced gastric cancer undetected on gastroscopy. *Korean J Gastroenterol* 2011; **57**: 288-293 [PMID: 21623137 DOI: 10.4166/kjg.2011.57.5.288]

8 **Mori Y**, Kudo SE, Misawa M, Saito Y, Ikematsu H, Hotta K, Ohtsuka K, Urushibara F, Kataoka S, Ogawa Y, Maeda Y, Takeda K, Nakamura H, Ichimasa K, Kudo T, Hayashi T, Wakamura K, Ishida F, Inoue H, Itoh H, Oda M, Mori K. Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study. *Ann Intern Med* 2018; **169**: 357-366 [PMID: 30105375 DOI: 10.7326/M18-0249]

9 **Shi XJ**, Wei Y, Ji B. Systems Biology of Gastric Cancer: Perspectives on the Omics-Based Diagnosis and Treatment. *Front Mol Biosci* 2020; **7**: 203 [PMID: 33005629 DOI: 10.3389/fmolb.2020.00203]

10 **Zhou YH**, Gallins P. A Review and Tutorial of Machine Learning Methods for Microbiome Host Trait Prediction. *Front Genet* 2019; **10**: 579 [PMID: 31293616 DOI: 10.3389/fgene.2019.00579]

11 **Qin J**, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Doré J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J; MetaHIT Consortium, Bork P, Ehrlich SD, Wang J. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 2010; **464**: 59-65 [PMID: 20203603 DOI: 10.1038/nature08821]

12 **Kim Y**, Koh I, Rho M. Deciphering the human microbiome using next-generation sequencing data and bioinformatics approaches. *Methods* 2015; **79-80**: 52-59 [PMID: 25448477 DOI: 10.1016/j.ymeth.2014.10.022]

13 **Zhang C**, Cleveland K, Schnoll-Sussman F, McClure B, Bigg M, Thakkar P, Schultz N, Shah MA, Betel D. Identification of low abundance microbiome in clinical samples using whole genome sequencing. *Genome Biol* 2015; **16**: 265 [PMID: 26614063 DOI: 10.1186/s13059-015-0821-z]

14 **Fricker AM**, Podlesny D, Fricke WF. What is new and relevant for sequencing-based microbiome research? A mini-review. *J Adv Res* 2019; **19**: 105-112 [PMID: 31341676 DOI: 10.1016/j.jare.2019.03.006]

15 **Kopylova E**, Navas-Molina JA, Mercier C, Xu ZZ, Mahé F, He Y, Zhou HW, Rognes T, Caporaso JG, Knight R. Open-Source Sequence Clustering Methods Improve the State Of the Art. *mSystems* 2016; **1** [PMID: 27822515 DOI: 10.1128/mSystems.00003-15]

16 **Lupolova N**, Lycett SJ, Gally DL. A guide to machine learning for bacterial host attribution using genome sequence data. *Microb Genom* 2019; **5** [PMID: 31778355 DOI: 10.1099/mgen.0.000317]

17 **Sheh A**, Fox JG. The role of the gastrointestinal microbiome in Helicobacter pylori pathogenesis. *Gut Microbes* 2013; **4**: 505-531 [PMID: 23962822 DOI: 10.4161/gmic.26205]

18 **Klein S**, Gildenblat J, Ihle MA, Merkelbach-Bruse S, Noh KW, Peifer M, Quaas A, Büttner R. Deep learning for sensitive detection of Helicobacter Pylori in gastric biopsies. *BMC Gastroenterol* 2020; **20**: 417 [PMID: 33308189 DOI: 10.1186/s12876-020-01494-7]

19 **Li Z**, Ma J, Ruan J, Zhuang X. Using Positively Charged Magnetic Nanoparticles to Capture Bacteria at Ultralow Concentration. *Nanoscale Res Lett* 2019; **14**: 195 [PMID: 31165285 DOI: 10.1186/s11671-019-3005-z]

