ISSN 1881-7815 Online ISSN 1881-7823

BST BioScience Trends

Volume 19, Number 2
April 2025

IRCA BSSA

IACMHR

www.biosciencetrends.com

BST
BioScience Trends

ISSN: 1881-7815
Online ISSN: 1881-7823
CODEN: BTIRCZ
Issues/Year: 6
Language: English
Publisher: IACMHR Co., Ltd.

BioScience Trends is one of a series of peer-reviewed journals of the International Research and Cooperation Association for Bio & Socio-Sciences Advancement (IRCA-BSSA) Group. It is published bimonthly by the International Advancement Center for Medicine & Health Research Co., Ltd. (IACMHR Co., Ltd.) and supported by the IRCA-BSSA.

BioScience Trends devotes to publishing the latest and most exciting advances in scientific research. Articles cover fields of life science such as biochemistry, molecular biology, clinical research, public health, medical care system, and social science in order to encourage cooperation and exchange among scientists and clinical researchers.

BioScience Trends publishes Original Articles, Brief Reports, Reviews, Policy Forum articles, Communications, Editorials, News, and Letters on all aspects of the field of life science. All contributions should seek to promote international collaboration.

Editorial Board

Editorial and Head Office

Pearl City Koishikawa 603,
2-4-5 Kasuga, Bunkyo-ku, Tokyo 112-0003, Japan
E-mail: office@biosciencetrends.com

BioScience Trends

Editorial and Head Office

Pearl City Koishikawa 603, 2-4-5 Kasuga, Bunkyo-ku,
Tokyo 112-0003, Japan

E-mail: office@biosciencetrends.com
URL: www.biosciencetrends.com

Editorial Board Members

Editorial

Review

Original Article

Editorial

DOI: 10.5582/bst.2024.01382

Combating syphilis resurgence: China's multifaceted approach

Rongfeng Zhou[1], Kai Sun[2], Ting Li[1], Hongzhou Lu[3,4,*]

[1] Hospital administrative office, National Clinical Research Center for Infectious Diseases, the Third People's Hospital of Shenzhen, Shenzhen, Guangdong, China;
[2] Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China;
[3] Shenzhen Clinical Research Center for Infectious Disease, State Key Discipline of Infectious Disease, Shenzhen Third People's Hospital, Second Hospital Affiliated to Southern University of Science and Technology, Shenzhen, Guangdong, China;
[4] National Clinical Research Center for Infectious Disease, Shenzhen, Guangdong, China.

SUMMARY: Syphilis, a chronic infection caused by Treponema pallidum, is experiencing a global resurgence, posing significant public health challenges. This study examined the escalating trends of syphilis in the United States, China, and some other countries highlighting the impact of the COVID-19 pandemic, changes in sexual behavior, coinfection with the other infectious diseases such as AIDs, and the role of public health funding. The analysis revealed a stark increase in syphilis cases, particularly among high-risk groups such as men who have sex with men (MSM). China's National Syphilis Control Program (NSCP), initiated in 2010, is a comprehensive approach to syphilis management that incorporates health education, access to testing and treatment, partner notification, safe sex promotion, community interventions, and stigma reduction. The success of the NSCP in reducing early syphilis incidence rates and congenital syphilis in Guangdong Province, that may be served as a model for international syphilis control efforts. This study highlights the necessity for targeted public health interventions and the importance of robust healthcare infrastructure in combating the syphilis epidemic.

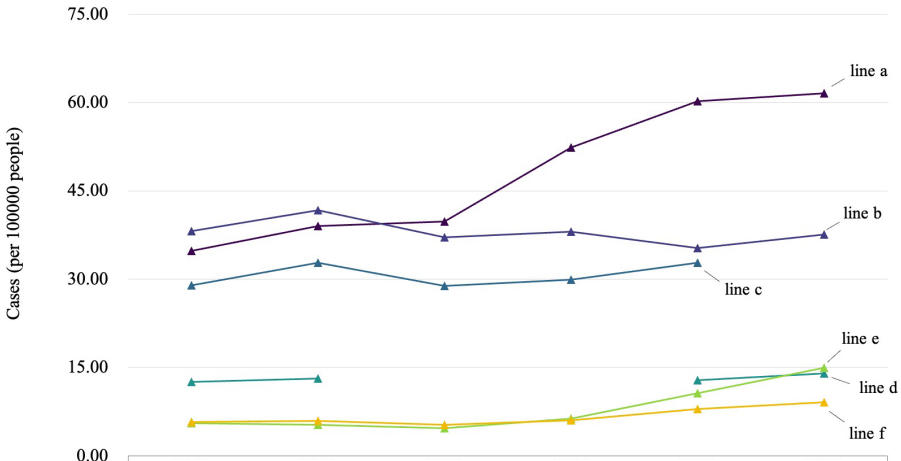
Keywords: syphilis, global resurgence, social determinations of health (SDoH), the National Syphilis Control Program (NSCP), China's syphilis management

1. Syphilis sweeping the world

Syphilis, a chronic systemic sexually transmitted infections(STIs) caused by Treponema pallidum, is on the rise globally and poses a major public health concern. It affects multiple organs and systems. Recently, the syphilis dilemma in countries such as the United States, Japan, and the European Union has been reported exhibiting a escalating trend. The United States reported 210,000 cases in 2023 (*1*), and Japan is facing a "once in fifty years" syphilis epidemic, which mainly involves heterosexual men, Unexpectedly, the number of cases of primary and secondary syphilis has increased significantly in young women. Japan's Ministry of Health has regarded the disease as a public health threat and has strengthened its preventive measures for young women (*2*). The state quo of the syphilis situation in China is not optimistic. Still, thanks to comprehensive syphilis control measures, such as a syphilis recording system and management strategies that combine traditional Chinese medicine with modern medicine, the National Syphilis Control Program (NSCP), launched in 2010, has played a crucial role. As a result, by 2020, the incidence rate of early syphilis has dropped significantly from 21.1 cases per 100,000 people to 8.8 cases. Guangdong Province, a province located at southern China, where syphilis was first recorded in China (contained in the Compendium of Materia Medica), has taken effective measures to reduce congenital syphilis. The reported sexual syphilis rate has dropped from 128.55 to 5.76 cases per 100,000 live births (*3*).

2. A scary upward trend

The global incidence of syphilis is on the rise, presenting a significant public health challenge. Figure 1 provides a detailed view of the syphilis prevalence, measured as cases per 10,000 individuals, across several countries from 2018 to 2023. The United States has the highest rates among the listed countries, with a substantial increase from 34.82 per 10,000 in 2018 to 61.55 in 2023, highlighting a severe and growing syphilis epidemic (*4*). China has seen a general increase in syphilis prevalence, starting at 38.19 per 10,000 in 2018, reaching a high of 41.73 in 2019, and then stabilizing around 37.61 in 2023 (*5*). Australia's syphilis rates have fluctuated,



	2018	2019	2020	2021	2022	2023
United States (line a)	34.82	38.98	39.75	52.36	60.16	61.55
China (line b)	38.19	41.73	37.06	38.07	35.27	37.61
Australia (line c)	28.93	32.81	28.84	29.94	32.74	
United Kingdom (line d)	12.56	13.11			12.86	13.96
Japan (line e)	5.53	5.25	4.65	6.35	10.61	14.96
European Union (line f)	5.77	5.95	5.26	6.06	7.92	9.08

Data sources:
Line a is based on the CDC's National Overview of STIs in 2023 (*4*).
Line b is attributed to the China Center for Disease Control and Prevention's national epidemic situation reports (*5*).
Line c is based on the surveillance data from the Australian government's health department and relevant health research publications (*6*).
Line d is sourced from the UK Health Security Agency's report "Tracking the syphilis epidemic in England: 2013 to 2023." (*7*)
Line e is based on a descriptive study published in Plos One(*6*).
Line f is sourced from the European Centre for Disease Prevention and Control's (ECDC) Surveillance Atlas of Infectious Diseases (*8*).

Figure 1. Progressive syphilis trends across nations: A comparative analysis.

beginning at 28.93 per 10,000 in 2018, peaking at 32.81 in 2019, and then slightly decreasing to 32.74 in 2023 (*6*). The United Kingdom's data indicates a moderate rise in syphilis prevalence, from 12.56 per 10,000 in 2018 to 13.96 in 2023 (*7*). Japan has experienced a notable increase, with syphilis rates growing from 5.53 per 10,000 in 2018 to 14.96 in 2023, which is approximately a 2.7-fold increase, indicating a significant upward trend (*6*). The European Union has shown a steady rise in syphilis rates, from 5.77 per 10,000 in 2018 to 9.08 in 2023 (*8*).

3. The resurgence of syphilis from the perspective of social determinations of health (SDoH)

A variety of factors cause the resurgence of syphilis, which is more obvious in high-risk groups such as MSM (*9*). The factors affecting the population can be summarized according to the social determinants of health (SDoH) model as follows:

1). *Changes in health behaviors*: As online interaction methods have become more popular, the incidence of casual sexual encounters facilitated by digital platforms has surged. This trend may contribute to enhancing syphilis transmission rates, consequently broadening the spread of the disease and increasing the number of vulnerable individuals (*10*). The adoption of pre-exposure prophylaxis (PrEP) has been associated with an uptick in cases of unprotected anal intercourse, consequently fueling a surge in syphilis diagnoses (*11*). This enhancement poses heightened risk of exposure and infection, particularly in several at-risk communities.

2). *Improved health awareness and diagnostic techniques*: It will lead to an increase in the self-detection rate of more suspected infected people and is more likely to lead to an increase in the reporting and identification of confirmed syphilis cases (*12*).

3). *The comprehensive impacts of COVID-19 pandemic*: On the one hand, during lockdown periods, diminished social interactions and sexual activities among at-risk populations might curtail syphilis proliferation, causing a downturn. However, on the other hand, the pandemic prompted a reallocation of healthcare resources, diminishing syphilis detection and patients' access to diagnostic and therapeutic services. This reduction in early case identification and intervention opportunities might fuel syphilis transmission. Conversely, congenital syphilis cases experienced an increase throughout the pandemic, with an upswing in vertical transmission. Post-pandemic, heightened economic and social strain may elevate population stress and alter relational dynamics, potentially inciting more hazardous sexual practices (*13*).

4). *Diminution in Public Health Financing*: Recently, the predominant allocation of public health budgets to combat the COVID-19 crisis may have sidetracked the populace's focus on STIs prevention and management. This reorientation of priorities could have resulted in a surge in syphilis cases (*14*).

5). *Historical and modern environmental impacts*: Historically, individuals with syphilis have been stigmatized as immoral, with artistic depictions from the

renaissance through the 18th century mirroring societal alarm towards the disease, categorizing it as a 'social affliction' (*15*). In the contemporary era, the rise of the adult entertainment sector has exposed performers to significant health jeopardy (*16*). The reluctance to consistently employ condoms has fueled the spread of STIs, and the proliferation of pornography has reshaped sexual norms, indirectly contributing to syphilis dissemination (*17*).

6). *Interplay with Other Ailments*: The coinfection of syphilis with HIV is a significant consequence, given that those living with HIV are at an elevated risk for syphilis. Additionally, an increase in syphilis has been observed among premenopausal women and in cases of congenital syphilis (*2*).

4. Lessons from China's Syphilis Management

Observing the data presented in Figure 1, it is notable that despite its substantial population, China has managed to sustain a declining trajectory in syphilis cases. The National Syphilis Control Program (NSCP), initiated in 2010, aims to manage syphilis via various levels of the Social-Ecological Model (SEM) (*18*) (Figure 2). Thus, prevention of syphilis can be divided into several levels:

4.1. Individual Level

1). *Health Education and Awareness*: NSCP encompasses extensive public outreach efforts aimed at enlightening the citizenry on syphilis, detailing its modes of transmission, and promoting preventive strategies.

2). *Access to Testing and Treatment*: Upon the roles of NSCP, there is a commitment to provide the populace with affordable access to syphilis screening and therapeutic interventions, pivotal for timely identification and medical care.

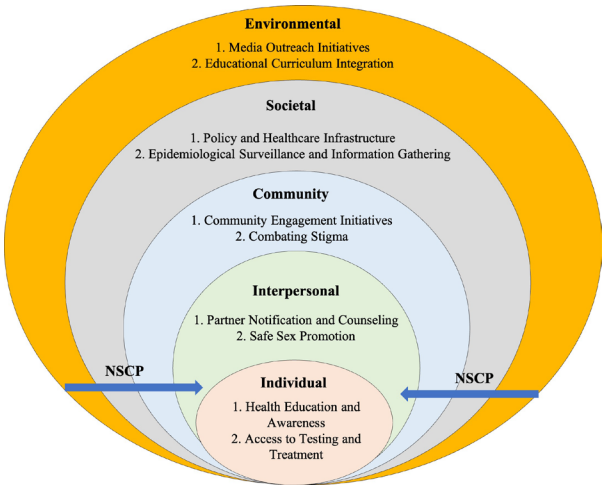


Figure 2. Syphilis control hierarchical: Integrating NSCP through social-ecological model (SEM).

4.2. Interpersonal Level

1). *Partner Notification and Counseling*: Within the framework of NSCP, proactive contact tracing is implemented to reach out to and offer screening and medical intervention to the intimate partners of those identified with syphilis, ensuring a comprehensive approach to the disease management.

2). *Safe Sex Promotion*: The NSCP actively encourages the use of condoms and advocates for safe sexual behaviors as a preventive measure.

4.3. Community Level

1). *Community Engagement Initiatives*: The NSCP bolsters grassroots efforts, including peer-led educational campaigns and community-driven health seminars, to involve local populations in the collective fight against syphilis.

2). *Combating Stigma*: As integral part of NSCP, they focus on mitigating the social stigma surrounding syphilis. This is aimed at creating an environment in which individuals feel empowered to access testing and treatment services without the dread of being subjected to prejudice or discrimination.

4.4. Societal Level

1). *Policy and Healthcare Infrastructure*: The NSCP bolsters management of syphilis through a robust policy framework, including mandatory case reporting and standardized STIs protocols. It also enhances the national healthcare system by expanding access to comprehensive STIs services, training medical staff, and upgrading diagnostic facilities.

2). *Epidemiological Surveillance and Information Gathering*: Central to NSCP is a sophisticated system for monitoring and data collection that is instrumental in tracking the spread of syphilis. This system is crucial for appraising the efficacy of control strategies and for guiding decisions regarding policy formulation and resource allocation.

4.5. Environmental Level

1). *Media Outreach Initiatives*: NSCP employs mass media channels to disseminate critical information on syphilis prevention and the significance of regular testing to a wide audience.

2). *Educational Curriculum Integration*: NSCP conducts sexual health education within the school system's curricula, aiming to instill a solid understanding of sexual health in the younger demographic and to cultivate a positive disposition toward it from an early age.

The resurgence of syphilis, often accompanied by HIV coinfection, poses a significant challenge to public

health. This intersection complicates treatment and increases transmission risks, highlighting the need for a comprehensive approach to syphilis management, including HIV coinfection (*19*). The NSCP addresses these issues, and the government has further efforts with the "Action Plan for Eliminating Mother-to-Child Transmission of HIV, Syphilis, and Hepatitis B (2022-2025) (*20*)", demonstrating a strategic commitment to tackling these intertwined epidemics.

5. Conclusions

The global syphilis resurgence highlights a critical public health challenge, with notable increases in countries such as the US, and Japan. China's National Syphilis Control Program (NSCP) has effectively mitigated this trend. NSCP serves as a model for addressing syphilis, emphasizing the importance of a multifaceted approach to controlling and preventing syphilis.

Funding: This work was supported by the Shenzhen Third People's Hospital project (No. 21250G1001 and No. 22240G1005).

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Harris E. "Alarming Trend" Persists as Syphilis Cases Swelled in the US in 2022. JAMA. 2024; 331:725.
2. Schmidt R, Carson PJ, Jansen RJ. Resurgence of Syphilis in the United States: An Assessment of Contributing Factors. Infect Dis (Auckl). 2019; 12:1178633719883282.
3. Tang X, Chen W, Tang SQ, Zhao PZ, Ling L, Wang C. The evaluation of preventive and control measures on congenital syphilis in Guangdong Province, China: A time series modeling study. Infection. 2022; 50:1179-1190.
4. CDC- STI Statistics. National Overview of STIs in 2023. *https://www.cdc.gov/sti-statistics/annual/summary.html* (accessd November 20 2024).
5. China Center for Disease Conrtol and Prevention. National epidemic situation of notifiable infectious diseases. 2024. *https://www.chinacdc.cn/jksj/jksj01/202410/t20241010_301346.html* (accessd November 20 2024).
6. Komori A, Mori H, Xie W, Valenti S, Naito T. Rapid resurgence of syphilis in Japan after the COVID-19 pandemic: A descriptive study. Plos One. 2024; 19:e0298288.
7. UK Health Security Agency. Tracking the syphilis epidemic in England: 2013 to 2023. *https://www.gov.uk/government/publications/tracking-the-syphilis-epidemic-in-england/tracking-the-syphilis-epidemic-in-england-2013-to-2023* (accessd November 20 2024).
8. ECDC. Surveillance Atlas of Infectious Diseases. 2024. *https://atlas.ecdc.europa.eu/public/index.aspx* (accessd November 20 2024).
9. Zheng Y, Ye K, Ying M, He Y, Yu Q, Lan L, Xu W. Syphilis epidemic among men who have sex with men: A global systematic review and meta-analysis of prevalence, incidence, and associated factors. J Glob Health. 2024; 14:04004.
10. Tripathy DM, Gupta S, Vasudevan B. Resurgence of syphilis, the great imitator. Med J Armed Forces India. 2022; 78:131-135.
11. Mendez-Lopez A, Stuckler D, Marcus U, Hickson F, Noori T, Whittaker RN, Jansen K, Diaz A, Henszel L, Velter A, Semenza JC, Schmidt AJ. Social and behavioural determinants of syphilis: Modelling based on repeated cross-sectional surveys from 2010 and 2017 among 278,256 men who have sex with men in 31 European countries. Lancet Reg Health Eur. 2022; 22:100483.
12. Ramchandani MS, Cannon CA, Marra CM. Syphilis: A Modern Resurgence. Infect Dis Clin North Am. 2023; 37:195-222.
13. Liu M, Zhou J, Lan Y, Zhang H, Wu M, Zhang X, Leng L, Mi X, Li J. A Neglected Narrative in the COVID-19 Pandemic: Epidemiological and Clinical Impacts of the COVID-19 Outbreak on Syphilis. Clin Cosmet Investig Dermatol. 2023; 16:2485-2496.
14. Sampath S, Renzu M, Clark P, Kelly J, Disandro D. Sexually transmitted disease rates are surging: A matter of life and death. In: Contemporary Issues in Clinical Bioethics-Medical, Ethical and Legal Perspectives *https://ouci.dntb.gov.ua/en/works/4LR2VYW9/* (accessd November 20 2024).
15. Morton RS. Syphilis in art: An entertainment in four parts. Part 1. Genitourin Med. 1990; 66:33-40.
16. Kluger N. Les infections sexuellement transmissibles et autres risques dans l'industrie du film adulte. In: Annales de Dermatologie et de Vénéréologie (Elsevier, 2014; pp. 122-129).
17. Ficarra G, Carlos R. Syphilis: the renaissance of an old disease with oral implications. Head Neck Pathol. 2009; 3:195-206.
18. The National Health Commission of the People's Republic of China. Syphilis Prevention and Control Plan in China (2010-2020). *https://www.gov.cn/gzdt/2010-06/21/content_1632301.htm* (accessd November 10 2024).
19. Wu MY, Gong HZ, Hu KR, Zheng HY, Wan X, Li J. Effect of syphilis infection on HIV acquisition: A systematic review and meta-analysis. Sex Transm Infect. 2021; 97:525-533.
20. The National Health Commission of the People's Republic of China. Action Plan for Eliminating Mother-to-Child Transmission of HIV, Syphilis, and Hepatitis B (2022-2025). *https://www.gov.cn/gongbao/content/2023/content_5741260.htm* (accessd November 20 2024). (in Chinese)

Received December 3, 2024; Revised January 20, 2025; Accepted January 24, 2025.

**Address correspondence to:*
Hongzhou Lu, Shenzhen Third People's Hospital, 29 Bulan Road, Shenzhen, Guangdong 518112, China.
E-mail: luhongzhou@fudan.edu.cn

Released online in J-STAGE as advance publication January 27, 2025.

Editorial

DOI: 10.5582/bst.2025.01108

Deep cervical lymphaticovenous anastomosis in Alzheimer's disease: A promising frontier or premature enthusiasm?

Ya-nan Ma[1], Zijie Wang[1], Wei Tang[2,*]

[1] Department of Neurosurgery, Haikou Affiliated Hospital of Central South University Xiangya School of Medicine, Haikou, China;
[2] International Health Care Center, National Center for Global Health and Medicine, Japan Institute for Health Security, Tokyo, Japan.

SUMMARY: Alzheimer's disease (AD) is a progressive neurodegenerative disorder marked by β-amyloid accumulation, tau pathology, and impaired metabolic waste clearance. Recent evidence suggests that meningeal lymphatic vessels (MLVs) contribute significantly to the drainage of cerebrospinal and interstitial fluid. Deep cervical lymphaticovenous anastomosis (LVA), a microsurgical technique designed to enhance this drainage, has been proposed as a potential therapeutic strategy for AD. Preliminary findings from exploratory studies in China indicate possible cognitive and biomarker improvements, but current evidence is limited by small sample sizes, non-randomized designs, and methodological variability. Without standardized protocols and rigorous clinical validation, the broader applicability of LVA remains uncertain. Further investigation through multicenter, controlled trials is essential to objectively assessing its safety, efficacy, and clinical relevance in the management of AD.

Keywords: meningeal lymphatic vessels, VEGF-C, Piezo1, metabolic clearance, neurodegenerative disease, surgical intervention

1. Introduction

Alzheimer's disease (AD) is the most prevalent form of dementia, accounting for 60-70% of all cases worldwide (*1*). As a progressive neurodegenerative condition, it clinically manifests as cognitive decline, behavioral disturbances, and impaired activities of daily living (*2*). Pathologically, it is characterized by extracellular deposition of β-amyloid (Aβ) plaques, intracellular neurofibrillary tangles consisting of hyperphosphorylated tau protein, synaptic dysfunction, and neuroinflammation (*3*). According to the World Alzheimer Report 2023, over 55 million people globally are living with dementia, and this number is projected to exceed 139 million by 2050 (*1*). In China, the prevalence of AD is increasing rapidly due to aging of the population and it affects more than 10 million individuals, making it one of the most burdensome chronic neurological diseases in the country (*4*).

Despite continuous progress in drug development, current treatments for AD remain largely symptomatic, aiming to temporarily improve cognition or delay decline. Cholinesterase inhibitors and NMDA receptor antagonists are commonly used, but they have limited impact on disease progression (*5*). Recently, the exploration of disease-modifying therapies has shifted attention to pathological mechanisms such as impaired clearance of brain-derived metabolic waste, and particularly Aβ and tau aggregates.

Stem cell-based therapies have shown promise in preclinical studies by modulating inflammation, enhancing neuroprotection, and promoting neurogenesis (*6*). In China, policy support for stem cell and exosome-based treatment of neurological diseases is growing. Notably, on March 22, 2025, during the Boao Lecheng Stem Cell Conference, Chinese regulatory authorities announced for the first time the official pathways for approval, pricing, admission criteria, and clinical translation of stem cells projects. Several innovative therapies were granted pilot application status. However, stem cell-based interventions for AD remain in the clinical trial stage and have yet to enter routine clinical practice (*2*).

In this context, a novel microsurgical approach known as deep cervical lymphaticovenous anastomosis (LVA) has garnered increasing attention in China. This technique aims to enhance the clearance of cerebrospinal fluid (CSF) and interstitial fluid (ISF) by reconstructing a drainage route between the meningeal lymphatics and venous system (*7*). Several clinical centers, including those in Hangzhou, Shanghai, Nanjing, Harbin, Zhengzhou, and Zunyi, have launched exploratory studies using LVA in patients with AD (Table 1) (*8-13*). Preliminary results suggest potential improvements in

Table 1. Summary of LVA performed at major clinical centers in China

Facility (*ref.*)	Number of cases (*n*)	Main inclusion criteria	Surgical approach	Postoperative follow-up and preliminary outcomes
Qiushi Hospital, Hangzhou (*8*)	200	Age 40–90; diagnosis >12 months; expected survival >6 months; severe cognitive impairment; imaging evidence of ventricular enlargement or cerebral atrophy; informed consent obtained	Deep cervical lymphaticovenous anastomosis in the level II/III neck region	Cognitive and self-care functions largely restored within 9 months postoperatively
The First People's Hospital of Zunyi (*9*)	100	Age 60-90; IWG-2021 AD diagnostic criteria; experimental (LVA + medication) *vs.* control (medication only); multidimensional cognitive, biomarker, and imaging assessment	Deep cervical lymph node/vessel–venous anastomosis	Experimental group displayed superior improvement on neuropsychological scales, biomarkers, and imaging over 24-month follow-up
Shanghai Ninth People's Hospital (*10*)	10	Age 50–75; MMSE > 10; AD confirmed by PET-MR and CSF testing; no major organ disease; informed consent obtained	Deep cervical lymphaticovenous anastomosis	Over 50% of patients exhibited marked alleviation of symptoms; follow-up of up to 5 months
Zhengzhou Central Hospital (*11*)	100	Not specified; all cases clinically diagnosed with AD	Deep cervical lymphaticovenous anastomosis	Most patients displayed significant alleviation of symptoms postoperatively
The Second Hospital Affiliated with Harbin Medical University (*12*)	100	Moderate-stage AD; exclusion of vascular, toxic, or hypoxic dementia; preoperative cognitive assessment, PET, and CSF Aβ/tau analysis	Deep cervical lymphaticovenous anastomosis	Reported efficacy of 80%; cognitive improvements observed in the majority of patients
Nanjing First Hospital (*13*)	26	PET/MRI-confirmed abnormal deposition of Aβ and tau proteins	Deep cervical lymphaticovenous anastomosis	Preliminary outcomes favorable; marked cognitive improvement tended to be observed

Abbreviations: AD: Alzheimer's disease; Aβ: β-amyloid; CSF: cerebrospinal fluid; IJV: internal jugular vein; IWG: International Working Group; LVA: lymphaticovenous anastomosis; MMSE: Mini-Mental State Examination; MRI: magnetic resonance imaging; PET: positron emission tomography; p-Tau: hyperphosphorylated tau protein.

cognition, imaging biomarkers, and metabolic clearance efficiency (*14*). While preliminary clinical findings from China are intriguing, the field now stands at a critical juncture: will LVA mark a paradigm shift in AD treatment, or have we gotten ahead of the evidence?

2. Anatomical basis and therapeutic mechanism of LVA

The clearance of intracranial metabolic waste has long been a central issue in understanding the pathogenesis of AD (*15*). Traditionally, the central nervous system was believed to lack a conventional lymphatic system, with metabolic waste being primarily eliminated through arachnoid granulations into the venous circulation (*16*). However, recent studies have noted the presence of specialized lymphatic structures in the meninges, known as meningeal lymphatic vessels (MLVs), which connect with the deep cervical lymph nodes (dcLNs) (*17,18*). This discovery provides anatomical evidence supporting the drainage of CSF and ISF from the brain (Figure 1).

Imaging data suggest that CSF efflux *via* perisinusal and paravascular meningeal lymphatic pathways may be significantly greater – potentially up to 180% – than drainage through basal dural lymphatic routes (*19*). Approximately 50% of CSF clearance is believed to occur through cervical lymphatic drainage into the cervical lymph nodes (*18,20,21*). The remaining CSF is routed through the spinal cord to mediastinal, iliac, and sacral lymph nodes (*22,23*), or drains *via* perivascular spaces (*21*). Toxic molecules including Aβ, hyperphosphorylated tau, inflammatory mediators, and other metabolic byproducts are transported out of the brain through these lymphatic channels (*17,24-26*). When the meningeal lymphatic system is impaired or obstructed, clearance efficiency declines, resulting in the accumulation of neurotoxic waste in the brain, activation of neuroinflammation, and the progression of neurodegeneration (*27*). These processes may play a key role in the pathogenesis and exacerbation of AD (*26,28*).

LVA has been proposed as a surgical intervention grounded in the clearance pathway hypothesis mentioned earlier. Using microsurgical techniques, LVA establishes an anastomosis between downstream lymphatic structures – such as deep cervical lymphatic vessels or nodes – and adjacent venous branches (*e.g.*, the internal jugular vein, IJV) to create a low-resistance drainage route, thereby enhancing the efflux of brain-derived metabolic waste (*7,29*). Intraoperatively, sodium fluorescein or indocyanine green (ICG) is often used in lymphatic mapping to assist in identifying functional lymphatic vessels (*30*). Under a high-magnification microsurgical field, lymphaticovenous anastomoses are typically performed using vessels with diameters between 0.5-0.8 mm (*14*). The most commonly employed techniques include end-to-side and end-to-end anastomoses, both of which are designed to minimize venous reflux and

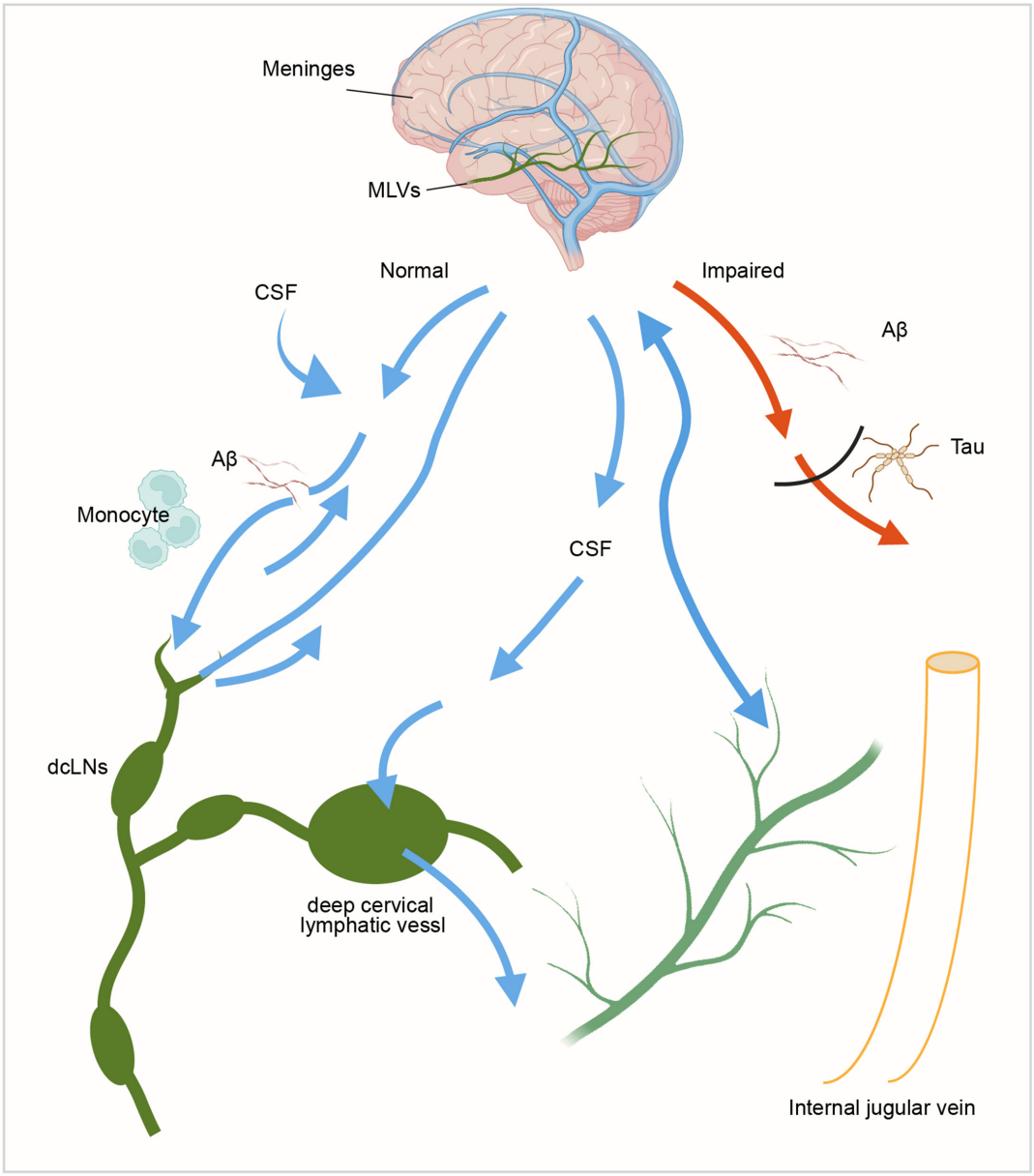


Figure 1. Schematic diagram of the MLVs and pathways of metabolic waste clearance. *Notes*: Green lines indicate lymphatic routes; blue, venous structures; arrows, waste flow direction. Aβ, β-amyloid; CSF, cerebrospinal fluid; dcLNs, deep cervical lymph nodes; MLVs, meningeal lymphatic vessels; p-Tau, hyperphosphorylated tau protein; IJV, internal jugular vein.

maintain the long-term patency of the connection (*7,14*). With the development of supermicrosurgery, some studies have explored finer anastomoses involving vessels smaller than 0.5 mm, which may enhance conduit stability and reduce tissue reactivity.

In patients with AD, functional evaluation of the MLV system can be performed using contrast-enhanced magnetic resonance imaging (MRI) (*31*). Imaging studies have demonstrated that MLV function is significantly impaired in AD, and particularly in its moderate to advanced stages, with lymphatic flow decreasing by nearly 40% compared to age-matched controls (*26,32*). Further evidence suggests that CSF flow disturbances are strongly correlated with cognitive decline and that the efficiency of Aβ clearance is positively associated with MLV functionality (*26*).

In an animal model of AD, reduced MLV function is directly associated with increased Aβ accumulation in the cortex and hippocampus, leading to neuronal damage and cognitive impairment (*33*). The decline in MLV function not only impairs the clearance of Aβ and other metabolic waste but also triggers neuroinflammatory responses. Studies have shown that during the progression of AD, VEGF-C expression in meningeal tissues declines significantly, leading to impaired lymphangiogenesis, reduced Aβ clearance capacity, and localized neurotoxic inflammation (*33*). Treatment with exogenous VEGF-C has been shown to improve Aβ clearance by over 40%, along with a significant enhancement in cognitive performance (*34*).

Additionally, mechanical stress induced by CSF flow dynamics may regulate VEGF-C expression and thereby modulate lymphatic function indirectly (*35*). In 5xFAD mice, ablation of MLVs results in a significant increase in meningeal macrophage populations within just one week, indicating that Aβ accumulation in dysfunctional MLVs induces local inflammation (*26*). Subsequent studies have demonstrated that restoring MLV function can effectively suppress microglial overactivation and attenuate chronic neuroinflammation within the brain (*36*). Other findings suggest that CSF hydrodynamic abnormalities associated with AD may impair Piezo1 channel activation, thereby reducing the mechanosensory capacity of MLVs, impairing waste clearance, and promoting Aβ deposition in brain tissues (*37*).

In summary, dysfunction of the MLV system is closely associated with impaired metabolic waste clearance and heightened neuroinflammation in AD. Given this mechanism, LVA represents a novel surgical intervention aimed at reestablishing lymphatic-venous drainage and improving intracranial metabolic homeostasis. It is rapidly gaining attention as a promising exploratory approach in the treatment of AD.

3. Clinical challenges and future directions

Currently, the use of LVA to treat AD remains in an exploratory phase. Most clinical studies conducted to date are observational or non-randomized, with small sample sizes, heterogeneity in study design, and inconsistent outcome measures. Critically, there is a lack of high-quality, multicenter, double-blind, prospective randomized controlled trials (RCTs), which significantly limits the robustness and generalizability of the evidence base. Given that AD is a slowly progressive neurodegenerative condition, short-term follow-up is insufficient to fully evaluate the long-term effects of a surgical intervention on disease trajectory. Large-scale, methodologically rigorous, and long-duration studies need to be promptly conducted to assess the sustainability of therapeutic benefits, clarify patient eligibility criteria, compare the efficacy of different surgical techniques, and document postoperative complications. A structured and scientifically sound research framework is essential.

Notably, there is substantial variability among clinical centers in terms of preoperative assessment and inclusion criteria. The absence of standardized patient selection protocols and clearly defined inclusion and exclusion criteria undermines the reliability of the current findings. Given the substantial clinical and pathological heterogeneity inherent in AD, whether LVA offers universal benefit remains unclear. Future efforts should prioritize the development of individualized screening models incorporating pathology subtypes, neuroimaging profiles, CSF dynamics, biomarker levels, and cognitive assessments. Such precision-based approaches would optimize patient selection, enhance treatment efficacy, and minimize unnecessary or ineffective interventions.

From a technical perspective, LVA is substantiated by a well-defined anatomical rationale, but there are still inconsistencies in its implementation. The type of anastomosis (*e.g.*, end-to-end, end-to-side, lymphatic valve reconstruction), choice of target vessels, intraoperative imaging techniques, and postoperative assessment protocols vary among facilities. Given that the procedure requires high-level supermicrosurgical skills and is technically demanding, differences in the surgeon's experience and technique may directly impact outcomes. To ensure safety, reproducibility, and broader adoption, a unified set of technical guidelines and a formalized training and credentialing system should be established.

Ethical and regulatory considerations are equally critical. As an invasive intervention, and particularly one in a population with cognitive impairment, the ethical performance of LVA must be rigorously upheld. Comprehensive informed consent procedures are essential, ensuring that patients and their caregivers fully understand the purpose, anticipated benefits, uncertainties, and potential risks of the surgery.

Despite these challenges, LVA represents a novel intervention that seeks to restore the brain's metabolic clearance pathways—an emerging paradigm in the management of AD. Future studies should explore the potential synergy between LVA and other therapeutic strategies, including stem cell therapy (*6*), anti-Aβ monoclonal antibodies (*38,39*), and neurorehabilitation techniques such as sensory-paired associative stimulation (SPA) (*40*). International research, including work by Louveau *et al*. (*41*) and Iliff *et al*. (*42*), has laid a theoretical foundation for the role of the MLV system in neurological disease. Future clinical trials of LVA should adopt internationally benchmarked methodologies, including multicenter, double-blind, and stratified randomization designs, to elevate the level of evidence and facilitate a global consensus.

4. Conclusion

LVA has emerged as a promising microsurgical technique based on the MLV system's role in clearing brain metabolic waste. Early clinical studies suggest that LVA may improve waste drainage and delay disease progression in AD, with preliminary evidence supporting its short-term efficacy and safety. China is at the forefront of global exploration in this area, with multiple centers reporting initial procedural experience and technical innovation.

Nevertheless, LVA remains in the early stages of clinical validation. Most existing studies are limited in sample size and methodological rigor and lack standardized protocols or high-quality evidence to support widespread clinical adoption.

In summary, LVA introduces a novel therapeutic

concept centered on reconstructing brain clearance pathways. It expands the scope of AD treatment beyond conventional pharmacology. As a technology still under evaluation, its future clinical relevance will depend on in-depth research into mechanisms, rigorous validation in clinical trials, and collaborative development of standardized procedures. Only through the convergence of verified efficacy, technical standardization, and robust regulatory oversight can LVA attain a clearly defined role in the evolving multimodal treatment landscape of AD.

Funding: This work was supported by grants from National Natural Science Foundation of China (no. 82460268), the Hainan Province Clinical Medical Research Center (no. LCYX202309), and the Hainan Province Postdoctoral Research Project (no. 403254).

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Collaborators GBDDF. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019. Lancet Public Health. 2022; 7:e105-e125.
2. Ma YN, Hu X, Karako K, Song P, Tang W, Xia Y. Exploring the multiple therapeutic mechanisms and challenges of mesenchymal stem cell-derived exosomes in Alzheimer's disease. Biosci Trends. 2024; 18:413-430.
3. Hu X, Ma YN, Xia Y. Association between abnormal lipid metabolism and Alzheimer's disease: New research has revealed significant findings on the *APOE4* genotype in microglia. Biosci Trends. 2024; 18:195-197.
4. Wang Gang, Qi Jinlei, Liu Xinya, *et al*. China Alzheimer's Disease Report 2024. Diagnostics Theory and Practice. 2024; 23:219-256. (in Chinese)
5. Ya-nan Ma XH, Ying Xia, Peipei Song, Wei Tang. Unraveling the mysteries of Alzheimer's disease: The critical role of the brain lymphatic system and therapeutic prospects. The New Perspectives Journal of Medicine. 2024; 1:257-259. (in Chinese)
6. Hu X, Ma YN, Peng J, Wang Z, Liang Y, Xia Y. Exosomes derived from olfactory mucosa mesenchymal stem cells attenuate cognitive impairment in a mouse model of Alzheimer's disease. Biosci Trends. 2025. doi: 10.5582/bst.2025.01065.
7. Hong JP, Chen WF, Nguyen DH, Xie Q. A proposed role for lymphatic supermicrosurgery in the management of Alzheimer's disease: A primer for reconstructive microsurgeons. Arch Plast Surg. 2025; 52:96-103.
8. Qiushi Hospital, Hangzhou. Exploring professor Xie Qingping's surgical team: The birthplace of a groundbreaking procedure to alleviate Alzheimer's disease. *https://www.qiushihospital.com/article/5387575015001878* (accessd March 20, 2025) (in Chinese)
9. The First People's Hospital of Zunyi. From forgetfulness to recalling memories: After surgical treatment for Alzheimer's disease, he called out his daughter's name. *http://www.zunyihospital.com/culture_xxzl/2024/xboQAYeg.html* (accessd March 20, 2025) (in Chinese)
10. Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine. Clinical study by Shanghai experts: Surgery improves cerebral lymphatic circulation and slows the progression of severe Alzheimer's disease. *https://www.9hospital.com.cn/djy_web/html/djy/jy_xwgg_mtjd/2024-09-24/Detail_474011.htm* (accessd March 20, 2025) (in Chinese)
11. Zhengzhou Central Hospital. 100 cases in 9 months! First in the province to pioneer minimally invasive surgery for Alzheimer's disease. *https://www.zzszxyy.com/detail/8416.html* (accessd March 20, 2025) (in Chinese)
12. The Second Hospital Affiliated with Harbin Medical University. Cervical deep lymphatic-venous anastomosis for the treatment of Alzheimer's disease. *https://www.hrbmush.edu.cn/info/1155/8534.htm* (accessd March 20, 2025) (in Chinese)
13. Nanjing First Hospital. Surgical treatment of Alzheimer's disease. *https://mp.weixin.qq.com/s/eF-Xk-gLdoRi6KokTzOdqw* (accessd March 20, 2025) (in Chinese)
14. Li Ke, Wen Ke, Ai Songtao, *et al*. Preliminary clinical observation on the treatment of Alzheimer's disease with cervical level II/III lymphatic-venous anastomosis. Tissue Engineer Reconstruct Surg. 2025; 21:10-13. (in Chinese)
15. Iliff JJ, Wang M, Liao Y, Plogg BA, Peng W, Gundersen GA, Benveniste H, Vates GE, Deane R, Goldman SA, Nagelhus EA, Nedergaard M. A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid beta. Sci Transl Med. 2012; 4:147ra111.
16. Sandrone S, Moreno-Zambrano D, Kipnis J, van Gijn J. A (delayed) history of the brain lymphatic system. Nat Med. 2019; 25:538-540.
17. Li W, Chen D, Liu N, Luan Y, Zhu S, Wang H. Modulation of lymphatic transport in the central nervous system. Theranostics. 2022; 12:1117-1131.
18. Proulx ST. Cerebrospinal fluid outflow: A review of the historical and contemporary evidence for arachnoid villi, perineural routes, and dural lymphatics. Cell Mol Life Sci. 2021; 78:2429-2457.
19. Yoon JH, Jin H, Kim HJ, Hong SP, Yang MJ, Ahn JH, Kim YC, Seo J, Lee Y, McDonald DM, Davis MJ, Koh GY. Nasopharyngeal lymphatic plexus is a hub for cerebrospinal fluid drainage. Nature. 2024; 625:768-777.
20. Ma Q, Ineichen BV, Detmar M, Proulx ST. Outflow of cerebrospinal fluid is predominantly through lymphatic vessels and is reduced in aged mice. Nat Commun. 2017; 8:1434.
21. Cserr HF, Harling-Berg CJ, Knopf PM. Drainage of brain extracellular fluid into blood and deep cervical lymph and its immunological significance. Brain Pathol. 1992; 2:269-276.
22. Jacob L, Boisserand LSB, Geraldo LHM, *et al*. Anatomy and function of the vertebral column lymphatic network in mice. Nat Commun. 2019; 10:4594.
23. Ma Q, Decker Y, Muller A, Ineichen BV, Proulx ST. Clearance of cerebrospinal fluid from the sacral spine through lymphatic vessels. J Exp Med. 2019; 216:2492-2502.
24. Zou W, Pu T, Feng W, Lu M, Zheng Y, Du R, Xiao M, Hu G. Blocking meningeal lymphatic drainage aggravates Parkinson's disease-like pathology in mice overexpressing mutated alpha-synuclein. Transl Neurodegener. 2019; 8:7.
25. Ding XB, Wang XX, Xia DH, *et al*. Impaired meningeal lymphatic drainage in patients with idiopathic Parkinson's

disease. Nat Med. 2021; 27:411-418.
26. Da Mesquita S, Louveau A, Vaccari A, *et al*. Functional aspects of meningeal lymphatics in ageing and Alzheimer's disease. Nature. 2018; 560:185-191.
27. Kim K, Abramishvili D, Du S, Papadopoulos Z, Cao J, Herz J, Smirnov I, Thomas JL, Colonna M, Kipnis J. Meningeal lymphatics-microglia axis regulates synaptic physiology. Cell. 2025. doi:10.1016/j.cell.2025.02.022
28. Laaker C, Baenen C, Kovacs KG, Sandor M, Fabry Z. Immune cells as messengers from the CNS to the periphery: The role of the meningeal lymphatic system in immune cell migration from the CNS. Front Immunol. 2023; 14:1233908.
29. Xie Q, Louveau A, Pandey S, Zeng W, Chen WF. Rewiring the brain: The next frontier in supermicrosurgery. Plast Reconstr Surg. 2024; 153:494e-495e.
30. Gan Cheng, Kong Zhengdong, Ran Xiaoye, *et al*. Short-term outcomes of deep cervical lymph node–venous anastomosis in the treatment of Alzheimer's disease. Chin J Plastic Surg. 2025; 41:130-143. (in Chinese)
31. Ringstad G, Vatnehol SAS, Eide PK. Glymphatic MRI in idiopathic normal pressure hydrocephalus. Brain. 2017; 140:2691-2705.
32. Kress BT, Iliff JJ, Xia M, Wang M, Wei HS, Zeppenfeld D, Xie L, Kang H, Xu Q, Liew JA, Plog BA, Ding F, Deane R, Nedergaard M. Impairment of paravascular clearance pathways in the aging brain. Ann Neurol. 2014; 76:845-861.
33. Da Mesquita S, Papadopoulos Z, Dykstra T, *et al*. Meningeal lymphatics affect microglia responses and anti-Abeta immunotherapy. Nature. 2021; 593:255-260.
34. Wen YR, Yang JH, Wang X, Yao ZB. Induced dural lymphangiogenesis facilities soluble amyloid-beta clearance from brain in a transgenic mouse model of Alzheimer's disease. Neural Regen Res. 2018; 13:709-716.
35. Nikolenko VN, Oganesyan MV, Vovkogon AD, *et al*. Current understanding of central nervous system drainage systems: Implications in the context of neurodegenerative diseases. Curr Neuropharmacol. 2020; 18:1054-1063.
36. Virenque A, Koivisto H, Antila S, Zub E, Rooney EJ, Miszczuk D, Muller A, Stoka E, Marchi N, Alitalo K, Tanila H, Noe FM. Significance of developmental meningeal lymphatic dysfunction in experimental post-traumatic injury. Brain Behav Immun Health. 2022; 23:100466.
37. Xiao B. Levering mechanically activated Piezo channels for potential pharmacological intervention. Annu Rev Pharmacol Toxicol. 2020; 60:195-218.
38. Lu D, Dou F, Gao J. Development of amyloid beta-directed antibodies against Alzheimer's disease: Twists and turns. Drug Discov Ther. 2023; 17:440-444.
39. Yang P, Sun F. Aducanumab: The first targeted Alzheimer's therapy. Drug Discov Ther. 2021; 15:166-168.
40. Peng Y, Zou Y, Asakawa T. The glamor of and insights regarding hydrotherapy, from simple immersion to advanced computer-assisted exercises: A narrative review. Biosci Trends. 2025; 19:10-30.
41. Louveau A, Herz J, Alme MN, *et al*. CNS lymphatic drainage and neuroinflammation are regulated by meningeal lymphatic vasculature. Nat Neurosci. 2018; 21:1380-1391.
42. Zeppenfeld DM, Simon M, Haswell JD, D'Abreo D, Murchison C, Quinn JF, Grafe MR, Woltjer RL, Kaye J, Iliff JJ. Association of perivascular localization of Aquaporin-4 with cognition and Alzheimer disease in aging brains. JAMA Neurol. 2017; 74:91-99.