20 **Companioni O**, Bonet C, Muñoz X, Weiderpass E, Panico S, Tumino R, Palli D, Agnoli C, Vineis P, Boutron-Ruault MC, Racine A, Clavel-Chapelon F, Travis RC, Khaw KT, Riboli E, Murphy N, Vergnaud AC, Trichopoulou A, Benetou V, Trichopoulos D, Lund E, Johansen D, Lindkvist B, Johansson M, Sund M, Ardanaz E, Sánchez-Cantalejo E, Huerta JM, Dorronsoro M, Ramón Quirós J, Tjonneland A, Mortensen LM, Overvad K, Chang-Claude J, Rizzato C, Boeing H, Bueno-de-Mesquita HB, Siersema P, Peeters PH, Numans ME, Carneiro F, Licaj I, Freisling H, Sala N, González CA. Polymorphisms of Helicobacter pylori signaling pathway genes and gastric cancer risk in the European Prospective Investigation into Cancer-Eurgast cohort. *Int J Cancer* 2014; **134**: 92-101 [PMID: 23824692 DOI: 10.1002/ijc.28357]

21 **Malfertheiner P**, Sipponen P, Naumann M, Moayyedi P, Mégraud F, Xiao SD, Sugano K, Nyrén O; Lejondal H. pylori-Gastric Cancer Task Force. Helicobacter pylori eradication has the potential to prevent gastric cancer: a state-of-the-art critique. *Am J Gastroenterol* 2005; **100**: 2100-2115 [PMID: 16128957 DOI: 10.1111/j.1572-0241.2005.41688.x]

22 **Schulz C**, Schütte K, Mayerle J, Malfertheiner P. The role of the gastric bacterial microbiome in gastric cancer: *Helicobacter pylori* and beyond. *Therap Adv Gastroenterol* 2019; **12**: 1756284819894062 [PMID: 31897087 DOI: 10.1177/1756284819894062]

23 **Liang W**, Yang Y, Wang H, Wang H, Yu X, Lu Y, Shen S, Teng L. Gut microbiota shifts in patients with gastric cancer in perioperative period. *Medicine (Baltimore)* 2019; **98**: e16626 [PMID: 31464899 DOI: 10.1097/MD.0000000000016626]

24 **Matson V**, Fessler J, Bao R, Chongsuwat T, Zha Y, Alegre ML, Luke JJ, Gajewski TF. The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science* 2018; **359**: 104-108 [PMID: 29302014 DOI: 10.1126/science.aao3290]

25 **Sudhakar P,** Machiels K, Verstockt B, Korcsmaros T, Vermeire S. Computational Biology and Machine Learning Approaches to Understand Mechanistic Microbiome-Host Interactions. *Front Microbiol* 2021; **12:** 618856 [PMID: 34046017 DOI: 10.3389/fmicb.2021.618856]

**Footnotes**

**Conflict-of-interest statement:** The authors declare that they have no competing interests to disclose.

**Open-Access:** This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/Licenses/by-nc/4.0/

**Manuscript source:** Invited manuscript

**Peer-review started:** March 16, 2021

**First decision:** April 15, 2021

**Article in press:**

**Specialty type:** Gastroenterology and hepatology

**Country/Territory of origin:** China

**Peer-review report’s scientific quality classification**

Grade A (Excellent): 0

Grade B (Very good): B

Grade C (Good): C, C, C

Grade D (Fair): 0

Grade E (Poor): 0

**P-Reviewer:** Abreu de Melo MI, Kinami S, Moradi L, Sharma J **S-Editor:** Fan JR **L-Editor:** Wang TQ **P-Editor:**

**Figure Legends**



**Figure 1 Introducing artificial intelligence and microbiome study to precision medicine for gastric cancer**. The sequencing profiles of individual patient microbiomes are analyzed by artificial intelligence (AI), which helps patients to be classified into sub-groups. At the molecular level, AI reveals the molecular mechanisms of microbe-host interactions. At the individual level, AI allows gastric cancer patients to be treated with effective drugs, such as supplementing commensal bacteria, engineered bacteria, and microbiome-targeted drugs.