Received March 14, 2025; Revised April 8, 2025; Accepted April 12, 2025.

**Address correspondence to:*
Wei Tang, International Health Care Center, National Center for Global Health and Medicine, Japan Institute for Health Security, Tokyo 162-8655, Japan.
E-mail: politang-tky@umin.ac.jp

Released online in J-STAGE as advance publication April 15, 2025.

Review

DOI: 10.5582/bst.2025.01045

Artificial intelligence in colorectal cancer liver metastases: From classification to precision medicine

Runze Huang[1,2,§], Xin Jin[1,2,§], Qinyu Liu[1,2], Xuanci Bai[3], Kenji Karako[4], Wei Tang[4,5], Lu Wang[1,2,*], Weiping Zhu[1,2,*]

[1] Department of Hepatic Surgery, Fudan University Shanghai Cancer Center, Shanghai Medical College, Fudan University, Shanghai, China;
[2] Department of Oncology, Shanghai Medical College Fudan University, Shanghai, China;
[3] Department of Clinical Medicine, Shanghai Medical College, Fudan University, Shanghai, China;
[4] Hepato-Biliary-Pancreatic Surgery Division, Department of Surgery, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan;
[5] National Center for Global Health and Medicine, Japan Institute for Health Security, Tokyo, Japan.

SUMMARY: Colorectal cancer liver metastasis (CRLM) remains the leading cause of mortality among colorectal cancer (CRC) patients, with more than half eventually developing hepatic metastases. Achieving long-term survival in CRLM necessitates early detection, robust stratification, and precision treatment tailored to individual classifications. These processes encompass critical aspects such as tumor staging, predictive modeling of therapeutic responses, and risk stratification for survival outcomes. The rapid evolution of artificial intelligence (AI) has ushered in unprecedented opportunities to address these challenges, offering transformative potential for clinical oncology. This review summarizes the current methodologies for CRLM grading and classification, alongside a detailed discussion of the machine learning models commonly used in oncology and AI-driven applications. It also highlights recent advances in using AI to refine CRLM subtyping and precision medicine approaches, underscoring the indispensable role of interdisciplinary collaboration between clinical oncology and the computational sciences in driving innovation and improving patient outcomes in metastatic colorectal cancer.

Keywords: colorectal cancer liver metastasis, artificial intelligence, classification, precision medicine

1. Introduction

Colorectal cancer (CRC) is among the most prevalent malignancies globally, with over 50% of patients eventually developing colorectal liver metastasis (CRLM) (*1,2*). The high incidence of CRLM and its pivotal role in degrading patient survival underscores the importance of early and accurate classification within this population. Precise classification of CRLM serves as a cornerstone for optimizing therapeutic strategies. Moreover, it plays a pivotal role in predicting treatment responses and patient outcomes, thereby enabling more personalized and effective clinical management (*3*).

Conventional classification methods, such as histopathological analysis, imaging evaluation, and clinical risk scoring (CRS), while valuable, have notable limitations, including subjectivity, time spent, and dependency on expert interpretation (*4*). In contrast, artificial intelligence (AI) offers the potential for automated, efficient, and scalable classification, addressing the constraints of conventional approaches. AI excels in handling multimodal data, integrating information from imaging, genomics, and clinical parameters to enhance the accuracy of classification models (*5,6*).

This review provides a comprehensive overview of the role of machine learning (ML) in CRLM classification, focusing on current methodologies, data applications, and future directions. Specifically, the discussion covers established classification frameworks for CRLM, including intrapatient stratification (*e.g.* sensitivity to treatment) and interpatient subgrouping (*e.g.* distinguishing CRLM from liver metastases of non-CRC origins). By integrating AI advances in clinical use, this review aims to highlight the transformative potential of AI in CRLM management, promoting the advancement of precision medicine in oncology.

2. The concept of integrating AI into CRLM classification

In clinical practice, metastatic liver cancer staging primarily relies on the TNM system established by the AJCC in 2017, which classifies cancer based on

tumor invasion, lymph node involvement, and distant metastases. However, growing evidence suggests that this pathological classification often fails to fully reflect patient heterogeneity at specific stages. For example, with current advances in medical care, many patients classified as having advanced-stage disease under the TNM system still demonstrate the potential for long-term survival (*7*). Consequently, clinical guidelines are increasingly reevaluating its role in determining surgical indications.

A more comprehensive and precise staging system is urgently needed to guide personalized cancer treatment. Advances in AI, multi-omics sequencing, and clinical data integration are enabling more accurate and efficient classification models. ML-based systems outperform conventional methods in predicting treatment responses and prognoses, while AI-driven clinical decision support systems (CDSS) are transforming oncology care. This review aims to explore and validate these emerging possibilities (Figure 1)

2.1 Current methods of clinical classification of CRLM

2.1.1. Histopathological classification (HGP)

As early as 2001, Vermeulen *et al*. identified three histopathological growth patterns (HGPs) in HE-stained sections of CRLM: desmoplastic (dHGP), pushing (pHGP), and replacement (rHGP) (*8*). In dHGP, the metastatic lesion is separated from the liver parenchyma by a stromal layer, with tumor cells infiltrating the matrix but not directly contacting hepatocytes. In pHGP, only a thin reticulin fiber layer separates tumor cells from hepatocytes, with metastatic lesions compressing and displacing hepatic plates. Unlike these patterns, rHGP preserves liver architecture, as tumor cells replace hepatocytes within hepatic plates while maintaining direct contact with normal hepatocytes. Notably, pHGP exhibits a higher ratio of proliferative endothelial cells compared to the other two (*9*).

Studies have demonstrated that different HGP patterns have a significant impact on patient prognosis. For instance, pHGP is associated with poorer survival outcomes (*10*), whereas dHGP correlates with better survival (*11,12*). Moreover, HGP classification also aids in developing various clinical strategies. As an example, Lazaris *et al*. demonstrated that bevacizumab is more effective in treating dHGP-type CRLM with abundant angiogenesis compared to rHGP-type CRLM (*13*). However, despite being a valuable classification system, HGPs has several limitations. For example, the growth patterns of tumor tissues may be altered by chemotherapy, and the classification still relies on postoperative histological analysis of tissue sections.

2.1.2. MMR/MSI classification

The DNA mismatch repair (MMR) system plays a critical role in correcting base mismatches or insertion/deletion errors that occur during DNA replication, and it was first identified as being associated with the progression of CRC (*14*). Defects in the MMR system led to the microsatellite instability (MSI) phenotype, also known as deficient MMR (dMMR). Based on the MMR/

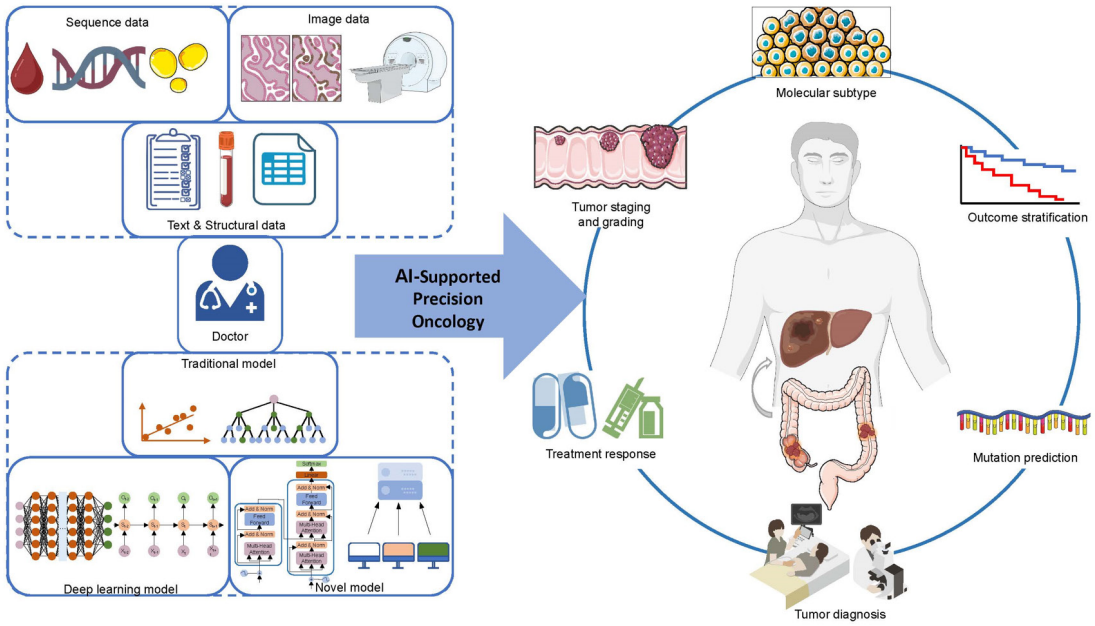


Figure 1. AI-driven Framework for Precision Oncology in Colorectal Cancer Liver Metastases. This figure depicts the integration of sequence data, imaging, and structured data with AI models to enhance tasks such as tumor classification, molecular subtyping, mutation prediction, and evaluation of treatment response. Central to this workflow is the use of advanced machine learning and deep learning techniques to facilitate personalized clinical decision-making.

microsatellite status, metastatic CRC patients can be classified into microsatellite stable (MSS, also referred to as proficient MMR, pMMR) and MSI (dMMR). Studies have shown that mCRC patients with dMMR generally have poorer survival outcomes compared to those with pMMR (*15,16*). However, these dMMR patients represent a small subset, accounting for only 3–5% of cases. Recent research has highlighted that monoclonal antibodies targeting immune checkpoints such as programmed cell death protein-1 and cytotoxic T-lymphocyte-associated protein 4 exhibit remarkable and durable benefits in this minority of MSI patients (*17-20*). Additionally, MSI status serves as a predictive biomarker for sensitivity to immune checkpoint blockade (ICB) therapy (*21*). In contrast, ICB therapies have not demonstrated superior efficacy over standard treatments in pMMR mCRC patients, underscoring the need for further exploration in this area (*22*).

2.1.3. Clinical staging (TNM)

The TNM staging system is based on the tumor, lymph node, and metastasis (TNM) concept first proposed by Pierre Denoix in the 1940s and 1950s. It remains the most commonly used staging system in the clinical management of CRLM (*23,24*). This system classifies cancer based on three key parameters: T refers to the size and depth of tumor invasion; N describes the involvement of regional lymph nodes; M indicates the presence of distant metastases (*25*). The TNM system provides a comprehensive framework for evaluating the severity and extent of tumor spread. However, due to its relatively narrow evaluation criteria, recent studies have suggested incorporating additional factors, such as tumor burden and the number of metastatic lesions, to improve the TNM system's prognostic accuracy for CRC patients and to better guide treatment strategies (*26,27*).

2.1.4. Clinical risk score

The clinical risk score has emerged in recent years as one of the most prominent tools for evaluation of colorectal cancer liver metastases (CRLM). A clinical risk score was initially proposed by Fong *et al*. in 1999. This landmark study analyzed clinical and pathological data from 1,001 consecutive patients and identified five key clinical indicators for the scoring system: the nodal status of the primary tumor, a disease-free interval of less than 12 months between the primary tumor and the detection of liver metastases, the presence of more than one tumor, a preoperative CEA level exceeding 200 ng/ml, and a maximum tumor diameter greater than 5 cm. Each criterion is assigned 1 point, and the total score stratifies patients by risk. The aforementioned study demonstrated that patients with lower CRS scores had significantly better 5-year survival rates compared to those with higher scores, and the CRS outperformed other scoring combinations available at the time. However, despite its pivotal role in clinical practice over the years, an increasing number of studies have sought to enhance the predictive power of the CRS system. This is being explored through the incorporation of additional clinical or molecular indicators, as well as by integrating CRS with other classification systems to address its limitations in specific contexts (*28,29*).

2.2. AI models commonly used in tumor classification

The standard workflow in ML research typically involves key steps such as data preprocessing, model construction, model training, parameter optimization, and external validation. Model selection plays a pivotal role in both data analysis and the advancement of research. An appropriate model not only significantly enhances research efficiency but also improves the accuracy and reliability of analytical results. This section systematically describes and discusses ML models, which are categorized into three types: conventional models, deep learning models, and emerging models.

2.2.1. Conventional ML models

Support vector machines: Support vector machines (SVM) were first developed by Cortes *et al*. in 1995 as a method for binary classification (*30*). Before the resurgence of deep learning, SVM was one of the most widely used ML models in various domains. The fundamental principle of SVM is to identify an optimal separating hyperplane in the feature space that divides data into distinct classes while maximizing the margin between them. To achieve this, SVM uses kernel methods, utilizing a mapping function to transform input data from its original feature space into a higher-dimensional space, where a separating hyperplane can be more easily identified. The strengths of SVM lie in its ability to handle high-dimensional data, its suitability for small sample sizes, and its excellent generalization capabilities. SVM has been widely applied to tumor subtyping and classification, including applications that incorporate imaging data (*31*) or transcriptomic data (*32*). However, SVM has certain limitations, such as longer training times for large datasets or datasets with high feature dimensions, and its sensitivity to noise and outliers, which can lead to overfitting.

Random forest: Random forest (RF), first developed by Breiman in 2001, is an ensemble learning method that solves classification and regression problems by constructing multiple independent decision trees (*33*). In the structure of RF, each internal node of a decision tree represents a feature, while each leaf node corresponds to a classification (category) or regression (numerical) outcome. The final prediction is generated by aggregating the outputs of all trees using a voting mechanism (for classification problems) or an averaging mechanism

(for regression problems). RF is highly robust against overfitting and can be parallelized to efficiently handle large-scale datasets. As a non-parametric approach, it effectively models complex nonlinear relationships and high-dimensional feature data, and it has been extensively applied to tumor subtyping and classification, including studies on breast cancer and pancreatic cancer (*34,35*). However, the ensemble nature of RF reduces its interpretability, and processing large datasets with many decision trees can require substantial time and computational resources.

Regression models: Regression models are widely used in statistics and ML to analyze relationships between a dependent variable and one or more independent variables. While regression methods can predict outcomes and explain variable influences, they are primarily statistical tools rather than standalone ML models. The core of regression models lies in finding a function that optimally maps the input values of independent variables to the output values of the dependent variable, typically by minimizing the error between predicted and observed values. Regression models encompass various types, including linear regression, logistic regression, lasso regression, and Cox regression. Cox regression, developed by Cox in 1972 (*36*), and lasso regression, developed by Tibshirani in 1996 (*37*), are the most commonly utilized in oncology research. For instance, Liu *et al*. utilized Cox regression to investigate the relationship between metabolic-associated fatty liver disease (MAFLD) and multiple cancers (*38*). Li *et al*. used a Bayesian lasso model to integrate multi-omics data for lung cancer classification (*39*). Each regression model has its own strengths and limitations. In general, regression models are simple, interpretable, and computationally efficient, with various regularization methods available to enhance their generalizability. However, they are also sensitive to outliers, heavily dependent on data characteristics, and often less effective when dealing with complex high-dimensional datasets.

Gradient boosting algorithm: Gradient boosting machine (GBM) is an ensemble learning method developed by Friedman in 2001 (*40*). It iteratively optimizes a target function to achieve the best possible solution by sequentially combining multiple weak learners (typically decision trees) into a strong learner. Each weak learner focuses on correcting the prediction errors of the previous model. GBM demonstrates exceptional performance in handling nonlinear, high-dimensional, and large-scale datasets, effectively capturing complex data patterns while maintaining robustness against noise and outliers. Popular implementations of gradient boosting include XGBoost, LightGBM, and CatBoost. For example, Rodriguez *et al*. used XGBoost combined with imaging and clinical parameters for risk stratification of hepatocellular carcinoma (HCC) patients (*41*), while Qi *et al*. used LightGBM as the optimal algorithm for predicting cardiovascular disease (CVD)-cancer comorbidity (*42*). However, GBM has some limitations, such as a lengthy training time for large datasets and a high dependence on hyperparameter tuning, which often requires extensive optimization to achieve peak performance.

k-nearest neighbors: The k-nearest neighbors (kNN) algorithm is an instance-based, non-parametric learning method known for its simplicity and intuitive nature. It predicts outcomes by measuring the similarity between samples. Specifically, kNN calculates the distance between an input sample and all training samples, selects the k nearest neighbors, and infers the target category (for classification problems) or value (for regression problems) based on the labels or values of these neighbors. A defining feature of kNN is its lack of an explicit training phase, as it relies primarily on the stored training data and distance computations. This simplicity makes it easy to implement and adapt. The algorithm has been used in various medical research fields. For instance, Wang *et al*. used kNN for lung cancer subtype classification (*43*), and modified kNN methods have been used to classify CRC tissues (*44*).

2.2.2. Deep learning models

Deep learning models (DLMs) are an extension of artificial neural networks (ANNs) and represent a more advanced and sophisticated branch of ML. Broadly, neural networks can be categorized into shallow neural networks (typically consisting of one or two hidden layers) and deep neural networks (DNNs, generally comprising three or more hidden layers). The latter forms the foundation and most prevalent framework for DLMs.

DLMs excel in non-linear modeling, making them effective for pattern recognition and predictive tasks. These networks act as multi-layer feature extractors, transforming input data (*e.g.* images or text) into abstract representations. Using these features, models interpret and process data for various applications, such as analyzing histopathological slides to distinguish tumor from non-tumor regions.

The following sections will delve into the foundational concepts of DLMs, focusing on deep neural networks as a paradigm of DLMs and their commonly implemented architectures, such as convolutional neural networks and recurrent neural networks.

Deep neural network: Deep neural networks (DNNs) are fundamental in deep learning, enabling hierarchical feature extraction and complex pattern recognition. A typical DNN consists of an input layer, multiple hidden layers, and an output layer, with each neuron in a layer connected to all neurons in the previous layer. DNNs operate through two key processes: forward propagation and backpropagation. In forward propagation, input data is passed layer by layer, with each neuron computing a weighted sum of its inputs, followed by a non-linear

activation function. This allows the network to model complex patterns. Backpropagation is the cornerstone of training DNNs. It calculates the error between predicted and true values, propagates it backward, and updates weights using optimization algorithms. This iterative process minimizes errors and refines model performance.

DNNs have displayed exceptional performance in handling highly complex datasets. For instance, Khan *et al*. integrated over 23,000 CT and pathology images to develop a multimodal DNN for predicting metastasis and variant classification of liver tumors, achieving an accuracy of 96.06% and an AUC of 0.832 (*45*). Nevertheless, DNNs have notable limitations. In addition to their high demands for large-scale data and computational resources, they are inherently "black box" models, making their learning processes and decision logic difficult to interpret—an issue that constrains their broader adoption in medicine. Moreover, conventional DNNs may suffer from shallow feature loss when dealing with high-dimensional, non-linearly distributed complex data, often requiring compensation or refinement through the introduction of attention mechanisms or skip connections.

Convolutional neural networks: Convolutional neural networks (CNN) are a class of DNNs that incorporate convolutional layers and that are particularly suited for processing data with grid-like topology, such as images and time-series data. Derived from multi-layer perceptrons, CNNs are designed to efficiently capture spatial locality in data by introducing specialized structures like convolutional and pooling layers.

A typical CNN architecture consists of five main components: an input layer, convolutional layers, pooling layers, fully connected layers, and an output layer. The convolutional layer is the backbone of CNNs, utilizing convolutional kernels (or filters, typically small matrices) to slide over the input data and extract local features such as edges, textures, and shapes. Each kernel learns specific feature patterns, with its parameters optimized through backpropagation. The pooling layer reduces the dimensions of the feature maps using down-sampling techniques, thereby decreasing computational complexity and enhancing translational invariance. Finally, the fully connected layer maps the extracted high-dimensional features into specific outputs, such as classification or regression predictions.

Thanks to their powerful feature extraction capabilities, CNNs excel in various computer vision tasks. For instance, Cho *et al*. used deep convolutional neural networks (DCNNs) in conjunction with image data to distinguish between benign and malignant lip skin lesions (*46*). Similarly, Chang *et al*. utilized CNNs combined with self-attention mechanisms to analyze histopathological slides in order to predict MSI status in CRC patients (*47*). However, CNNs also have notable limitations. They require large-scale training datasets and significant computational resources to achieve optimal performance. Moreover, CNNs are sensitive to hyperparameter settings, often necessitating extensive tuning to refine the model for specific applications.

Recurrent neural networks: Recurrent neural networks (RNNs) are specifically designed to process sequential data and can trace their origins back to Hopfield Networks of Associative Memory, developed by Hopfield in 1982 (*48*). Unlike conventional feedforward neural networks (*e.g.* DNNs), RNNs possess recurrent connections and memory capabilities, allowing them to retain information across time steps and respond to current inputs in the context of past information. This unique structure makes RNNs particularly suitable for tasks involving temporal dependencies, such as speech recognition, natural language processing, and time-series forecasting.

Despite their advantages, RNNs face challenges such as vanishing and exploding gradients, particularly when processing long sequences. To address these issues, several variants of RNNs have been developed, with the most prominent being the long short-term memory (LSTM) network (*49*). LSTM introduces a gating mechanism that regulates the retention and forgetting of information, overcoming the limitations of conventional RNNs in learning long-term dependencies. LSTM features three core gates—input, forget, and output gates—that collectively govern the flow and storage of information within the hidden states, enabling it to effectively capture long-range dependencies.

RNNs have accelerated advances in oncology research. For example, Yun *et al*. developed a transfer recurrent feature learning framework for intraoperative imaging and diagnosis of epithelial cancers (*50*). Similarly, a study combined CNNs with RNNs to differentiate benign from malignant fibroepithelial breast lesions, achieving promising results (*51*).

2.2.3. Emerging models and learning strategies

With the rapid advancement of AI technologies, an increasing number of emerging models and learning strategies are being applied to tumor classification and other related medical tasks. These approaches place greater emphasis on multimodal data integration, few-shot learning, and model interpretability, addressing the limitations of conventional models while driving innovation in the use of ML in medicine.

Transformer: The Transformer is a deep learning architecture based on attention mechanisms, initially designed for natural language processing tasks. With its typically deeper and more sophisticated layer design, the Transformer is categorized as a type of DNN. It processes input text or data sequences by dividing them into segments and using attention scores to determine the weight of each segment in the output module.

Compared to previous models, the Transformer uses multi-head self-attention mechanisms to process

input sequences in parallel, significantly improving computational efficiency and enabling effective modeling of long-range dependencies. The Transformer architecture consists of two main components: the encoder and the decoder. The encoder transforms input data into abstract contextual representations, while the decoder generates target sequences based on these contextual representations.

In tasks such as image segmentation and tumor classification, Transformers demonstrate exceptional performance. For instance, Xin *et al*. developed an improved Transformer model for skin cancer classification, achieving an accuracy exceeding 94% (*52*). Similarly, Xu *et al*. proposed a Transformer-based model, Prov-GigaPath, which not only classified subtypes across multiple cancer types but also identified molecular expression patterns and predicted gene mutations from histopathological slides, outperforming conventional models in various aspects (*53*). Nevertheless, the Transformer architecture faces certain challenges. Its complex and opaque internal mechanisms make the decision-making process difficult to interpret, and its high demand for computational resources remains a significant obstacle to its widespread use in oncology.

Multimodal ML: Multimodal ML is an approach that integrates information from different data sources to extract complementary features from multiple types of data simultaneously. Given the diverse data involved in oncology research—such as imaging, genomic, and transcriptomic data—multimodal ML has high compatibility and significant potential for advancing cancer research.

Just like DNNs, multimodal learning does not refer to a specific model but rather demonstrates the variety of data. For instance, Qian *et al*. reported the development of a multimodal model named BMU-Net, which integrates clinical data, mammographic images, and trimodal ultrasound data to diagnose benign and malignant breast tumors, achieving an overall diagnostic accuracy exceeding 90% (*54*). Multimodal ML models are evolving toward more efficient data fusion, improved interpretability, and enhanced clinical applicability. By integrating data from multiple modalities, these models can capture deeper insights that are often unavailable from a single data source, thereby offering more accurate and comprehensive support for tumor classification and diagnosis.

Self-supervised learning: In AI, ML approaches can be categorized into supervised learning, unsupervised learning, and reinforcement learning, based on whether the analyzed data includes specific labels or annotations. Self-supervised learning (SSL) is considered an extension of unsupervised learning. Unlike supervised methods, SSL does not require extensive labeled datasets. Instead, its core principle is to construct pretext tasks that enable models to extract meaningful feature representations from unlabeled data for downstream tasks. SSL methods are broadly divided into two main categories: generative methods and contrastive methods. Generative methods train models by reconstructing data, such as completing images or predicting missing words, making them ideal for reconstruction tasks. Contrastive methods, in contrast, use positive and negative sample pairs to help models distinguish similarities and differences.

These methods are particularly effective in image classification and data representation tasks. For example, Schirris *et al*. developed DeepSMILE, a contrastive SSL framework for classifying whole slide images of HE-stained tissue sections (*55*). Similarly, Zhang *et al*. developed SANDI, a model capable of spatial cellular classification, by first learning pairwise similarities among unlabeled data and subsequently incorporating reference data (*56*). While SSL eliminates the dependency on large-scale labeled datasets, developing high-performance SSL models requires carefully designed pretext tasks and significant computational resources.

In conclusion, while ML models face challenges in data dependency and interpretability, AI integration into medicine is a key direction for the future. By combining AI with conventional methodologies, particularly in cancer detection, diagnosis, subtyping, and personalized treatment, AI-driven research is advancing precision medicine and overcoming technical barriers in healthcare.

3. Application of ML to CRLM classification

Based on task requirements and technical methodologies, the application of ML to CRC liver metastasis classification can be divided into two major categories: single-modality task classification and multi-modality task classification. The former can be further subdivided into three subcategories: classification based on imaging data, classification based on omics data, and classification based on structured data.

3.1. Single-modality task classification

3.1.1. Based on imaging data

ML has been extensively applied to imaging data for CRLM classification, with researchers exploring both conventional ML models and advanced deep learning frameworks. The following studies demonstrate the diversity of approaches and highlight their respective strengths and limitations (onlin data: Table 1, *https://www.biosciencetrends.com/supplementaldata/252*).

Tharmaseelan *et al*. conducted a study using CT imaging data from 78 patients, encompassing 1,296 metastatic liver lesions, to evaluate the performance of various ML models (*57*). These models included conventional ML classifiers such as XGBoost and kNN, as well as a DLM based on CNN. The CNN model was

derived from the DenseNet-121 architecture and trained using the PyTorch platform. The objective was to identify the primary tumor site in gastrointestinal cancer patients with liver metastases. Interestingly, the kNN model achieved the highest discriminative ability (AUC: 0.87), outperforming the CNN model (AUC: 0.80). However, the CNN model demonstrated superior accuracy (0.83 *vs.* 0.67). These findings suggest that conventional ML models may, in certain classification tasks, perform comparably or even better than advanced DLMs, and especially with limited datasets.

Building upon this, Jia *et al*. proposed a DLM based on CT imaging to identify the primary tumor sites in patients with liver metastases (*58*). Their study included imaging data from 489 patients and a total of 769 metastatic liver lesions. To provide a comparative analysis, the researchers also developed conventional ML models, including Decision Tree, RF, and kNN. With a larger sample size, the DLM outperformed conventional ML models in all metrics. Specifically, the DLM achieved an accuracy of 0.714 and an AUC of 0.811 on the validation set, and external validation yielded an accuracy of 0.667 and an AUC of 0.784. In comparison, the best-performing conventional model, RF, achieved a maximum AUC of 0.775 and an accuracy of 0.655. These results highlight the advantages of DLMs in using larger datasets for superior performance.

Moving into histopathological growth pattern analysis, Höppener *et al*. developed a CNN-based model for binary classification of HGPs in liver metastases (desmoplastic *vs.* non-desmoplastic) using digitalized whole-slide images (*59*). Their algorithm, neural image compression (NIC), is a multi-task learning framework that compresses high-dimensional image patches into low-dimensional embeddings while preserving spatial information and suppressing noise. The study used 3,641 slides from 932 patients for training and 870 slides for external validation. The model achieved outstanding results, with an AUC of 0.93 on the training set and 0.95 on the validation set. By using supervised training across multiple histopathological tasks, NIC demonstrated the potential of multi-task learning in extracting transferable features for robust classification.

Similarly, Starmans *et al*. explored the use of CNNs to classify the HGPs of CRLM using CT data (*60*). The study used multi-observer segmentation, combining data from three human observers to train the model. Each lesion was segmented three times, effectively tripling the training sample size. Interestingly, the performance of the multi-observer model (AUC: 0.69) was comparable to the single-observer models (maximum AUC: 0.72). Despite exploring ICC-based feature selection and ComBat for further analysis, these methods did not significantly improve performance. The aforementioned study highlights the challenges of utilizing multi-observer data and suggests the importance of optimizing segmentation techniques for better performance.

Turning to genetic mutation prediction, Wesdorp *et al*. developed models based on RF and gradient boosting algorithms to identify KRAS mutation status using CT imaging data (*61*). The study included 255 CRLM patients, split into training (n = 204) and test (n = 51) sets. While the ensemble classifier performed well on the test set (AUC: 0.86), it underperformed in external validation (AUC: 0.47). In contrast, RF demonstrated relatively better external performance (AUC: 0.54). These results reflect ongoing challenges in linking imaging features to genetic mutations, exacerbated by small sample sizes and insufficient preprocessing.

Similarly, Granata *et al*. utilized CT data to predict RAS mutations (*62*). They extracted 851 radiomic features from 77 liver metastases in 28 patients and constructed multiple ML models, including logistic regression, decision trees, kNN, and SVM. Multivariable analysis using logistic regression achieved superior performance (AUC: 0.953, accuracy: 98%), especially after z-score normalization. However, the authors noted no significant improvements when applying normalization techniques, raising questions about their utility in radiomic analysis.

Finally, Li *et al*. developed a comprehensive platform, the Radiomics Intelligent Analysis Toolkit (RIAT), for predicting liver metastasis risk (*63*). By integrating multiple ML methods and clinical data, RIAT demonstrated the value of combining advanced statistical and ML techniques for robust diagnostic tool development. Similarly, Kim *et al*. applied YOLO-based deep learning to large-scale CT imaging (*64*), achieving sensitivity comparable to radiologists but emphasizing its role as an assistive, rather than standalone, diagnostic tool.

Together, these studies highlight the diversity of imaging-based ML applications in CRLM classification, emphasizing the importance of task-specific adaptations, model optimization, and data integration.

3.1.2. Based on omics data

In the context of CRLM classification, omics data provide a rich source of biological insights, enabling ML models to predict risk, classify subtypes, and identify molecular features with significant diagnostic and prognostic implications. The integration of multi-omics datasets with ML not only offers enhanced classification accuracy but also deepens our understanding of the underlying molecular mechanisms driving CRLM.

Yu *et al*. used the AdaBoost algorithm to predict the risk of liver metastases in CRC patients using blood test markers (*65*). The study compared AdaBoost to five other algorithms, including Extremely Randomized Trees (ERT), Multilayer Perceptron, Stochastic Gradient Descent (SGD), RF, and XGBoost. AdaBoost, which dynamically adjusts sample weights to optimize weak

learners, achieved the highest diagnostic accuracy (89.3%) and precision (89.4%). Interestingly, MLP demonstrated the weakest performance, with an accuracy of 79.6% and a precision of 80.1%. The superior performance of AdaBoost was attributed to its robustness with small datasets, whereas MLP's reliance on larger, high-dimensional data likely limited its effectiveness in that study. This underscores the potential of simpler, adaptive algorithms in data-limited clinical settings.

Extending the analysis to tissue-level investigations, Kiritani *et al*. developed a logistic regression model using mass spectrometry data from 103 CRLM samples and 80 normal tissue samples to distinguish metastatic from non-malignant tissues (*66*). The model underwent validation using leave-one-out cross-validation (LOOCV), 10-fold cross-validation, and an independent cohort of 40 samples (20 CRLM and 20 non-cancerous tissues). Phosphatidylcholine, phosphatidylethanolamine, and monounsaturated fatty acids were found to be enriched in CRLM tissues, with the model achieving an exceptional accuracy of 99.5% and an AUC of 0.9999. These findings highlight the synergy of mass spectrometry and ML in identifying molecular markers for CRLM diagnosis.

Taking a step further into molecular subtyping, Katipally *et al*. utilized data from the Phase 3 new EPOC randomized clinical trial to construct a neural network model for CRLM molecular subtyping (*29*). Sequencing data from 93 patients revealed 31 optimal features, including 24 mRNAs and 7 miRNAs, which were used for subtyping. In a validation cohort of 147 patients, the model classified CRLM into canonical, immune, and stromal subtypes, with immune subtype patients having the best 5-year OS (63%) and canonical subtype patients having the worst prognosis (43%). Incorporating molecular subtypes into clinical risk scores improved predictive performance (OS AUC increased from 0.59 to 0.63). The aforementioned study demonstrates how molecular subtyping can enhance both prognostic stratification and personalized therapeutic strategies.

Finally, Moosavi *et al*. developed an RF-based CRLM classification model using transcriptomic data from 171 patients (*67*). The study compared the new LMS subtyping framework to the CMS and CRIS classification systems, using 829 CRC samples, including CRLM, primary CRC tumors, non-malignant liver tissues, organoids, and cell lines. Unlike CMS, which struggled to classify CRLM and which was influenced by prior systemic treatments, LMS effectively stratified samples into five subtypes (LMS1-5). LMS1 was associated with the poorest prognosis (5-year OS of 15%, HR = 2.2, $p = 9 \times 10^{-4}$), while LMS5 exhibited stromal-like characteristics. LMS demonstrated superior prognostic stratification and independence from treatment-related biases, outperforming CMS and CRIS in this regard.

Together, these studies illustrate the potential of integrating omics data with ML for CRLM classification. From simple blood markers to comprehensive transcriptomic analyses, omics-driven ML approaches offer unparalleled opportunities to provide precision oncology, unravel molecular complexities, and provide robust frameworks for diagnosis and prognosis.

3.1.3. Based on structured data

Building on the success of imaging-based approaches, the application of ML to omics data has opened up new avenues for CRLM classification. By using molecular and biological datasets, these studies aim to glean deeper insights into tumor biology while improving diagnostic accuracy and prognostic predictions.

Nemlander *et al*. developed a gradient boosting model to identify non-metastatic colorectal cancer (NMCRC) patients during their first clinical visit (*68*). The study included 2,681 participants, consisting of 542 NMCRC patients and 2,139 matched controls. Clinical data used for model construction included age, sex, primary healthcare (PHC) unit, NMCRC stage (I-III), the number of general practitioner consultations in the previous year, and all diagnoses reported in VEGA within the preceding year. The dataset contained 360 different ICD-10 or KSH97-P diagnostic codes. Of the participants, 75% were used for training, while 25% were used for validation.

The model was constructed using the GBM package in R, with class-stratified 10-fold cross-validation. The final model correctly identified 99 out of 135 NMCRC cases, achieving a sensitivity of 73.3%, a specificity of 83.5%, and an AUC of 0.832. Among the 361 predictors, 184 variables were found to have predictive value, with 16 factors showing a normalized relative influence (NRI) >1%. Notable predictors included changes in bowel habits, other diseases of the anus and rectum, iron deficiency anemia, and other and unspecified non-infective gastroenteritis and colitis. These findings suggest that such symptoms may indicate an elevated risk of NMCRC.

Although studies utilizing structured data for CRLM classification are relatively scarce, this research highlights the potential of such data to contribute to early cancer detection. Structured data offers a non-invasive, cost-effective means of identifying diagnostic patterns that can complement other ML approaches in precision oncology.

3.2. Multimodal task classification

Moving beyond single-modality approaches, multimodal task classification integrates diverse datasets to improve predictive accuracy and uncover complex patterns in CRLM. By combining biological, clinical, and electronic health record (EHR) data, these models provide a comprehensive framework for understanding and

predicting disease progression.

Krishnan *et al*. developed a model using a Bayesian regularized neural network (BRANN) and sparse multilinear regression to classify CRC patients (*69*). The study integrated multiple biological datasets, including plasma lipid and protein levels, chemokines, gene mutation status, and clinical information. Initially, a regression model, MLR-EM, was constructed to extract key feature data, identifying 9 lipids as significant predictors for distinguishing CRLM patients. Using these features, the BRANN model, a variant of ANN with Bayesian regularization, successfully classified cancer-free individuals, CRC patients, and CRLM patients. The model had an R^2 of 0.77 and an accuracy of 87% on the training set and an R^2 of 0.68 and an accuracy of 77% on the test set. The aforementioned study demonstrates the potential of integrating biochemical and clinical data for accurate classification of disease stages in CRC patients.

Li *et al*. combined EHR information and laboratory data to construct NLP and ML models in order to predict the likelihood of postoperative liver metastases in CRC patients (*70*). The study included 1,463 CRC patients, 609 with CRLM and 854 without. A total of 18 features were analyzed using five conventional models and a bidirectional encoder representations from Transformer (BERT)-based NLP model. Among the conventional models, SVM demonstrated the best performance (AUC: 0.64, accuracy: 0.64), comparable to the NLP model (AUC: 0.676, accuracy: 0.636). When these two approaches were fused into a single model, the combined framework exhibited significantly enhanced performance, achieving an accuracy of 80.8% and precision of 80.3%. Moreover, the combined model outperformed physicians in an external validation cohort in both accuracy (0.760 *vs.* 0.658 and 0.640) and precision (0.763 *vs.* 0.697 and 0.670). These results highlight the potential of combining EHR data with advanced NLP and ML techniques to improve predictive accuracy for CRLM.

These studies underscore the advantages of multimodal approaches in CRLM classification, using complementary datasets to refine predictions and improve patient stratification. By integrating diverse data sources, multimodal models address the limitations of single-modality methods and pave the way for more robust and clinically actionable insights.

4. Clinical decision-making and treatment optimization based on cancer classification

4.1. Classification-guided personalized treatment

ML models are increasingly being used to guide clinical decision-making and optimize treatment strategies for CRLM patients. These models provide valuable tools for predicting therapeutic responses, stratifying patients, and personalizing treatment approaches. The following studies illustrate how classification results can inform clinical decisions and improve patient outcomes (onlin data: Table 2, *https://www.biosciencetrends.com/supplementaldata/252*).

To begin with, Karagkounis *et al*. developed an RF model to evaluate the pathological response of CRLM patients to chemotherapy (*71*). The study included 85 patients and 95 liver metastases, with 63 lesions classified as responders and 32 as non-responders based on histopathological assessments. To address a data imbalance, the authors implemented cost-sensitive learning by assigning higher penalties for misclassifying non-responders. The model outperformed conventional methods, including RECIST and morphological evaluation, achieving an AUC of 0.87 compared to 0.53 and 0.56, respectively. This demonstrates the potential of ML models to provide more accurate and nuanced assessments of chemotherapy responses.

Building on the use of CT data to predict chemotherapy response, Maaref *et al*. utilized CNNs to predict treatment responses in CRLM patients (*72*). The study included 202 patients with 444 lesions, where 230 had previously undergone FOLFOX-based chemotherapy. The CNN model achieved outstanding performance in distinguishing treated from untreated lesions (AUC: 0.97) and predicting chemotherapy responses (AUC: 0.88, sensitivity: 98.1%). These findings highlight the ability of CNNs to handle large imaging datasets and assist in managing metastatic lesions.

Expanding on the prediction of chemotherapy response, Davis *et al*. used an attention-based deep learning framework to analyze CT images and predict responses to neoadjuvant chemotherapy in CRLM patients (*73*). Using a dataset of 33,619 CT images from 95 patients, the model assigned attention weights to different image patches and achieved an AUC of 0.77, far surpassing the logistic regression model based on Fong scores (AUC: 0.41). These results emphasize the utility of attention mechanisms and multi-instance learning frameworks when analyzing complex imaging data with weak annotations.

Taking the next step toward multi-modal modeling, Qi *et al*. developed an artificial neural network (ANN) model to predict the sensitivity of unresectable CRLM patients to irinotecan-based chemotherapy (*74*). The study included 116 patients, randomly divided into training (n = 81) and validation (n = 35) sets. Feature selection using Pearson correlation and the MRMR algorithm identified key imaging and clinical variables for model construction. The primary ANN model (p-model) integrated multi-scale resampling of imaging features with clinical data, while three variant ANN models used only single-scale inputs. The p-model achieved an AUC of 0.754 (training) and 0.752 (validation), surpassing the best conventional model, XGBoost (AUC: 0.718 and 0.704). Further intra-ANN comparisons confirmed the superiority of multi-modal

integration, with the p-model outperforming single-scale ANN variants. The aforementioned study underscores the value of combining multi-scale imaging and clinical data to enhance chemotherapy response prediction, offering a promising tool for optimizing CRLM treatment strategies.

Focusing on precision medicine, Lu *et al*. developed a hybrid CNN-RNN model to predict VEGF therapy sensitivity in mCRC patients based on the VELOUR trial (*75*). By combining CNN-based feature extraction with RNN-based temporal sequence analysis, the model demonstrated superior performance in predicting early treatment responses (AUC: 0.76) compared to conventional RECIST (AUC: 0.66) and ETS (AUC: 0.60) standards. Moreover, responders identified by the model had a significantly longer median OS (18.0 months *vs.* 10.4 months for non-responders, HR = 0.49, $p = 1\times10^{-6}$). The aforementioned study highlights the potential of combining dynamic imaging data and ML for real-time therapeutic decision-making.

In terms of survival stratification, Endo *et al*. developed a decision-tree-based model to predict postoperative chemotherapy responses in CRLM patients (*76*). The study analyzed data from 1,358 patients, incorporating 18 demographic and clinicopathologic variables, including T stage, primary tumor location, and tumor burden score (TBS). Patients with lymph node metastasis, specific tumor locations, and certain KRAS statuses displayed significant benefits from adjuvant chemotherapy. Subgroup analyses revealed that patients with lymph node metastasis, left-sided or rectal primary tumors with low/high TBS, and right-sided tumors with KRAS mutations benefited significantly from adjuvant chemotherapy. The model demonstrated good predictive performance, with a C-index of 0.68 for OS and 0.69 for RFS in both training and test sets. These findings highlight the utility of incorporating clinicopathologic data into predictive models to guide adjuvant chemotherapy decisions and improve patient outcomes

Shifting focus to imaging data and biological response, Zhu *et al*. developed a ML model using pre- and post-chemotherapy MRI images to predict pathological tumor regression grade (TRG) in CRLM patients (*77*). The study included 180 patients (389 lesions) divided into training, test, and external validation sets. Implemented with TensorFlow and Keras, the model utilized multi-stream inputs and center cropping to enhance CNN performance. Three models with varying input streams were compared: Model A (four input streams), Model B (pre-treatment images), and Model C (post-treatment images). Model A achieved the highest AUC (0.849) with the training set, significantly outperforming Models B, C ($p = 0.04$), and RECIST ($p = 0.03$). In external validation, Model A maintained superior performance (AUC: 0.833, accuracy: 0.885) compared to RECIST (AUC: 0.558, accuracy: 0.533). Additionally, Model A effectively stratified survival outcomes, while RECIST-defined groups displayed no significant differences (DFS and OS, $p = 0.12$, $p = 0.99$). The aforementioned study underscores the potential of CNN-based models in improving chemotherapy response prediction and survival stratification over conventional RECIST assessments.

Giannini *et al*. utilized imaging data to predict treatment responses in HER2-amplified CRLM patients receiving HER2 dual-targeted therapy (*78*). The study included CT data from 38 patients and 141 metastatic lesions, with 28 patients (108 lesions) in the training set and 10 patients (33 lesions) in the validation set. The authors extracted 24 radiomic features from CT images and applied a Gaussian Naïve Bayes (GNB) classifier for feature selection, ultimately retaining 12 significant features. The GNB model performed better on the training set compared to the validation set, particularly in sensitivity (training: a sensitivity of 0.89, a specificity of 0.85; validation: a sensitivity of 0.90, a specificity of 0.42). The model correctly classified 24 of the 38 patients, partially misclassified 12, and completely misclassified 2. The authors noted that while the model effectively predicted responsive lesions (R+), it struggled to accurately identify non-responsive lesions (R-). The aforementioned study underscores the potential of radiomic feature-based models to predict treatment response in HER2-targeted therapies, while highlighting challenges in generalizability and specificity.

Together, these studies demonstrate the pivotal role of ML models in guiding clinical decisions and optimizing treatment strategies for CRLM. By improving the accuracy of therapeutic response predictions, stratifying patients based on clinical and molecular characteristics, and integrating multi-modal data, these models are driving precision oncology forward.

4.2. Patient prognostic stratification

ML has significantly enhanced prognostic stratification for CRLM patients, utilizing diverse data types to improve survival predictions and patient management. From imaging-based models to multi-modal approaches, these studies illustrate the versatility of ML in addressing clinical challenges.

Wang *et al*. developed an unsupervised ML model based on preoperative CT imaging and clinical data to stratify survival risks in 197 CRLM patients (*79*). Using hierarchical clustering, the study filtered imaging features from 851 to 56 through Cox regression and divided patients into favorable and poor prognosis groups, with the latter exhibiting an OS HR of 1.78 (95% CI: 1.12–2.83). The model outperformed CRS and TBS scores in predicting long-term survival, with a time-dependent AUC of 0.66 compared to 0.58 and 0.55, respectively.

Building on this, Paro *et al*. used a tumor burden-focused ML model, ML-TB, to optimize thresholds for tumor size and number, maximizing five-year survival

stratification (*80*). The study analyzed 1,344 patients from five centers and noted superior OS stratification compared to conventional Fong scores, with Cohen's d values of 1.61, 0.84, and 2.73, highlighting the model's ability to redefine tumor burden parameters for better clinical outcomes.

In a similar vein, Lam *et al*. incorporated lasso regression and Cox models to identify key predictors from 36 clinical variables in 572 patients (*81*). Variables such as CEA levels, tumor size, and KRAS mutation status were critical for OS and RFS predictions, achieving a concordance index of 0.651 and significantly outperforming Fong CRS in one- and five-year OS predictions. This comprehensive analysis underscores the importance of integrating clinicopathologic and molecular data into ML models for precise risk stratification.

Adding a histopathological dimension, Elforaici *et al*. used deep learning frameworks with GANs and Vision Transformers to analyze 1,620 pathology slides from 258 patients (*82*). The model extracted tumor and peritumoral features, achieving a c-index of 0.804 for OS and 0.735 for time-to-recurrence. By using multi-task deep learning, this approach demonstrated the potential to enhance prognostic precision through advanced histological insights.

Moro *et al*. utilized a classification and regression tree (CART) model to identify risk factors for CRLM prognosis in 1,123 patients (*83*). Based on demographic and clinicopathologic data, the model revealed distinct survival profiles for wtKRAS and mtKRAS patients. For instance, wtKRAS patients with small (<4.3 cm) solitary metastases and no nodal involvement exhibited the highest five-year OS (68.5%). The CART model also outperformed conventional Fong scores, and particularly for wtKRAS patients (AIC 3334 *vs.* 3341).

Incorporating imaging and molecular characteristics, Saber *et al*. utilized an attention-based TabNet model to predict levels of CD73 expression in 122 patients (*84*). By integrating immunofluorescence and CT data, the model achieved an AUC of 0.95 and yielded significant prognostic implications, with high levels of CD73 expression linked to shorter recurrence (13.0 *vs.* 23.6 months, p: 0.0098) and disease-specific survival (53.4 *vs.* 126.0 months, p: 0.0222). The aforementioned study emphasizes the role of molecular markers in stratifying treatment responses and outcomes.

Expanding the focus to targeted therapy, Zhou *et al*. developed the DERBY+ model to predict bevacizumab response using PET-CT and clinical data (*85*). Trained on multi-center cohorts, the model achieved an AUC of 0.95 with independent datasets, outperforming individual predictors such as clinical (AUC: 0.66) and imaging features (AUC: 0.72). The identified responders exhibited prolonged OS (27.6 *vs.* 18.5 months, p = 0.010), underscoring the utility of integrated ML frameworks for precision oncology.

Turning to recurrence prediction, Zhao *et al*. designed a hybrid DLM combining 2D-CNN, Bi-LSTM, and attention modules to predict early recurrence after thermal ablation (*86*). Analyzing 13,248 ultrasound images and clinical data from 207 patients, the combined model achieved an AUC of 0.78 and demonstrated significant prognostic stratification. Notably, the DL model consistently outperformed clinical models in all datasets, with significantly lower false-positive rates and better high-risk group identification ($p < 0.001$).

In the realm of disease-free survival prediction, Luo *et al*. compared elastic net (EN) and random survival forest (RSF) models using contrast-enhanced CT imaging data from 180 patients (*87*). The EN model outperformed RSF in the test set (C-index = 0.78), while RSF excelled in the training set (C-index = 0.74). Both models effectively stratified DFS outcomes, illustrating the complementarity of regression- and forest-based approaches in survival analysis.

Finally, Amygdalos *et al*. developed a gradient-boosted decision tree model to predict OS in 487 CRLM patients (*88*). By focusing on six top-ranked predictors, such as CEA levels and metastatic lesion size, the GBDT-Top6 model achieved a superior C-index of 0.70, outperforming the original GBDT (C-index: 0.65). This highlights the potential of feature selection in enhancing ML model performance and clinical interpretability.

Together, these studies underscore the transformative potential of ML in CRLM prognostic stratification. By integrating diverse data sources and using cutting-edge algorithms, these models will pave the way for more personalized and effective patient care.

In conclusion, ML has advanced clinical decision-making and prognostic stratification for CRLM patients by integrating clinical, imaging, and molecular data. Techniques such as RFs, regression trees, and deep learning have demonstrated effectiveness in predicting chemotherapy responses, stratifying survival risks, and enhancing prognostic accuracy. These advances highlight AI's potential to optimize personalized treatment and improve patient outcomes in CRLM management.

4. Conclusion

AI has shown great promise in classifying and managing CRLM, yet challenges remain in its clinical integration. The complexity of multimodal data, limited access to large annotated datasets, and ethical concerns such as data privacy and model transparency hinder their widespread use. Additionally, CRLM's biological heterogeneity requires AI models that are both adaptive and interpretable.

To overcome these barriers, future research should focus on federated learning to enable secure multi-institution collaboration, self-supervised and transfer learning to reduce dependence on labeled data, and improved model interpretability to enhance clinical

trust. Longitudinal studies integrating AI into real-world workflows will be essential for validation.

Despite these challenges, AI is transforming CRLM management by integrating clinical, imaging, and omics data for personalized treatment strategies. Advancing AI-driven solutions through interdisciplinary collaboration will further enhance precision medicine, optimizing outcomes for CRLM patients.

Funding: This work was supported by the Shanghai Natural Science Foundation Project (22ZR1413300), the National Natural Science Foundation of China (81874056, 81874182), the National Key Research and Development Plan of the Ministry of Science and Technology (2022YFE0125300), and the Public Health Bureau Foundation of Shanghai (202240240, 201940043).

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021; 71:209-249.
2. Ganesh K, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, Diaz LA Jr. Immunotherapy in colorectal cancer: Rationale, challenges and potential. Nat Rev Gastroenterol Hepatol. 2019; 16:361-375.
3. Chen Q, Chen J, Deng Y, Bi X, Zhao J, Zhou J, Huang Z, Cai J, Xing B, Li Y, Li K, Zhao H. Personalized prediction of postoperative complication and survival among colorectal liver metastases patients receiving simultaneous resection using machine learning approaches: A multi-center study. Cancer Lett. 2024; 593:216967.
4. Rompianesi G, Pegoraro F, Ceresa CD, Montalti R, Troisi RI. Artificial intelligence in the diagnosis and management of colorectal cancer liver metastases. World J Gastroenterol. 2022; 28:108-122.
5. Wu X, Li W, Tu H. Big data and artificial intelligence in cancer research. Trends Cancer. 2024; 10:147-160.
6. Joshi RM, Telang B, Soni G, Khalife A. Overview of perspectives on cancer, newer therapies, and future directions. Oncol Transl Med. 2024; 10:105-109.
7. Yu C, Shi Z, Zhou G, Chang X. Revisiting the survival paradox between stage IIB/C and IIIA colon cancer. Sci Rep. 2024; 14:22133.
8. Vermeulen PB, Colpaert C, Salgado R, Royers R, Hellemans H, Van Den Heuvel E, Goovaerts G, Dirix LY, Van Marck E. Liver metastases from colorectal adenocarcinomas grow in three patterns with different angiogenesis and desmoplasia. J Pathol. 2001; 195:336-342.
9. van Dam PJ, van der Stok EP, Teuwen LA, *et al*. International consensus guidelines for scoring the histopathological growth patterns of liver metastasis. Br J Cancer. 2017; 117:1427-1441.
10. Falcão D, Alexandrino H, Caetano Oliveira R, Martins J, Ferreira L, Martins R, Serôdio M, Martins M, Tralhão JG, Cipriano MA, Castro E Sousa F. Histopathologic patterns as markers of prognosis in patients undergoing hepatectomy for colorectal cancer liver metastases - Pushing growth as an independent risk factor for decreased survival. Eur J Surg Oncol. 2018; 44:1212-1219.
11. Galjart B, Nierop PMH, van der Stok EP, van den Braak RRJC, Höppener DJ, Daelemans S, Dirix LY, Verhoef C, Vermeulen PB, Grünhagen DJ. Angiogenic desmoplastic histopathological growth pattern as a prognostic marker of good outcome in patients with colorectal liver metastases. Angiogenesis. 2019; 22:355-368.
12. Takasu C, Morine Y, Yoshikawa K, Tokunaga T, Nishi M, Kashihara H, Wada Y, Yoshimoto T, Shimada M. Impact of pure desmoplastic histological growth patterns in colorectal liver metastasis. BMC Cancer. 2024; 24:1528.
13. Lazaris A, Amri A, Petrillo SK, Zoroquiain P, Ibrahim N, Salman A, Gao ZH, Vermeulen PB, Metrakos P. Vascularization of colorectal carcinoma liver metastasis: Insight into stratification of patients for anti-angiogenic therapies. J Pathol Clin Res. 2018; 4:184-192.
14. Cunningham JM, Kim CY, Christensen ER, Tester DJ, Parc Y, Burgart LJ, Halling KC, McDonnell SK, Schaid DJ, Walsh Vockley C, Kubly V, Nelson H, Michels VV, Thibodeau SN. The frequency of hereditary defective mismatch repair in a prospective series of unselected colorectal carcinomas. Am J Hum Genet. 2001; 69:780-790.
15. Venderbosch S, Nagtegaal ID, Maughan TS, *et al*. Mismatch repair status and BRAF mutation status in metastatic colorectal cancer patients: A pooled analysis of the CAIRO, CAIRO2, COIN, and FOCUS studies. Clin Cancer Res. 2014; 20:5322-5330.
16. Tran B, Kopetz S, Tie J, Gibbs P, Jiang ZQ, Lieu CH, Agarwal A, Maru DM, Sieber O, Desai J. Impact of BRAF mutation and microsatellite instability on the pattern of metastatic spread and prognosis in metastatic colorectal cancer. Cancer. 2011; 117:4623-4632.
17. Overman MJ, McDermott R, Leach JL, *et al*. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): An open-label, multicentre, phase 2 study. Lancet Oncol. 2017; 18:1182-1191.
18. Overman MJ, Lonardi S, Wong KYM, *et al*. Durable clinical benefit with nivolumab plus ipilimumab in DNA mismatch repair-deficient/microsatellite instability-high metastatic colorectal cancer. J Clin Oncol. 2018; 36:773-779.
19. André T, Shiu KK, Kim TW, *et al*. Pembrolizumab in microsatellite-instability-high advanced colorectal cancer. N Engl J Med. 2020; 383:2207-2218.
20. Bi F, Dong J, Jin C, *et al*. Iparomlimab (QL1604) in patients with microsatellite instability-high (MSI-H) or mismatch repair-deficient (dMMR) unresectable or metastatic solid tumors: A pivotal, single-arm, multicenter, phase II trial. J Hematol Oncol. 2024; 17:109.
21. Le DT, Uram JN, Wang H, *et al*. PD-1 blockade in tumors with mismatch-repair deficiency. N Engl J Med. 2015; 372:2509-2520.
22. Kawazoe A, Xu RH, García-Alfonso P, *et al*. Lenvatinib plus pembrolizumab versus standard of care for previously treated metastatic colorectal cancer: Final analysis of the randomized, open-label, Phase III LEAP-017 study. J Clin Oncol. 2024; 42:2918-2927.
23. AJCC Cancer Staging Handbook. *https://link.springer.*

com/book/9780387884424 (accessed February 1, 2025)
24. Sobin LH, Gospodarowicz MK, Wittekind C, eds. TNM Classification of Malignant Tumours. 8th ed. Wiley-Blackwell, Oxford, UK, 2017.
25. Weiser MR. AJCC 8th Edition: Colorectal Cancer. Ann Surg Oncol. 2018; 25:1454-1455.
26. Jin X, Wu Y, Feng Y, Lin Z, Zhang N, Yu B, Mao A, Zhang T, Zhu W, Wang L. A population-based predictive model identifying optimal candidates for primary and metastasis resection in patients with colorectal cancer with liver metastatic. Front Oncol. 2022; 12:899659.
27. Takeda Y, Mise Y, Matsumura M, Hasegawa K, Yoshimoto J, Imamura H, Noro T, Yamamoto J, Ishizuka N, Inoue Y, Ito H, Takahashi Y, Saiura A. Accuracy of modern clinical risk score including RAS status changes based on whether patients received perioperative chemotherapy for colorectal liver metastases. World J Surg. 2021; 45:2176-2184.
28. Brudvik KW, Jones RP, Giuliante F, *et al*. RAS mutation clinical risk score to predict survival after resection of colorectal liver metastases. Ann Surg. 2019; 269:120-126.
29. Katipally RR, Martinez CA, Pugh SA, Bridgewater JA, Primrose JN, Domingo E, Maughan TS, Talamonti MS, Posner MC, Weichselbaum RR, Pitroda SP; with the S:CORT Consortium. Integrated cinical-molecular classification of colorectal liver metastases: A biomarker analysis of the Phase 3 new EPOC randomized clinical trial. JAMA Oncol. 2023; 9:1245-1254.
30. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995; 20:273-297.
31. Sun T, Wang J, Li X, Lv P, Liu F, Luo Y, Gao Q, Zhu H, Guo X. Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set. Comput Methods Programs Biomed. 2013; 111:519-524.
32. Chu F, Wang L. Applications of support vector machines to cancer classification with microarray data. Int J Neural Syst. 2005; 15:475-484.
33. Breiman L. Random forests. Mach Learn. 2001; 45:5-32.
34. Zare A, Postovit LM, Githaka JM. Robust inflammatory breast cancer gene signature using nonparametric random forest analysis. Breast Cancer Res. 2021; 23:92.
35. Verschuur AVD, Hackeng WM, Westerbeke F, *et al*. DNA methylation profiling enables accurate classification of nonductal primary pancreatic neoplasms. Clin Gastroenterol Hepatol. 2024; 22:1245-1254.e10.
36. Cox D. Regression models and life-tables. J R Stat Soc Ser B-Stat Methodol. 1972; 34:187-220.
37. Tibshirani R. Regression shrinkage and selection *via* the lasso. J R Stat Soc Ser B (Methodol). 1996; 58:267-288.
38. Liu Z, Lin C, Suo C, Zhao R, Jin L, Zhang T, Chen X. Metabolic dysfunction-associated fatty liver disease and the risk of 24 specific cancers. Metabolism. 2022; 127:154955.
39. Li J, Wang Y, Song X, Xiao H. Adaptive multinomial regression with overlapping groups for multi-class classification of lung cancer. Comput Biol Med. 2018; 100:1-9.
40. Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Stat. 2001; 29:1189-1232.
41. Rodriguez LA, Schmittdiel JA, Liu L, Macdonald BA, Balasubramanian S, Chai KP, Seo SI, Mukhtar N, Levin TR, Saxena V. Hepatocellular carcinoma in metabolic dysfunction-associated steatotic liver disease. JAMA Netw Open. 2024; 7:e2421019.
42. Qi X, Wang S, Fang C, Jia J, Lin L, Yuan T. Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. Redox Biol. 2024; 79:103470.
43. Wang C, Long Y, Li W, Dai W, Xie S, Liu Y, Zhang Y, Liu M, Tian Y, Li Q, Duan Y. Exploratory study on classification of lung cancer subtypes through a combined K-nearest neighbor classifier in breathomics. Sci Rep. 2020; 10:5880.
44. Li Q, Hao C, Kang X, Zhang J, Sun X, Wang W, Zeng H. Colorectal cancer and colitis diagnosis using Fourier transform infrared spectroscopy and an improved k-nearest-neighbour classifier. Sensors (Basel). 2017; 17:2739.
45. Khan RA, Fu M, Burbridge B, Luo Y, Wu FX. A multi-modal deep neural network for multi-class liver cancer diagnosis. Neural Netw. 2023; 165:553-561.
46. Cho SI, Sun S, Mun JH, Kim C, Kim SY, Cho S, Youn SW, Kim HC, Chung JH. Dermatologist-level classification of malignant lip diseases using a deep convolutional neural network. Br J Dermatol. 2020; 182:1388-1394.
47. Chang X, Wang J, Zhang G, Yang M, Xi Y, Xi C, Chen G, Nie X, Meng B, Quan X. Predicting colorectal cancer microsatellite instability with a self-attention-enabled convolutional neural network. Cell Rep Med. 2023; 4:100914.
48. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A. 1982; 79:2554-2558.
49. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997; 9:1735-1780.
50. Yun G, Vyas K, Yang J, Yang GZ. Transfer recurrent feature learning for endomicroscopy image recognition. IEEE Trans Med Imaging. 2019; 38:791-801.
51. Cheng CL, Md Nasir ND, Ng GJZ, *et al*. Artificial intelligence modelling in differentiating core biopsies of fibroadenoma from phyllodes tumor. Lab Invest. 2022; 102:245-252.
52. Xin C, Liu Z, Zhao K, Miao L, Ma Y, Zhu X, Zhou Q, Wang S, Li L, Yang F, Xu S, Chen H. An improved transformer network for skin cancer classification. Comput Biol Med. 2022; 149:105939.
53. Xu H, Usuyama N, Bagga J, *et al*. A whole-slide foundation model for digital pathology from real-world data. Nature. 2024; 630:181-188.
54. Qian X, Pei J, Han C, *et al*. A multimodal machine learning model for the stratification of breast cancer risk. Nat Biomed Eng. 2025; 9:356-370.
55. Schirris Y, Gavves E, Nederlof I, Horlings HM, Teuwen J. DeepSMILE: Contrastive self-supervised pre-training benefits MSI and HRD classification directly from H&E whole-slide images in colorectal and breast cancer. Med Image Anal. 2022; 79:102464.
56. Zhang H, AbdulJabbar K, Grunewald T, *et al*. Self-supervised deep learning for highly efficient spatial immunophenotyping. EBioMedicine. 2023; 95:104769.
57. Tharmaseelan H, Vellala AK, Hertel A, Tollens F, Rotkopf LT, Rink J, Woźnicki P, Ayx I, Bartling S, Nörenberg D, Schoenberg SO, Froelich MF. Tumor classification of gastrointestinal liver metastases using CT-based radiomics and deep learning. Cancer Imaging. 2023; 23:95.
58. Jia W, Li F, Cui Y, Wang Y, Dai Z, Yan Q, Liu X, Li Y, Chang H, Zeng Q. Deep learning radiomics model of contrast-enhanced CT for differentiating the primary source of liver metastases. Acad Radiol. 2024; 31:4057-

4067.
59. Höppener DJ, Aswolinskiy W, Qian Z, Tellez D, Nierop PMH, Starmans M, Nagtegaal ID, Doukas M, de Wilt JHW, Grünhagen DJ, van der Laak JAWM, Vermeulen P, Ciompi F, Verhoef C. Classifying histopathological growth patterns for resected colorectal liver metastasis with a deep learning analysis. BJS Open. 2024; 8:zrae127.
60. Starmans MPA, Buisman FE, Renckens M, Willemssen FEJA, van der Voort SR, Groot Koerkamp B, Grünhagen DJ, Niessen WJ, Vermeulen PB, Verhoef C, Visser JJ, Klein S. Distinguishing pure histopathological growth patterns of colorectal liver metastases on CT using deep learning and radiomics: A pilot study. Clin Exp Metastasis. 2021; 38:483-494.
61. Wesdorp N, Zeeuw M, van der Meulen D, *et al*. Identifying genetic mutation status in patients with colorectal cancer liver metastases using radiomics-based machine-learning models. Cancers (Basel). 2023; 15:5648.
62. Granata V, Fusco R, Setola SV, Brunese MC, Di Mauro A, Avallone A, Ottaiano A, Normanno N, Petrillo A, Izzo F. Machine learning and radiomics analysis by computed tomography in colorectal liver metastases patients for RAS mutational status prediction. Radiol Med. 2024; 129:957-966.
63. Li M, Li X, Guo Y, Miao Z, Liu X, Guo S, Zhang H. Development and assessment of an individualized nomogram to predict colorectal cancer liver metastases. Quant Imaging Med Surg. 2020; 10:397-414.
64. Kim K, Kim S, Han K, Bae H, Shin J, Lim JS. Diagnostic performance of deep learning-based lesion detection algorithm in CT for detecting hepatic metastasis from colorectal cancer. Korean J Radiol. 2021; 22:912-921.
65. Yu Z, Li G, Xu W. Rapid detection of liver metastasis risk in colorectal cancer patients through blood test indicators. Front Oncol. 2024; 14:1460136.
66. Kiritani S, Yoshimura K, Arita J, Kokudo T, Hakoda H, Tanimoto M, Ishizawa T, Akamatsu N, Kaneko J, Takeda S, Hasegawa K. A new rapid diagnostic system with ambient mass spectrometry and machine learning for colorectal liver metastasis. BMC Cancer. 2021; 21:262.
67. Moosavi SH, Eide PW, Eilertsen IA, Brunsell TH, Berg KCG, Røsok BI, Brudvik KW, Bjørnbeth BA, Guren MG, Nesbakken A, Lothe RA, Sveen A. De novo transcriptomic subtyping of colorectal cancer liver metastases in the context of tumor heterogeneity. Genome Med. 2021; 13:143.
68. Nemlander E, Ewing M, Abedi E, Hasselström J, Sjövall A, Carlsson AC, Rosenblad A. A machine learning tool for identifying non-metastatic colorectal cancer in primary care. Eur J Cancer. 2023; 182:100-106.
69. Krishnan ST, Winkler D, Creek D, Anderson D, Kirana C, Maddern GJ, Fenix K, Hauben E, Rudd D, Voelcker NH. Staging of colorectal cancer using lipid biomarkers and machine learning. Metabolomics. 2023; 19:84.
70. Li J, Wang X, Cai L, Sun J, Yang Z, Liu W, Wang Z, Lv H. An interpretable deep learning framework for predicting liver metastases in postoperative colorectal cancer patients using natural language processing and clinical data integration. Cancer Med. 2023; 12:19337-19351.
71. Karagkounis G, Horvat N, Danilova S, *et al*. Computed tomography-based radiomics with machine learning outperforms radiologist assessment in estimating colorectal liver metastases pathologic response after chemotherapy. Ann Surg Oncol. 2024; 31:9196-9204.
72. Maaref A, Romero FP, Montagnon E, Cerny M, Nguyen B, Vandenbroucke F, Soucy G, Turcotte S, Tang A, Kadoury S. Predicting the response to FOLFOX-based chemotherapy regimen from untreated liver metastases on baseline CT: A deep neural network approach. J Digit Imaging. 2020; 33:937-945.
73. Davis JMK, Niazi MKK, Ricker AB, Tavolara TE, Robinson JN, Annanurov B, Smith K, Mantha R, Hwang J, Shrestha R, Iannitti DA, Martinie JB, Baker EH, Gurcan MN, Vrochides D. Predicting response to neoadjuvant chemotherapy for colorectal liver metastasis using deep learning on prechemotherapy cross-sectional imaging. J Surg Oncol. 2024; 130:93-101.
74. Qi W, Yang J, Zheng L, Lu Y, Liu R, Ju Y, Niu T, Wang D. CT-based radiomics for the identification of colorectal cancer liver metastases sensitive to first-line irinotecan-based chemotherapy. Med Phys. 2023; 50:2705-2714.
75. Lu L, Dercle L, Zhao B, Schwartz LH. Deep learning for the prediction of early on-treatment response in metastatic colorectal cancer from serial medical imaging. Nat Commun. 2021; 12:6654.
76. Endo Y, Alaimo L, Moazzam Z, *et al*. Optimal policy tree to assist in adjuvant therapy decision-making after resection of colorectal liver metastases. Surgery. 2024; 175:645-653.
77. Zhu HB, Xu D, Ye M, Sun L, Zhang XY, Li XT, Nie P, Xing BC, Sun YS. Deep learning-assisted magnetic resonance imaging prediction of tumor response to chemotherapy in patients with colorectal liver metastases. Int J Cancer. 2021; 148:1717-1730.
78. Giannini V, Rosati S, Defeudis A, *et al*. Radiomics predicts response of individual HER2-amplified colorectal cancer liver metastases in patients treated with HER2-targeted therapy. Int J Cancer. 2020; 147:3215-3223.
79. Wang Q, Nilsson H, Xu K, Wei X, Chen D, Zhao D, Hu X, Wang A, Bai G. Exploring tumor heterogeneity in colorectal liver metastases by imaging: Unsupervised machine learning of preoperative CT radiomics features for prognostic stratification. Eur J Radiol. 2024; 175:111459.
80. Paro A, Hyer MJ, Tsilimigras DI, Guglielmi A, Ruzzenente A, Alexandrescu S, Poultsides G, Aucejo F, Cloyd JM, Pawlik TM. Machine learning approach to stratifying prognosis relative to tumor burden after resection of colorectal liver metastases: An international cohort analysis. J Am Coll Surg. 2022; 234:504-513.
81. Lam CSN, Bharwani AA, Chan EHY, Chan VHY, Au HLH, Ho MK, Rashed S, Kwong BMH, Fang W, Ma KW, Lo CM, Cheung TT. A machine learning model for colorectal liver metastasis post-hepatectomy prognostications. Hepatobiliary Surg Nutr. 2023; 12:495-506.
82. Elforaici MEA, Montagnon E, Romero FP, Le WT, Azzi F, Trudel D, Nguyen B, Turcotte S, Tang A, Kadoury S. Semi-supervised ViT knowledge distillation network with style transfer normalization for colorectal liver metastases survival prediction. Med Image Anal. 2025; 99:103346.
83. Moro A, Mehta R, Tsilimigras DI, Sahara K, Paredes AZ, Bagante F, Guglielmi A, Alexandrescu S, Poultsides GA, Sasaki K, Aucejo FN, Pawlik TM. Prognostic factors differ according to KRAS mutational status: A classification and regression tree model to define prognostic groups after hepatectomy for colorectal liver metastasis. Surgery. 2020; 168:497-503.
84. Saber R, Henault D, Messaoudi N, Rebolledo R,

Montagnon E, Soucy G, Stagg J, Tang A, Turcotte S, Kadoury S. Radiomics using computed tomography to predict CD73 expression and prognosis of colorectal cancer liver metastases. J Transl Med. 2023; 21:507.
85. Zhou S, Sun D, Mao W, Liu Y, Cen W, Ye L, Liang F, Xu J, Shi H, Ji Y, Wang L, Chang W. Deep radiomics-based fusion model for prediction of bevacizumab treatment response and outcome in patients with colorectal cancer liver metastases: A multicentre cohort study. EClinicalMedicine. 2023; 65:102271.
86. Zhao QX, He XL, Wang K, Cheng ZG, Han ZY, Liu FY, Yu XL, Hui Z, Yu J, Chao A, Liang P. Deep learning model based on contrast-enhanced ultrasound for predicting early recurrence after thermal ablation of colorectal cancer liver metastasis. Eur Radiol. 2023; 33:1895-1905.
87. Luo X, Deng H, Xie F, Wang L, Liang J, Zhu X, Li T, Tang X, Liang W, Xiang Z, He J. Prognostication of colorectal cancer liver metastasis by CE-based radiomics and machine learning. Transl Oncol. 2024; 47:101997.
88. Amygdalos I, Müller-Franzes G, Bednarsch J, Czigany Z, Ulmer TF, Bruners P, Kuhl C, Neumann UP, Truhn D, Lang SA. Novel machine learning algorithm can identify patients at risk of poor overall survival following curative resection for colorectal liver metastases. J Hepatobiliary Pancreat Sci. 2023; 30:602-614.

Received February 2, 2025; Revised March 30, 2025; Accepted April 12, 2025.

[§]These authors contributed equally to this work.
*Address correspondence to:
Lu Wang and Weiping Zhu, Department of Hepatic Surgery, Fudan University Shanghai Cancer Center, Shanghai Medical College, Fudan University, Shanghai 200032, China.
E-mail: wangluzl@fudan.edu.cn (LW), wpzhush@hotmail.com (WZ)

Released online in J-STAGE as advance publication April 15, 2025.

Review

DOI: 10.5582/bst.2025.01046

Advances in research on receptor heterogeneity in breast cancer liver metastasis

Qinyu Liu[1,2,§], Runze Huang[1,2,§], Xin Jin[1,2], Xuanci Bai[3], Wei Tang[4,5], Lu Wang[1,2], Kenji Karako[4,*], Weiping Zhu[1,2,*]

[1] Department of Hepatic Surgery, Fudan University Shanghai Cancer Center, Shanghai Medical College, Fudan University, Shanghai, China;
[2] Department of Oncology, Shanghai Medical College Fudan University, Shanghai, China;
[3] Department of Clinical Medicine, Shanghai Medical College, Fudan University, Shanghai, China;
[4] Hepato-Biliary-Pancreatic Surgery Division, Department of Surgery, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan;
[5] National Center for Global Health and Medicine, Japan Institute for Health Security, Tokyo, Japan.

SUMMARY: Breast cancer liver metastasis (BCLM) presents a critical challenge in breast cancer treatment and has substantial epidemiological and clinical significance. Receptor status is pivotal in managing both primary breast cancer and its liver metastases. Moreover, shifts in these statuses can have a profound impact on patient treatment strategies and prognoses. Research has indicated that there is significant heterogeneity in receptor status between primary breast cancer and liver metastases. This variation may be influenced by a multitude of factors, such as therapeutic pressure, inherent tumor heterogeneity, clonal evolution, and the unique microenvironment of the liver. Changes in the receptor status of BCLM are crucial for adjusting treatment strategies, and liver biopsy plays an important role in the treatment process. Directions for future research targeting changes in receptor status include in-depth study of molecular mechanisms, combined treatment strategies for receptor status reversal, development of artificial intelligence deep learning models to predict receptor status in liver metastases, and clinical research on new drug development and combination therapies. That research will provide more precise treatment strategies for patients with BCLM and improve their prognosis.

Keywords: breast cancer liver metastasis, receptor heterogeneity, influencing factors, molecular mechanisms, treatment strategies

1. Introduction

According to the latest data from the International Agency for Research on Cancer (*1*), breast cancer has become the most commonly diagnosed cancer type in women worldwide, surpassing lung cancer. The latest statistics from the American Cancer Society show that the incidence of breast cancer has been continuously rising and those affected have become younger over the past decade (*2*). The survival rate of breast cancer varies depending on the stage at diagnosis, molecular subtypes, and other clinical pathological characteristics, with a 5-year relative survival rate of 99% for localized disease and only 32% for distant metastatic disease (*2*). Distant metastasis of breast cancer is the leading cause of death in patients with breast cancer. The liver ranks among the primary targets of breast cancer metastasis. In patients with advanced breast cancer, liver metastasis occurs in 20-30% of cases. This makes the liver the third most common site of distant metastasis, following bone and the lungs (*3,4*). Notably, breast cancer liver metastasis (BCLM) also tends to develop at a younger age, with a higher incidence of liver metastatic breast cancer in young women compared to older women (*5,6*). Thus, focusing on the prognosis for patients with BCLM is crucial.

Based on molecular biological characteristics, breast cancer can be classified into Luminal A, Luminal B, Triple-negative, HER2-positive, and HER2-negative types. The status of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) is crucial in guiding clinical treatment decisions (*7*). The American Society of Clinical Oncology (ASCO) and the European Society for Medical Oncology (ESMO) have recently updated their clinical practice guidelines for metastatic breast cancer. These updates emphasize receptor-dependent treatment strategies, similar to those used for in situ breast cancer. Specifically, they recommend formulating personalized treatment plans based on the receptor status

of liver metastases. This approach ensures that treatment is tailored to the specific characteristics of the metastatic disease, potentially improving outcomes for patients with BCLM (*8,9*).

Clinical research has demonstrated that there is a certain degree of expression discrepancy in ER, PR, and HER2 between primary breast cancer and liver metastases (*10-12*). Therefore, re-evaluating the receptor status of BCLM is crucial to formulating precise personalized treatment plans. The molecular mechanisms of and new drug targets associated with changes in the receptor status of BCLM need to be urgently examined. The current review meticulously synthesizes recent research findings to provide a summary of the situation and potential factors influencing receptor heterogeneity in BCLM. It also delves into the impact of these factors on the development of diagnostic and treatment strategies. Additionally, it explores and discusses promising directions for future research in this critical field in order to shed light on new avenues for advancing our understanding and management of this complex condition.

2. Heterogeneity of receptor status in BCLM

Clinical studies have shown that there is a significant degree of temporal and spatial heterogeneity in the expression of ER, PR, and HER2 during the process of breast cancer metastasis (*13-20*) (Tables 1 and 2). A study by Sundén *et al*. (10) on a cohort of 132 BCLM patients registered in two Swedish national cancer registries indicated that the discordance rates for ER, PR, and HER2 status between the primary tumor and liver metastasis were 17%, 33%, and 10%, respectively; among the cases with changes in receptor status, the proportion in which ER changed from positive to negative was 72.7%, and for PR it was 86.5%. Chen *et al*. (*21*) assessed a cohort of 390 paired primary and distant metastasis cases and found that the discordance rates for ER, PR, and HER2 between the primary and metastatic sites were 20%, 41.4%, and 14.1%, respectively; among all cases with receptor changes in breast cancer distant metastasis, the proportion in which ER changed from positive to negative was 85.9%, the proportion in which PR changed from positive to negative was 77.0%, and the proportion in which HER2 changed from positive to negative was 56.8%, but this study did not specify the individual cases in which each receptor changed from positive to negative. A meta-analysis performed by Schrijver *et al*. (*11*), which encompassed 39 studies, revealed notable discordance rates for ER, PR, and HER2 of 14.3%, 47.0%, and 12.1%, respectively, in BCLM. The researchers further observed that the random effect percentages for ER, PR, and HER2 changing from positive to negative were 22.5%, 49.4%, and 21.3%, respectively. Conversely, the percentages for these receptors changing from negative to positive were found to be 21.5%, 15.9%, and 9.5%, respectively. Together, the aforementioned studies demonstrate that among the receptors in BCLM, the discordance rate for PR is the highest, while that for HER2 is the lowest. Notably, a greater proportion of patients experience a change in ER and PR expression from positive to negative, as compared to those who undergo a change from negative to positive. In contrast, the proportion of patients whose HER2 status changes from positive to negative is relatively similar to those whose status changes from negative to positive.

Interestingly, almost all studies on changes in receptor status in BCLM have indicated that the discordance rate for HER2 is the lowest between the primary breast cancer and liver metastasis, but nearly one-third of patients with BCLM have their HER2 status change from no HER2 expression in the primary tumor to low HER2 expression in the liver metastasis (*22,23*). For example, a study by Almstedt *et al*. (*24*) showed that during the process of BCLM, the discordance rate for HER2 status was 40.9%, with 72.2% changing from no HER2 expression to low HER2 expression.

In addition, certain studies have indicated that alterations in HER2 status are intimately linked to the patient's ER status. Specifically, a HER2 status of 0 is predominantly associated with ER negativity, whereas low expression of HER2 tends to occur more frequently

Table 1. Breast cancer liver metastasis receptor status conversion

	Rate of discrepancy (%)(Event/Sample size)		
	ER	PR	HER2
Curigliano *et al*., 2011	14.5 (37/255)	48.6 (124/255)	14.0 (24/172)
Hoefnagel *et al*., 2012	12.7 (8/63)	41.3 (26/63)	9.5 (6/63)
Botteri *et al*., 2012	15.2 (15/99)	-	13.3 (8/60)
Nakamura *et al*., 2013	-	-	10.0 (2/20)
Woo *et al*., 2019	16.7 (4/24)	33.3 (8/24)	16.7 (4/24)
Chen *et al*., 2020	20.0 (16/80)	41.4 (29/70)	14.1 (10/71)
Sundén *et al*., 2023	16.9 (22/130)	32.5 (37/114)	9.9 (10/101)
Procházková *et al*., 2024	20.0 (2/10)	40.0 (4/10)	0 (0/10)

Abbreviations: ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor.

Table 2. Basic characteristics of patients

	Total number	Age(years)	Gender	Adjuvant endocrine therapy (BC)	Adjuvant chemotherapy (BC)	T stage of BC (%)					N stage of BC (%)				M stage of BC (%)	
						Tis	1	2	3	4	0	1	2	3	0	1
Curigliano *et al*., 2011	255	45 (26-75)	Female	69.8	51.7	-	46.7	42.5	Total 10.8		40.2	49.2	10.6		91.2	8.8
Hoefnagel *et al*., 2012	233	-	Female	-	-	-	-	-	-	-	-	-	-	-	-	-
Botteri *et al*., 2012 (biopsied)	100	-	Female	79.0	72.0	-	45.0	45.0	Total 5		-	-	-	-	-	-
Nakamura *et al*., 2013	156	-	Female	-	-	-	-	-	-	-	-	-	-	-	-	-
Woo *et al*., 2019	152	49 (26-83)	Female	49.3	70.4	-	24.3	41.4	17.1	17.1	25.0	33.6	21.7	19.7	90.8	9.2
Chen *et al*., 2020	348	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sundén *et al*., 2023	132	60 (27-84)	Female	69.8	62.4	7.9	30.2	49.2	9.5	3.2	67.7	29.1	1.6	1.6	100.0	0.0
Procházková *et al*., 2024	50	52 (26-78)	Female	-	-	4.0	40.0	48.0	2.0	6.0	52.0	30.0	10.0	8.0	98.0	2.0

Abbreviations: BC, primary breast cancer.

in tumors that are ER-positive (*22*).

3. Factors influencing receptor heterogeneity in BCLM

3.1. Selection pressure from treatment

A whole-exome sequencing analysis of primary tumors and matched metastases (*25*) revealed that untreated metastases typically originate from the main clone of the primary tumor, while treated metastases often harbor driver mutations specific to the metastasis, mainly due to the selection pressure of drug treatment that causes metastases to derive from rare clones in the primary tumor. Several studies (*26-28*) have shown that breast cancer patients who have undergone chemotherapy or endocrine therapy have a higher rate of changes in ER or PR status when they develop distant metastases compared to those who have not received drug treatment. Niikura *et al*. (*29*) investigated the relationship between HER2-targeted therapy and HER2 changes, and their results indicated that the inconsistency between the HER2 status in primary and metastatic lesions in breast cancer is related to whether the patient received chemotherapy. Zhao *et al*. (*30*) discovered a correlation between hormone receptor conversion in distant metastases of breast cancer and prior adjuvant endocrine therapy. Specifically, over 40% of patients who underwent adjuvant endocrine therapy experienced a loss of PR in the distant metastases of breast cancer. Additionally, more than 20% of patients who had previously received adjuvant endocrine therapy exhibited a loss of ER at the metastatic sites. In addition, the aforementioned study also found a positive correlation between adjuvant chemotherapy and the loss of PR at recurrence. These statistical results are similar to those of several previous statistics (*31-33*), suggesting that receptor heterogeneity in BCLM may be associated with the selection pressure of treatment.

3.2. Clonal evolution and tumor heterogeneity

Clonal evolution refers to the process in which some mutated subclones expand under the pressure of natural selection while others may perish as the tumor cell population evolves over time. Sprouffske *et al*. (*34*) confirmed the clonal evolution process of primary breast tumors in the development of metastatic dissemination. They achieved this by tracking genetic changes in breast cancer tumor xenograft models during metastasis. In addition, several studies have proposed that distinct tumor microenvironments can exert different selective pressures, thereby influencing tumor clonal evolution (*35,36*).

Tumor genetic heterogeneity refers to the diversity of genetic variations and gene expression patterns among different cells within a tumor during its development,

which may arise through complex genetic, epigenetic, and protein modifications. Genetic heterogeneity within tumors has been extensively documented, serving as a reflection of potential clonal evolution occurring within the tumor (*37-44*). A clinical study has shown that patients with high tumor heterogeneity are more likely to have adverse prognostic outcomes (*45*). In the progression of BCLM, the diversity in receptor expression status is indicative of the high degree of tumor heterogeneity present in the metastatic lesions. A study has indicated that there may be subclones in the primary breast tumor that cannot be detected by current technical means and that changes in receptor status occur during the spread to the liver due to various factors (*46*). Moreover, successful BCLM requires multiple steps (*47-50*), each of which can produce a population bottleneck, leading to differences in receptor status between the metastatic and primary lesions.

3.3. Influence of the metastatic microenvironment

Changes in the liver microenvironment may also affect receptor heterogeneity in BCLM. These changes, such as the presence of inflammatory responses and cytokines in BCLM, may influence the phenotype of breast cancer tumor cells, including receptor status. For example, studies have shown that inflammatory factors such as IL-6 may affect cell adhesion and the expression of E-cadherin, thereby influencing tumor metastasis and receptor status (*51,52*).

4. Impact of receptor heterogeneity in BCLM on treatment strategies

Zhao *et al*. (*30*) found that patients experiencing a change in hormone receptor status from negative to positive tend to have longer survival times than those with a persistently hormone receptor-negative status. Moreover, multivariate survival analysis has revealed that patients whose ER status changes from positive to negative face a significantly elevated risk of death compared to those with a stable ER-positive status. A large cohort study (*53*) indicated that patients with low HER2 expression have improved survival rates compared to those with no HER2 expression, regardless of ER status. This phenomenon is also reflected in other studies (*54,55*). Clearly, changes in receptor status during the progression of breast tumors have a significant impact on survival rates. Both the ASCO (*56*) and the ESMO (*57*) underscore the importance of basing treatment strategies for initially diagnosed BCLM on the ER, PR, and HER2 status of liver metastatic lesions. They also highlight the necessity of evaluating other treatment-related biomarkers in order to optimize therapeutic approaches. Therefore, evaluating the receptor status of BCLM is of great clinical significance to guiding the formulation of personalized treatment strategies.

The liver is one of the primary targets of distant metastasis in breast cancer cases. Unfortunately, patients with BCLM generally face a rather grim prognosis (*58*). Research by Botteri *et al*. (*15*) has shown that early BCLM patients (within 3 years) who undergo a liver biopsy have higher survival rates than those who do not. Compared to other target organs for distant metastasis of breast cancer, the liver is relatively accessible for biopsy. Thus, a comprehensive and timely assessment of the receptor status and related biomarkers of BCLM according to the latest clinical practice guidelines is crucial to guiding treatment decisions.

In response to changes in the receptor status of BCLM, the latest clinical practice guidelines state that classifying treatment based on molecular subtypes remains the general principle. A point worth highlighting is that nearly one-third of patients with BCLM exhibit a change in HER2 status, changing from no HER2 expression in the primary lesion to low HER2 expression in the liver metastasis. As low HER2-expressing breast cancer targets is researched further, this group of patients will become a potentially targetable population (*59*). Patients with low HER2-expressing BCLM also have new treatment options such as anti-HER2 antibody-drug conjugates (ADCs), and studies on the treatment of low HER2-expressing advanced breast cancer with the HER2 ADC drug T-DXd have become a focus of recent clinical research (*60*).

5. Future prospects

5.1. Molecular mechanisms of receptor heterogeneity in BCLM

In research on the molecular mechanisms of receptor heterogeneity in BCLM, the bidirectional crosstalk between ER and HER2 receptors has been widely reported in the context of endocrine or anti-HER2 treatment resistance in hormone receptor-positive and HER2-positive breast cancer (*61*). Studies have found that ER expression can modulate the activity of the PI3K pathway, thereby influencing the activation of the HER2 pathway. Conversely, HER2 overexpression, often driven by copy number amplification, can lead to the loss of *ER* gene expression. Moreover, multi-omics analysis of metastatic luminal-type primary breast tumors has shown that the transition from the luminal subtype to the HER2-enriched subtype is associated with the expression of ESR1, basal-like molecules, and the activation of related signaling pathways (*62-64*).

Nevertheless, the precise mechanisms driving the changes in receptor status between primary breast cancer and liver metastasis have yet to be fully understood. The precise molecular mechanisms involved in the process of breast cancer liver metastasis need to be explored further, and that effort will lay the foundation for the development of new treatment strategies.

5.2. Research on the reversal of receptor status in BCLM

Schade *et al*. (*65*) examined combined EZH2/AKT inhibitor therapy for triple-negative breast cancer and found that EZH2 and AKT inhibitors induce the expression of GATA3, promoting the transformation of triple-negative breast cancer from a basal-like state to a luminal-like state. Their findings indicate that the receptor status of breast cancer can be reversed under certain conditions, but whether the receptor status of liver metastases can be reversed and whether the specific mechanisms are consistent with those in the primary tumor require further research.

5.3. Artificial intelligence deep learning prediction models for receptor heterogeneity in BCLM

The advent of deep learning has driven the artificial intelligence (AI) revolution, increasing the use of AI in predictive modeling. Today, in relation to breast cancer, many AI models have been developed. For example, Bitencourt *et al*. used magnetic resonance imaging to assess *HER2* gene amplification and predict pathological response after neoadjuvant chemotherapy in HER2-positive breast cancer cases (*66*). Additionally, AI-driven digital pathology has demonstrated effectiveness in tumor diagnosis and treatment.

However, there is still a scarcity of AI models specifically tailored to BCLM. Current guidelines for BCLM typically recommend re-biopsy of liver metastases to re-evaluate their pathological status. Nevertheless, some patients with BCLM cannot tolerate punctures or surgical procedures. This hampers the accurate determination of the receptor status of liver metastases in those patients. Therefore, non-invasive methods of determining the receptor status of BCLM need to be urgently explored. The latest breakthroughs in deep learning technology allow algorithms to learn from clinical data to predict the receptor status of BCLM (*67,68*). On this basis, researchers can train AI models by collecting information on the primary lesion and liver metastasis of patients with BCLM to predict the receptor status of liver metastases and formulate personalized treatment plans based on the predicted receptor status (Figure 1).

5.4. New drug development and clinical evaluation

Considering the liver's pivotal role in detoxification and drug metabolism, a growing number of conventional therapeutics may rapidly lose their efficacy within the liver. Future research should therefore focus on developing new drugs that target molecular markers specific to BCLM, as well as optimizing drug delivery routes to the liver (*69,70*). Additionally, a study has found that a high proportion of ER and PR change from positive to negative in BCLM *(11*), that is, there is a high proportion of conversion from the luminal subtype to the

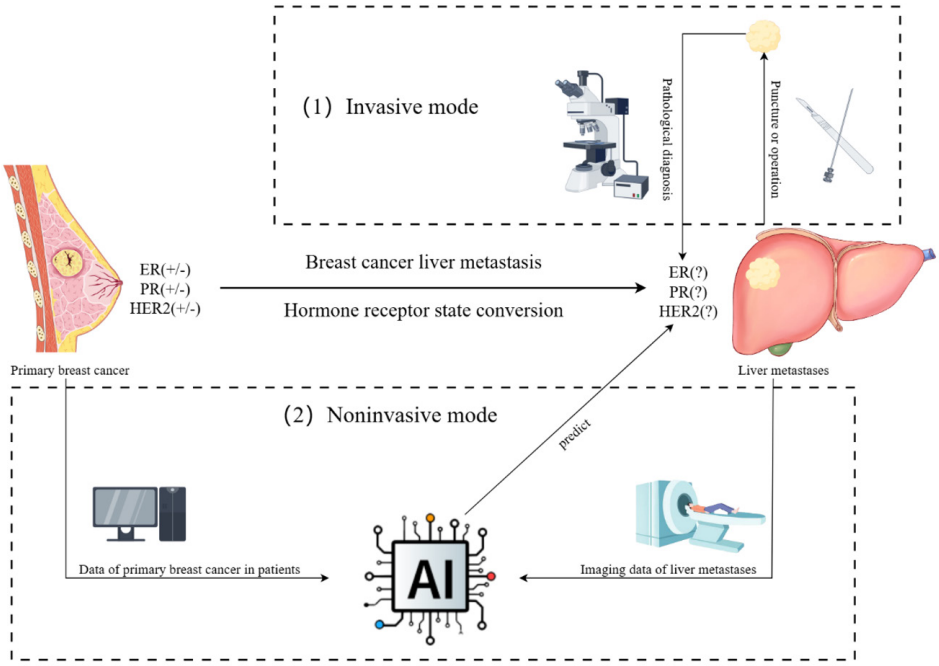


Figure 1. AI-assisted Framework for Predicting Hormone Receptor Status Conversion in Breast Cancer Liver Metastases. This figure illustrates the invasive and noninvasive methods for assessing hormone receptor (HR) status in breast cancer liver metastases. The invasive approach involves biopsy or surgical procedures to obtain pathological diagnoses of estrogen receptor (ER), progesterone receptor (PR), and HER2 status, capturing potential receptor conversions. In contrast, the noninvasive approach uses artificial intelligence to predict HR status changes using data from primary breast cancer and imaging of liver metastases, offering a less invasive alternative for clinical decision-making.

triple-negative subtype, and this has a negative impact on patient prognosis. Therefore, new drugs to reverse the triple-negative subtype of liver metastases to the luminal subtype could be explored and then used to treat those metastases based on ER and PR receptors. This approach has already yielded promising results in the treatment of primary breast cancer (*59*). However, whether it is equally applicable to the treatment of BCLM remains to be determined through large-scale clinical studies.

6. Conclusion

In summary, the changes in receptor status of BCLM represent a complex and pivotal clinical challenge. These changes not only influence the range of treatment options available to patients but also have a direct bearing on prognosis and survival rates. As we gain a better understanding of the molecular mechanisms underlying changes in receptor status and as AI technology is increasingly used in predictive modeling, we can anticipate the development of more precise and targeted treatment strategies.

Future research must concentrate on combination therapies aimed at reversing receptor status, the development of novel drugs, and large-scale clinical studies to assess the tangible impact of treatment modifications on patient survival. These efforts will pave the way for more personalized and effective treatment plans for individuals suffering from BCLM. Ultimately, this will lead to enhanced quality of life and improved survival rates for those patients. With ongoing advances in research, we eagerly anticipate further breakthroughs in the treatment of BCLM. Such progress holds the promise of bringing new hope and better outcomes to patients affected by this condition.

Funding: This work was jointly supported by the Public Health Bureau Foundation of Shanghai (202240240, 201940043), the National Natural Science Foundation of China (81874056, 81874182), the National Key Research and Development Plan of the Ministry of Science and Technology (2022YFE0125300), and the Shanghai Natural Science Foundation Project (22ZR1413300).

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024; 74:229-263.
2. Giaquinto AN, Sung H, Newman LA, Freedman RA, Smith RA, Star J, Jemal A, Siegel RL. Breast cancer statistics 2024. CA Cancer J Clin. 2024; 74:477-495.
3. Wang R, Zhu Y, Liu X, Liao X, He J, Niu L. The clinicopathological features and survival outcomes of patients with different metastatic sites in stage IV breast cancer. BMC Cancer. 2019; 19:1091.
4. Falck AK, Fernö M, Bendahl PO, Ryden L. St Gallen molecular subtypes in primary breast cancer and matched lymph node metastases - Aspects on distribution and prognosis for patients with luminal A tumours: Results from a prospective randomised trial. BMC Cancer. 2013; 13:558.
5. De Ridder J, De Wilt JHW, Simmer F, Overbeek L, Lemmens V, Nagtegaal I. Incidence and origin of histologically confirmed liver metastases: An explorative case-study of 23,154 patients. Oncotarget. 2016; 7:55368-55376.
6. Cummings MC, Simpson PT, Reid LE, Jayanthan J, Skerman J, Song S, McCart Reed AE, Kutasovic JR, Morey AL, Marquart L, O'Rourke P, Lakhani SR. Metastatic progression of breast cancer: insights from 50 years of autopsies. J Pathol. 2014; 232:23-31.
7. Van Poznak C, Harris LN, Somerfield MR. Use of biomarkers to guide decisions on systemic therapy for women with metastatic breast cancer: American Society of Clinical Oncology Clinical Practice Guideline. J Oncol Pract. 2015; 11:514-516.
8. Rugo HS, Rumble RB, Macrae E, *et al*. Endocrine therapy for hormone receptor-positive metastatic breast cancer: American Society of Clinical Oncology Guideline. *J Clin Oncol*. 2016; 34:3069-3103.
9. Cardoso F, Paluch-Shimon S, Senkus E, *et al*. 5th ESO-ESMO international consensus guidelines for advanced breast cancer (ABC 5). Ann Oncol. 2020; 31:1623-1649.
10. Sundén M, Norgren S, Lundqvist R, Andersson A, Sund M, Hemmingsson O. Receptor conversion and survival in breast cancer liver metastases. Breast Cancer Res. 2023; 25(1):105.
11. Schrijver WAME, Suijkerbuijk KPM, van Gils CH, van der Wall E, Moelans CB, van Diest PJ. Receptor conversion in distant breast cancer metastases: A systematic review and meta-analysis. J Natl Cancer Inst. 2018; 110:568-580.
12. Qi G, Zhang X, Gai X, Yan X. Retrospective analysis of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor-2 (HER2), Ki67 changes and their clinical significance between primary breast cancer and metastatic tumors. PeerJ. 2024; 12:e17377.
13. Wolff AC, Somerfield MR, Dowsett M, Hammond MEH, Hayes DF, McShane LM, Saphner TJ, Spears PA, Allison KH. Human epidermal growth factor receptor 2 testing in breast cancer: ASCO-College of American Pathologists Guideline Update. J Clin Oncol. 2023; 41:3867-3872.
14. Allison KH, Hammond MEH, Dowsett M, *et al*. Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP Guideline Update. J Clin Oncol. 2020; 38:1346-1366.
15. Botteri E, Disalvatore D, Curigliano G, Brollo J, Bagnardi V, Viale G, Orsi F, Goldhirsch A, Rotmensz N. Biopsy of liver metastasis for women with breast cancer: Impact on survival. Breast. 2012; 21:284-288.
16. Curigliano G, Bagnardi V, Viale G, Fumagalli L, Rotmensz N, Aurilio G, Locatelli M, Pruneri G, Giudici S, Bellomi M, Della Vigna P, Monfardini L, Orsi F, Nolè F, Munzone E, Goldhirsch A. Should liver metastases of breast cancer be biopsied to improve treatment choice? Ann Oncol. 2011; 22:2227-2233.
17. Woo JW, Chung YR, Ahn S, Kang E, Kim EK, Kim SH, Kim JH, Kim IA, Park SY. Changes in biomarker status

in metastatic breast cancer and their prognostic value. J Breast Cancer. 2019; 22:439-452.
18. Hoefnagel LD, Moelans CB, Meijer SL, van Slooten HJ, Wesseling P, Wesseling J, Westenend PJ, Bart J, Seldenrijk CA, Nagtegaal ID, Oudejans J, van der Valk P, van Gils CH, van der Wall E, van Diest PJ. Prognostic value of estrogen receptor α and progesterone receptor conversion in distant breast cancer metastases. Cancer. 2012; 118:4929-4935.
19. Nakamura R, Yamamoto N, Onai Y, Watanabe Y, Kawana H, Miyazaki M. Importance of confirming HER2 overexpression of recurrence lesion in breast cancer patients. Breast Cancer. 2013; 20(4):336-341.
20. Procházková K, Vojtíšek R, Vodička J, Horová J, Hošek P, Skála M, Šebek J, Dostál J, Přibáň V, Pivovarčíková K, Hes O, Třeška V, Moláček J. Hormone receptor conversion in metastatic breast cancer. Rep Pract Oncol Radiother. 202; 28:746-755.
21. Chen R, Qarmali M, Siegal GP, Wei S. Receptor conversion in metastatic breast cancer: Analysis of 390 cases from a single institution. Mod Pathol. 2020; 33:2499-2506.
22. Almstedt K, Krauthauser L, Kappenberg F, *et al*. Discordance of HER2-low between primary tumors and matched distant metastases in breast cancer. Cancers (Basel). 2023; 15:1413.
23. Boman C, Liu X, Eriksson Bergman L, Sun W, Tranchell C, Toli MA, Acs B, Bergh J, Foukakis T, Matikas A. A population-based study on trajectories of HER2 status during neoadjuvant chemotherapy for early breast cancer and metastatic progression. Br J Cancer. 2024; 131:718-728.
24. Modi S, Jacot W, Yamashita T, *et al*. Trastuzumab deruxtecan in previously treated HER2-low advanced breast cancer. N Engl J Med. 2022; 387:9-20.
25. Hu Z, Li Z, Ma Z, Curtis C. Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. *Nat Genet*. 2020; 52:701-708.
26. Curtit E, Nerich V, Mansi L, Chaigneau L, Cals L, Villanueva C, Bazan F, Montcuquet P, Meneveau N, Perrin S, Algros MP, Pivot X. Discordances in estrogen receptor status, progesterone receptor status, and HER2 status between primary breast cancer and metastasis. Oncologist. 2013; 18:667-674.
27. Ongaro E, Gerratana L, Cinausero M, Pelizzari G, Poletto E, Giangreco M, Andreetta C, Pizzolitto S, Di Loreto C, Minisini AM, Mansutti M, Russo S, Fasola G, Puglisi F. Comparison of primary breast cancer and paired metastases: Biomarkers discordance influence on outcome and therapy. Future Oncol. 2018; 14:849-859.
28. Yi ZB, Yu P, Zhang S, *et al*. Profile and outcome of receptor conversion in breast cancer metastases: A nation-wide multicenter epidemiological study. Int J Cancer. 2021; 148:692-701.
29. Okamura T, Kumaki N, Tsuda B, Niikura N. Stability of HER2 status by dual-color in situ hybridization before and after neoadjuvant chemotherapy in breast cancer. Tokai J Exp Clin Med. 2020; 45:176-181.
30. Zhao W, Sun L, Dong G, Wang X, Jia Y, Tong Z. Receptor conversion impacts outcomes of different molecular subtypes of primary breast cancer. Ther Adv Med Oncol. 2021; 13:17588359211012982.
31. Yang Z, Li N, Li X, Lei L, Wang X. The prognostic impact of hormonal receptor and HER-2 expression discordance in metastatic breast cancer patients. Onco Targets Ther. 2020; 13:853-863.
32. Lindström LS, Karlsson E, Wilking UM, Johansson U, Hartman J, Lidbrink EK, Hatschek T, Skoog L, Bergh J. Clinically used breast cancer markers such as estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 are unstable throughout tumor progression. J Clin Oncol. 2012; 30:2601-2608.
33. Ongaro E, Gerratana L, Cinausero M, Pelizzari G, Poletto E, Giangreco M, Andreetta C, Pizzolitto S, Di Loreto C, Minisini AM, Mansutti M, Russo S, Fasola G, Puglisi F. Comparison of primary breast cancer and paired metastases: biomarkers discordance influence on outcome and therapy. Future Oncol. 2018; 14:849-859.
34. Sprouffske K, Kerr G, Li C, Prahallad A, Rebmann R, Waehle V, Naumann U, Bitter H, Jensen MR, Hofmann F, Brachmann SM, Ferretti S, Kauffmann A. Genetic heterogeneity and clonal evolution during metastasis in breast cancer patient-derived tumor xenograft models. Comput Struct Biotechnol J. 2020; 18:323-331.
35. Fidler IJ. The pathogenesis of cancer metastasis: The 'seed and soil' hypothesis revisited. Nat Rev Cancer. 2003; 3:453-458.
36. Alzubi MA, Turner TH, Olex AL, Sohal SS, Tobin NP, Recio SG, Bergh J, Hatschek T, Parker JS, Sartorius CA, Perou CM, Dozmorov MG, Harrell JC. Separation of breast cancer and organ microenvironment transcriptomes in metastases. Breast Cancer Res. 2019; 21:36.
37. Navin N, Krasnitz A, Rodgers L, *et al*. Inferring tumor progression from genomic heterogeneity. Genome Res. 2010; 20:68-80.
38. Andor N, Graham TA, Jansen M, Xia LC, Aktipis CA, Petritsch C, Ji HP, Maley CC. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. Nat Med. 2016; 22:105-113.
39. Mroz EA, Tward AD, Hammon RJ, Ren Y, Rocco JW. Intra-tumor genetic heterogeneity and mortality in head and neck cancer: Analysis of data from the Cancer Genome Atlas. PLoS Med. 2015; 12:e1001786.
40. Jamal-Hanjani M, Wilson GA, McGranahan N, *et al*. Tracking the evolution of non-small-cell lung cancer. N Engl J Med. 2017; 376:2109-2121.
41. Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012; 481:306-313.
42. Sprouffske K, Merlo LM, Gerrish PJ, Maley CC, Sniegowski PD. Cancer in light of experimental evolution. Curr Biol. 2012; 22:R762-R771.
43. Aparicio S, Caldas C. The implications of clonal genome evolution for cancer medicine. N Engl J Med. 2013; 368:842-851.
44. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: Past, present, and the future. Cell. 2017; 168:613-628.
45. Ramón Y Cajal S, Sesé M, Capdevila C, Aasen T, De Mattos-Arruda L, Diaz-Cano SJ, Hernández-Losa J, Castellví J. Clinical implications of intratumor heterogeneity: Challenges and opportunities. J Mol Med (Berl). 2020; 98:161-177.
46. Aurilio G, Disalvatore D, Pruneri G, Bagnardi V, Viale G, Curigliano G, Adamoli L, Munzone E, Sciandivasci A, De Vita F, Goldhirsch A, Nolè F. A meta-analysis of oestrogen receptor, progesterone receptor and human epidermal growth factor receptor 2 discordance between primary breast cancer and metastases. Eur J Cancer. 2014; 50:277-289.

47. Nguyen DX, Bos PD, Massagué J. Metastasis: From dissemination to organ-specific colonization. Nat Rev Cancer. 2009; 9:274-284.
48. Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. Science. 2011; 331:1559-1564.
49. Massagué J, Obenauf AC. Metastatic colonization by circulating tumour cells. Nature. 2016; 529:298-306.
50. Welch DR, Hurst DR. Defining the hallmarks of metastasis. Cancer Res. 2019; 79:3011-3027.
51. Peiffer DS, Zhao F, Chen N, Hahn OM, Nanda R, Olopade OI, Huo D, Howard FM. Clinicopathologic characteristics and prognosis of ERBB2-low breast cancer among patients in the National Cancer Database. JAMA Oncol. 2023; 9:500-510.
52. Ma R, Feng Y, Lin S, Chen J, Lin H, Liang X, Zheng H, Cai X. Mechanisms involved in breast cancer liver metastasis. J Transl Med. 2015; 13:64.
53. Liu H, Li X, Li H, Feng L, Sun G, Sun G, Wu L, Hu Y, Liu L, Wang H. Potential molecular mechanisms and clinical progress in liver metastasis of breast cancer. Biomed Pharmacother. 2022; 149:112824.
54. Denkert C, Seither F, Schneeweiss A, *et al*. Clinical and molecular characteristics of HER2-low-positive breast cancer: Pooled analysis of individual patient data from four prospective, neoadjuvant clinical trials. Lancet Oncol. 2021; 22:1151-1161.
55. Molinelli C, Jacobs F, Agostinetto E, Nader-Marta G, Ceppi M, Bruzzone M, Blondeaux E, Schettini F, Prat A, Viale G, Del Mastro L, Lambertini M, de Azambuja E. Prognostic value of HER2-low status in breast cancer: A systematic review and meta-analysis. ESMO Open. 2023; 8:101592.
56. Van Poznak C, Somerfield MR, Bast RC, Cristofanilli M, Goetz MP, Gonzalez-Angulo AM, Hicks DG, Hill EG, Liu MC, Lucas W, Mayer IA, Mennel RG, Symmans WF, Hayes DF, Harris LN. Use of biomarkers to guide decisions on systemic therapy for women with metastatic breast cancer: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol. 2015; 33:2695-2704.
57. Gennari A, André F, Barrios CH, *et al*. ESMO Clinical Practice Guideline for the diagnosis, staging and treatment of patients with metastatic breast cancer. Ann Oncol. 2021; 32:1475-1495.
58. Ji L, Cheng L, Zhu X, Gao Y, Fan L, Wang Z. Risk and prognostic factors of breast cancer with liver metastases. BMC Cancer. 2021; 21:238.
59. Tarantino P, Viale G, Press MF, *et al*. ESMO expert consensus statements (ECS) on the definition, diagnosis, and management of HER2-low breast cancer. Ann Oncol. 2023; 34:645-659.
60. Bevers TB, Anderson BO, Bonaccio E, *et al*. NCCN clinical practice guidelines in oncology: Breast cancer screening and diagnosis. J Natl Compr Canc Netw. 2009; 7:1060-1096.
61. Giuliano M, Trivedi MV, Schiff R. Bidirectional crosstalk between the estrogen receptor and human epidermal growth factor receptor 2 signaling pathways in breast cancer: Molecular basis and clinical implications. Breast Care (Basel). 2013; 8:256-262.
62. Varešlija D, Priedigkeit N, Fagan A, *et al*. Transcriptome characterization of matched primary breast and brain metastatic tumors to detect novel actionable targets. J Natl Cancer Inst. 2019; 111:388-398.
63. Aftimos P, Oliveira M, Irrthum A, *et al*. Genomic and transcriptomic analyses of breast cancer primaries and matched metastases in AURORA, the Breast International Group (BIG) Molecular Screening Initiative. Cancer Discov. 2021; 11:2796-2811.
64. Garcia-Recio S, Thennavan A, East MP, *et al*. FGFR4 regulates tumor subtype differentiation in luminal breast cancer and metastatic disease. J Clin Invest. 2020; 130:4871-4887.
65. Schade AE, Perurena N, Yang Y, *et al*. AKT and EZH2 inhibitors kill TNBCs by hijacking mechanisms of involution. Nature. 2024; 635:755-763.
66. Bitencourt AGV, Gibbs P, Rossi Saccarelli C, Daimiel I, Lo Gullo R, Fox MJ, Thakur S, Pinker K, Morris EA, Morrow M, Jochelson MS. MRI-based machine learning radiomics can predict HER2 expression level and pathologic response after neoadjuvant therapy in HER2 overexpressing breast cancer. EBioMedicine. 2020; 61:103042.
67. Tan Y, Zhang WH, Huang Z, Tan QX, Zhang YM, Wei CY, Feng ZB. AI models predicting breast cancer distant metastasis using LightGBM with clinical blood markers and ultrasound maximum diameter. Sci Rep. 2024; 14:15561.
68. Goodin DA, Chau E, Zheng J, O'Connell C, Tiwari A, Xu Y, Niravath P, Chen SH, Godin B, Frieboes HB. Characterization of the breast cancer liver metastasis microenvironment *via* machine learning analysis of the primary tumor microenvironment. Cancer Res Commun. 2024; 4:2846-2857.
69. Rashid NS, Grible JM, Clevenger CV, Harrell JC. Breast cancer liver metastasis: Current and future treatment approaches. Clin Exp Metastasis. 2021; 38:263-277.
70. Dong M, Li C, Jiapaer P, Li X, Xue X, Jiang K, Sheng J. KIF15 expression characteristics: Relevance to neo-adjuvant chemotherapy efficacy in breast cancer. Oncol Transl Med. 2021; 7:69-75.

Received February 11, 2025; Revised March 31, 2025; Accepted April 12, 2025.

[§]These authors contributed equally to this work.
**Address correspondence to:*
Kenji Karako, Hepato-Biliary-Pancreatic Surgery Division, Department of Surgery, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan.
E-mail: tri.leafs@gmail.com

Weiping Zhu, Department of Hepatic Surgery, Fudan University Shanghai Cancer Center, Shanghai Medical College, Fudan University, Shanghai 200032, China.
E-mail: wpzhush@hotmail.com

Released online in J-STAGE as advance publication April 15, 2025.

Original Article

DOI: 10.5582/bst.2025.01013

Investigating perioperative pressure injuries and factors influencing them with imbalanced samples using a Synthetic Minority Over-sampling Technique

Yiwei Zhou[1,2,3], Jian Wu[1], Xin Xu[4], Guirong Shi[4], Ping Liu[4,*], Liping Jiang[4,*]

[1] Business School, University of Shanghai for Science and Technology, Shanghai, China;
[2] School of Intelligent Emergency Management, University of Shanghai for Science and Technology, Shanghai, China;
[3] Smart Urban Mobility Institute, University of Shanghai for Science and Technology, Shanghai, China;
[4] Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China.

SUMMARY: This study investigates the use of machine learning (ML) models combined with a Synthetic Minority Over-sampling Technique (SMOTE) and its variants to predict perioperative pressure injuries (PIs) in an imbalanced dataset. PIs are a significant healthcare problem, often leading to prolonged hospitalization and increased medical costs. Conventional risk assessment scales are limited in their ability to predict PIs accurately, prompting the exploration of ML techniques to address this challenge.We utilized data from 7,292 patients admitted to a tertiary care hospital in Shanghai between May 2017 and July 2023, with a final dataset of 2,972 patients, including 158 with PIs. Seven ML algorithms—Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Extra Trees (ET), K-Nearest Neighbors (KNN), and Decision Trees (DT)—were used in conjunction with SMOTE, SMOTE+ENN, Borderline-SMOTE, ADASYN, and GAN to balance the dataset and improve model performance.Results revealed significant improvements in model performance when SMOTE and its variants were used. For instance, the XGBoost model hadan AUC of 0.996 with SMOTE, compared to 0.800 on raw data. SMOTE+ENN and Borderline-SMOTE further enhanced the models' ability to identify minority classes. External validation indicatedthat XGBoost, RF, and ET exhibited the highest stability and accuracy, with XGBoost having an AUC of 0.977. SHAP analysis revealed that factors such as anesthesia grade, age, and serum albumin levels significantly influenced model predictions.In conclusion, integrating SMOTE with ML algorithms effectively addressed a data imbalance and improved the prediction of perioperative PIs. Future work should focus on refining SMOTE techniques and exploring their application to larger, multi-center datasets to enhance the generalizability of these findings, and especially for diseaseswith a lowincidence.

Keywords: machine learning, pressure injuries, SMOTE, predictive modelling, data imbalance

1. Introduction

Machine learning (ML) is playing an increasingly important role in data processing in clinical care, showing great potential for improving patient prognosis and optimizing healthcare management. However, a data imbalance is prevalent in clinical healthcare data, which is mainly evidentin the uneven distribution of case samples, where the number of samples for common diseases far exceeds that for rare diseases. This imbalance may lead to bias in the construction of ML models, making the models tend to predict categories with a higher frequency of occurrence while ignoring categories with smaller sample sizes. This bias not only affects the recall and accuracy of the model, but also limits the effective application of conventional classification algorithms in disease diagnosis, and especially in healthcare domains that require precise identification of a small number of cases (*1*).

An imbalance in clinical care data is a pervasive challenge in healthcare analytics, and especially for datasets with skewed class distributions. This issue presents significant hurdles in training classifiers for predictive modeling tasks, as highlighted by Kumar *et al.*(*2*). To address this, researchers have delved into a variety of solutions, with the Synthetic Minority Oversampling Technique (SMOTE) being a prominent one. SMOTE aims to enhance the predictive performance of ML models by rectifying class imbalances in clinical outcome prediction. In a pivotal study by Ishaq *et al.*,

the emphasis was on refining the survival prediction of heart failure patients through the useof SMOTE and sophisticated data mining. This studyuseda suite of nine classification models, underscoring the critical role of mitigating adata imbalanceto bolster the predictive accuracy of clinical datasets (*3*). Parallel to this, Goorbergh *et al.* conducted a case study on prediction modeling for ovarian cancer diagnosis. They scrutinized the repercussions of imbalance correction on the effectivenessof logistic regression models. Their findings underscored the potential detrimental effects of class imbalance corrections on risk prediction models, underscoring the necessity for judicious useof balancing techniques toanalyzeclinical data (*4*). Ridwan *et al*. usedML techniques and SMOTE to uncover patterns and risk factors within the Pima Indian diabetes dataset. By adeptly using SMOTE to detect diabetes, their study significantly advanced the scientific understandingof usingML algorithms to analyze clinical data (*5*). In a related vein, Javid *et al.*tackledthe issue of a class imbalance in the early diagnosis of Alzheimer's disease based on MRI images. They usedSMOTE to ensure an equitable distribution of samples across each class, thereby enhancing the performance of deep learning models in medical imaging data for disease diagnosis (*6*). This study highlightsthe significance of balancing techniques like SMOTE in increasing the effectiveness of deep learning models in this area.

In summary, a data imbalance is a pervasive challenge in ML, and particularly in healthcare datasets where minority class samples (*e.g*., patients with pressure injuries, or PIs) are often critical for accurate predictions. Various methods of data enhancement have been developed to address this issue, each with its own strengths and limitations. One of the most widely used techniques is SMOTE, which generates synthetic samples by interpolating between existing minority class samples. While SMOTE has shown significant success in improving model performance, other methods such as Adaptive Synthetic Sampling (ADASYN), Borderline-SMOTE, SMOTE+ENN (Edited Nearest Neighbors), and Generative Adversarial Networks(GAN) techniques have also emerged as promising alternatives. The current study aims to explore the effectiveness of these methods in predicting perioperative PIs and to compare their performance to conventional SMOTE.

PIs, also known as pressure ulcers, are a common and serious healthcare problem worldwide, leading to prolonged hospitalization, poor quality of life, increased mortality, and higher medical costs. They are defined as localized injuries to the skin and underlying tissue, usually over a bony prominence, resulting from pressure or a combination of pressure and shear(7). Given the preventable nature of these injuries, there is a critical need for effective early warning models to assist clinicians and nurses in making timely predictions and taking preventive action. SMOTE is used to enhance the predictive power of the model by balancing the class distribution by adding a few class samples to the training data, thus improving the model's prediction accuracy for PIs (*8,9*).

Conventional risk assessment scales (*e.g*., the Braden, Norton, and Waterlow scales) have been widely used but are limited in performance and are workload-intensive(*10*). As a result, artificial intelligence algorithms have been explored as they can capture patterns in complex data and have advantages in predicting time-to-event data, which is a common occurrence in clinical practice. ML models for predicting various medical outcomes, including PIs, have been developed by utilizing large datasets and algorithmic learning.

Nowadays, there are a growing number of instances whereMLis used in medicine, but the small amount of data has been a limitation in the aspects related to disease prediction, so the main aimof the current study was to evaluate the usefulness and effectiveness of the various resampling techniques in the prediction of PIs.Thegoal is to construct a ML model that can effectively predict PIs in emergency patients. To achieve this goal, seven ML algorithms were used in combination with the SMOTE algorithm and related methodsof extension to deal with the data imbalance problem.

2. Materials and Methods

2.1. Model selection

SMOTE is a technique for dealing with imbalanced datasets by generating synthetic samples of a few classes to balance the category distribution and was first proposed by Chawla *et al.* in 2002. This method creates new sample points by interpolating between the minority class samples and their k-nearest neighbors, thereby increasing sample diversity and reducing the risk of overfitting(*11*). In the current study, the SMOTE algorithm was used to enhance the model's ability to recognize the minority category (*i.e*., patients with PIs).

The basic steps of the underlying logic are as follows:

i) Select a minority sample X as the "root sample"for synthesizing a new sample.

ii) Find by Euclidean distance the k nearest neighboring samples (usually k is odd, *e.g.*, 5) of that sample, which also belong to the minority category. For two points X("x1,y1,z1,...") and O("x2,y2,z2,...") coordinates in n-dimensional space, the Euclidean distance d between them can be calculated withthe following formula.

$$d = \sqrt{(x2 - x1)^2 + (y2 - y1)^2 + (z2 - z1)^2 + \cdots} \quad (1)$$

iii) For each nearest-neighbor sample O, perform the following steps to generate a new sample point O_{new}. Calculate the root sample X and its nearest neighbor

samples O: dif = O -X; generate a random number between [0, 1] λ: and Use this formula to synthesize the value of each attribute of the new sample O_{new}.

$$O_{new} = O + \lambda \times (X - O)\,\lambda \quad (2)$$

iv) Repeat step 3 to produce the required number of new samples.

The key to the SMOTE algorithm is that instead of simply copying existing minority class samples, it creates new sample points by interpolating between the minority class samples, which increases the diversity of the samples and reduces the risk of overfitting. This approach is particularly useful in situations where the number of minority samples is small but each sample is important.A basic diagram of SMOTE is shown in Figure 1.

Similar to SMOTE, ADASYN generates synthetic samples but focuses more on the difficult-to-learn regions of the minority class, potentially improving model performance(*12*).

SMOTE+ENN is a hybrid technique that combines SMOTE with the ENN technique to efficiently deal with imbalanced datasets. First, a large amount of oversampled data is generated using the SMOTE method described above, and then ENN is used to clean the dataset by removing noisy samples, ENN works by identifying samples whose nearest neighbors belong to a different class and removing them. This helps toreduce noise and improve the quality of the dataset. This approach helps reduce overfitting and enhances the model's generalization ability (*13*).

Borderline-SMOTE is an enhanced version of SMOTE that generates synthetic samples specifically from minority class samples near the decision boundary to improve classification performance by targeting the most informative samples. Minority samples are categorized into three types: Safe (surrounded mostly by minority class samples), Danger (surrounded mostly by majority class samples and considered to be on the decision boundary), and Noise (surrounded entirely by majority class samples). Only Danger samples are used to create synthetic samples by selecting neighboring minority samples and interpolating between them using the same formula as SMOTE. There are two variants: Borderline-SMOTE1 generates synthetic samples using only minority class neighbors, whereas Borderline-SMOTE2 uses any neighbor (regardless of class) to introduce more diversity. The key advantage of Borderline-SMOTE is its focus on the decision boundary, which reduces the risk of generating noisy samples and enhances the effectiveness of synthetic samples (*14*).

GANs are advanced generative models that use a generator network to create synthetic samples and a discriminator network to distinguish between real and synthetic samples. GANs can produce high-quality synthetic data, potentially improving model performance by increasing the diversity of the minority class(*15*).

The current study used seven different ML models to predict PIs in emergency patients, each of which has its own unique strengths that make them perform well when dealing with specific types of data and problems.Support Vector Machine (SVM) is effective in dealing with high-dimensional spatial data and non-linear problems, being able to find hyperplanes that maximize the class interval. In PI prediction, SVM can help identify complex patterns, and especially when the feature space is large(*16*).Random Forest (RF), as an integrated learning method, improves the stability and accuracy of the model by constructing multiple decision trees and is very resistant to overfitting. When faced with imbalanced datasets, RF provides robust predictions and reduces the variance of predictions by integrating multiple models(*17*).Extreme Gradient Boosting (XGBoost) is an efficient gradient boosting framework that is capable of handling large-scale datasets and typically

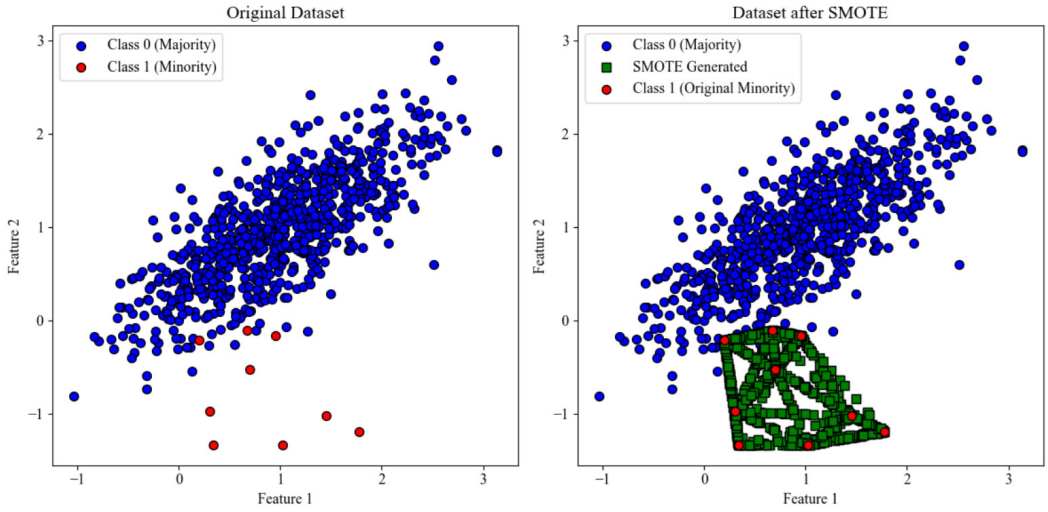


Figure 1. The basic working principle of SMOTE. Modelling SMOTE workings using randomly generated data.

outperforms conventional gradient boosting methods in terms of prediction performance. XGBoost performs well when dealing with datasets with a large number of features, which makes it suitable for extraction of key information from a large amount of patient datato predict PIs(*18*).Extra Trees (ET) is able to effectively deal with non-linear relationships and imbalanced datasets and improve the recognition of a few classes through its high stochasticity and integrated learning(*17*). K-Nearest Neighbors (KNN) is a simple instance-based learning algorithm that does not require a training phase and can directly use training data for prediction. KNN performs well on small datasets, so it is suitable when the sample size is not particularly large, as in the current study, and especially when SMOTE processingis used (*19*).Logistic regression (LR) is a linear model that is suitable for binary classification problems and can provide a probabilistic interpretation of the prediction results. When predicting PIs, LR can provide a direct interpretation of the probability of a patient developing a pressure injury, which is useful for clinical decision-making(*20*). Decision trees (DTs) are intuitive models that are easy to understand and interpret and can clearly demonstrate the relationship between features and target variables. DTs can help to identify the most important risk factors and can be used as a baseline for comparison to more complex models(*17*).

In the current study, the main challenge faced was the problem of a data imbalance, *i.e.*, the number of patients with PIs (positive sample) was much smaller than the number of patients without PIs (negative sample). To address this issue, the SMOTE algorithm was used to balance the dataset and the seven ML models described earlier were usedto construct predictive models. These models were chosen based on their extensive use and history of success in dealing with imbalanced datasets, handling high-dimensional data, providing predictive explanations, and in medical predictive modelling. Comparing the performance of these models enables the identification of the most appropriate model for the currentdata and problem, thereby improving the accuracy and reliability of predicting PIs. In addition, the diversity of these models allows evaluationand validation of predictions from different perspectives, ensuring that the findings are robust and reliable.

2.2. Participants

Data from a total of 7,292 patients consisting of7,171 indicators were selected from all recorded inpatient data ata tertiary care hospital in Shanghai during the period from May 2017 to July 2023 (numerous interfering items in data during the COVID-19 epidemic were not selected), and a total of 549 patients with PIs (7.53%) served as the initial screening subjects. After data processing, data from the remaining 2,972 patientsserved as the final data for this study and included 158patients with PIs (5.32%).

2.3. Data preprocessing

When dealing with the huge number of 7,171 feature variables, the XGBoost model was usedto identify the features that contribute most to the model performance. The advantage of XGBoost is that it is able to filter the features efficiently when there are missing values in the data, enabling the initialselectionof the top 32 feature variables that have the greatest impact on the model. Through further in-depth analyses, those features that were not strongly associated with PIswere eliminated and 27 key feature variables were ultimately selected, laying a solid foundation for building an accurate prediction model. In order to maintain the high quality of the dataset and reduce the noise interference in model training, a key decision was made to eliminate sets of data with more than 8 missing values among the 27 key feature variables. This strategy helps to maintain the integrity of the dataset while avoiding the uncertainty introduced by too many missing values, ensuring the reliability of the data and the stability of model training. After completing the screening of feature variables and the reduction of the dataset, in-depth data preprocessing was performedon the remaining data. This includes meticulous treatment of missing values, outliers, and duplicate records, steps that are critical to ensuring the quality of the data and the smooth running of subsequent experiments.

Data preprocessing consisted mainly of the following: *i*) Categorical variables. Missing values for characteristic variables in the data involving categorical variables are uniformly filled in using plurality in the current study; *ii*) Continuous variables. Missing values for continuous variables in this study were filled in using the mean of the age groups. Age groups were every 10 years, and 0-9 and10-19 were each averaged and populated within their age range.

These comprehensive data preprocessing measures ensured the cleanliness and consistency of the dataset, providing a solid data foundation for subsequent model training and analysis.

2.4. Evaluation metrics

In the model training phase, the datasetwas divided into training and validation sets at a ratio of 7:3, and multiple ML models were used to predict whether PIs occurred in emergency patients. In the model evaluation phase, two key evaluation metrics were used: the Confusion Matrix and ROC_AUC.

ROC curveswere also plotted and AUC values were calculated for each model; ROC curves demonstrate the model's performance under different thresholds, while AUC values quantify the model's ability to distinguish between positive and negative categories, with higher AUC values indicating better classification performance.

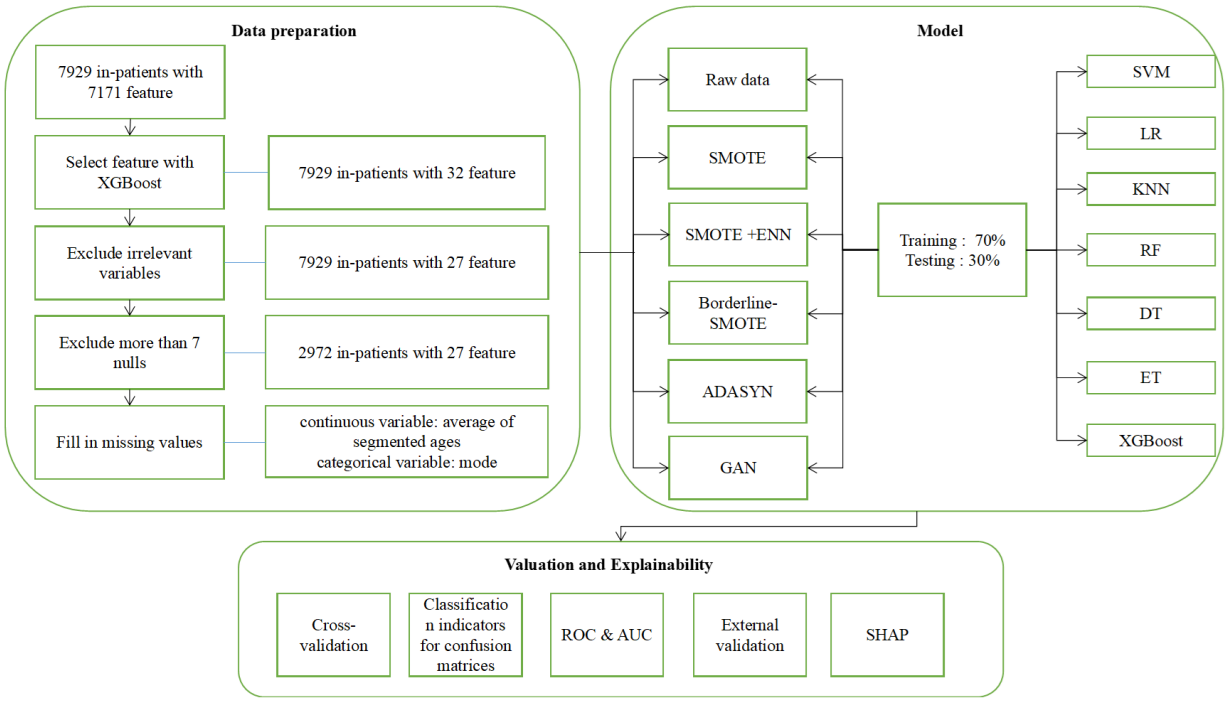


Figure 2. Flowchart for this study.

Finally, the confusion matrices and AUC values of the different modelswere compared to determine which model performed best in predicting PIs. This comprehensive assessment approach alloweda full understanding and comparison of the performance of each model in order to select the most appropriate model to aid inclinical decision-making.

2.5. Experimental design

The flow of this study is shown in Figure 2.

3. Results

3.1. Participants' characteristics

The distribution and comparison of several basic characteristics of patients with PIs (PI) and patients without PIs (Non-PI) is shown in Table 1. The table lists characteristics including sex, age, hypertension (HTN), hyperlipidemia, diabetes mellitus (DM), cardiovascular disease (CVD), history of malignancy, smoking status, drinking status, body temperature, pulse rate (PR), respiratory rate (RR), diastolic blood pressure (DBP), systolic blood pressure (SBP), body mass index (BMI), serum albumin, operating time, intra-operative blood transfusion, intra-operative hypotension (IH), surgical position (thisrefers to the specific position of the patient during surgery. 1-3 are supine, prone, and lateral positions, respectively. 0 is an undefined position), surgical dressing, dressing site, anesthesia grade, method of anesthesia, oxygen saturation (SpO_2), self-care competency grade, and blood glucose (BG). Categorical variables are expressed as the number (percentage) andcontinuous variables are expressed asthe mean (range). Comparison of these variables revealed significant differences in these characteristics between the two groups, with variables such as age, pulse rate, body mass index and method of anesthesia differing significantly between the two groups while variables such as sex, hypertension, and diabetes mellitus did not.

Given that the original dataset is multidimensional, visually depicting the newly generated positive samples presents a challenge. To overcome this, all variables wereprojected onto a single axis, thereby facilitating a clear visualization of the samples created by the SMOTE algorithm. For further details, refer to Figure 3.
Data after different methods of enhancement are shown in Tables 2-6.

3.2. Comparison of ML-based models

In this study, the confusion matrix of the model after using SMOTE and its variants revealed significant improvements as shown in Table 7. For example, the SMOTE-enhanced XGBoost model hadextremely high TP and TN values in internal validation while minimizing FP and FN values, indicating that the model performed well in identifying a small class of samples (patients with PIs). In addition, methods such as SMOTE+ENN and Borderline-SMOTE, although slightly inferior to SMOTE in some models, further improved the model's ability to identify minority classes by optimizing the sample quality or focusing on the borderline region.

Table 1. Basic characteristics of patients

Variables	Non-PI (*n*=2,814)	PI (*n*=158)	*p* value
Sex, *n* (%)	1,614 (57.4%)	69 (43.6%)	0.625
Age, years	54 [0-98]	74 [0-94]	< 0.001
Hypertension (HTN), *n* (%)	420 (14.9%)	26 (17.5%)	0.105
Hyperlipidemia, n (%)	12 (0.05%)	0 (0%)	0.445
Diabetes Mellitus (DM), *n* (%)	114 (4.1%)	14 (8.9%)	0.604
Cardiovascular Disease (CVD), *n* (%)	4 (0.2%)	0 (0%)	0.700
History of Malignant Tumor, *n* (%)	116 (4.1%)	14 (8.9%)	0.460
Smoking Status, *n* (%)	280 (10.0%)	13 (8.2%)	0.089
Alcohol Consumption Status, *n* (%)	2 (0.1%)	1 (0.6%)	0.709
Body Temperature, °C	36.6 [35.3-40.7]	36.6 [35.2-40.3]	0.684
Pulse Rate (PR), bpm	88 [38-198]	84 [52-165]	0.003
Respiratory Rate (RR), bpm	20 [10-78]	18 [12-33]	0.187
Diastolic Blood Pressure (DBP), mmHg	75 [24-152]	81 [54-134]	0.291
Systolic Blood Pressure (SBP), mmHg	124 [47-277]	135 [100-195]	0.278
Body Mass Index (BMI), kg/m²	21.7 [5.4-49]	22.7 [12.8-41.6]	< 0.001
Serum Albumin, g/L	36.7 [16.5-59]	32.9 [18.4-45.7]	0.488
Operating Time, h	2.90 [0.15-7.29]	2.82 [0.25-10.02]	0.586
Intraoperative Blood Transfusion, *n* (%)	262 (9.3%)	12 (7.6%)	0.693
Intraoperative Hypotension (IH), *n* (%)	246 (8.7%)	6 (3.8%)	0.912
Surgical Position	0.8 [0-3]	0.4 [0-3]	0.532
Surgical Dressing, *n* (%)	2,351 (83.5%)	125 (79.1%)	0.746
Dressing Site, *n* (%)	464 (16.5%)	33 (20.9%)	0.823
Anesthesia Grade	2.3 [0-5]	1.8 [0-5]	0.011
Anesthesia Method, *n* (%)	2,678 (95.2%)	26 (17.5%)	< 0.001
Oxygen Saturation (SpO2), %	97 [65-100]	96 [47-100]	0.567
Self-Care Ability Grade	2.8 [1-3]	2.9 [2-3]	0.006
Blood Glucose (BG)	7.9 [1.4-28.0]	8.0 [3.8-17.6]	0.243

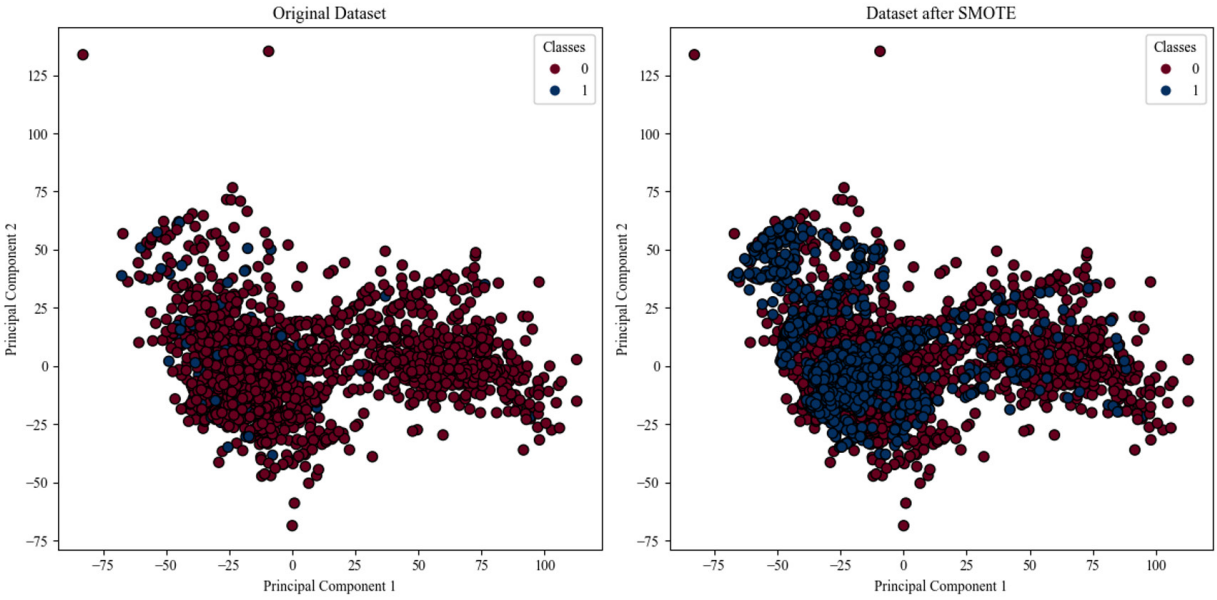


Figure 3. Status of data generated by SMOTE.

ADASYN and GAN also showed good performance in the confusion matrix, although they may face some challenges withhigh-dimensional data.

The analysis of the confusion matrix allows a more intuitive view of the impact of different methods of data enhancement on model performance. For example, SMOTE results in ahigh recall and precision in most models, while SMOTE+ENN performs well in removing noise, albeit possibly at the expense of some sample diversity.Borderline-SMOTE and ADASYN, in contrast, display better recognition of minority classes in specific models, although they have limited overall performance gains.GAN generated high-quality minority class samples, but its generated samples may be too close to the original samples, leading to an increased risk of overfitting.

As can be seen from Table 8, the performance metrics (*e.g.*, precision, recall, F1 score, accuracy, and AUC) of

Table 2. Supplementary data-enhanced dataset - SMOTE

Variables	Non-PI (*n*=2,252)	PI (*n*=2,252)	*p* value
Sex, *n* (%)	1,329 (59.0%)	821 (36.5%)	< 0.001
Age, years	53 [0-98]	74 [0-94]	< 0.001
Hypertension (HTN), *n* (%)	329 (14.6%)	778 (34.5%)	< 0.001
Hyperlipidemia, *n* (%)	11 (0.5%)	0 (0%)	0.011
Diabetes Mellitus (DM), *n* (%)	91 (4.0%)	39 (1.7%)	< 0.001
Cardiovascular Disease (CVD), *n* (%)	4 (0.2%)	0 (0%)	0.153
History of Malignant Tumor, *n* (%)	127 (5.6%)	22 (1.0%)	< 0.001
Smoking Status, *n* (%)	230 (10.2%)	32 (1.4%)	< 0.001
Alcohol Consumption Status, *n* (%)	2 (0.1%)	0 (0%)	0.515
Body Temperature, °C	36.6 [35.3-40.7]	36.5 [36.0-40.3]	0.247
Pulse Rate (PR), bpm	88 [18-198]	83 [52-165]	< 0.001
Respiratory Rate (RR), bpm	20 [10-78]	18 [12-33]	< 0.001
Diastolic Blood Pressure (DBP), mmHg	75 [24-152]	81 [56-134]	< 0.001
Systolic Blood Pressure (SBP), mmHg	124 [53-277]	135 [50-195]	0.604
Body Mass Index (BMI), kg/m²	21.6 [7.2-49]	22.7 [12.8-41.6]	< 0.001
Serum Albumin, g/L	36.8 [16.5-53.9]	33.2 [18.4-45.7]	0.886
Operating Time, h	2.89 [0.15-6.84]	2.76 [0.25-10.02]	< 0.001
Intraoperative Blood Transfusion, *n* (%)	222 (9.9%)	389 (17.3%)	< 0.001
Intraoperative Hypotension (IH), *n* (%)	198 (8.8%)	38 (1.7%)	0.017
Surgical Position	0.8 [0-3]	0.4 [0-3]	< 0.001
Surgical Dressing, *n* (%)	1,888 (83.8%)	1,315 (58.4%)	< 0.001
Dressing Site, *n* (%)	364 (16.2%)	159 (7.1%)	< 0.001
Anesthesia Grade	2.8 [1-5]	3.2 [2-5]	< 0.001
Anesthesia Method, *n* (%)	2,145 (95.2%)	1,519 (67.5%)	< 0.001
Oxygen Saturation (SpO2), %	97 [76-100]	97 [65-100]	0.003
Self-Care Ability Grade	2.9 [1-3]	2.9 [2-3]	< 0.001
Blood Glucose (BG)	7.9 [1.4-28.0]	7.9 [3.8-13.4]	0.575

Table 3. Supplementary data-enhanced dataset - ADASYN

Variables	Non-PI (*n*=2,252)	PI (*n*=2,252)	*p* value
Sex, *n* (%)	1,311 (58.2%)	733 (32.5%)	< 0.001
Age, years	54 [0-98]	73 [0-94]	< 0.001
Hypertension (HTN), *n* (%)	346 (15.3%)	64 (2.8%)	< 0.001
Hyperlipidemia, *n* (%)	9 (0.4%)	0 (0%)	0.101
Diabetes Mellitus (DM), *n* (%)	96 (4.3%)	36 (1.6%)	0.109
Cardiovascular Disease (CVD), *n* (%)	3 (0.1%)	0 (0%)	0.785
History of Malignant Tumor, *n* (%)	138 (6.1%)	34 (1.5%)	< 0.001
Smoking Status, *n* (%)	223 (9.9%)	33 (1.5%)	< 0.001
Alcohol Consumption Status, *n* (%)	1 (0.1%)	0 (0%)	0.305
Body Temperature, °C	36.6 [35.9-40.7]	36.6 [35.2-40.3]	0.079
Pulse Rate (PR), bpm	88 [18-198]	84 [52-165]	< 0.001
Respiratory Rate (RR), bpm	20 [10-78]	18 [12-33]	< 0.001
Diastolic Blood Pressure (DBP), mmHg	75 [24-152]	81 [54-134]	0.001
Systolic Blood Pressure (SBP), mmHg	124 [47-277]	134 [50-195]	0.987
Body Mass Index (BMI), kg/m²	21.7 [5.4-49]	22.7 [12.8-41.6]	< 0.001
Serum Albumin, g/L	36.8 [16.5-59.0]	33.0 [18.4-45.7]	0.119
Operating Time, h	2.89 [0.17-6.84]	2.80 [0.25-10.02]	< 0.001
Intraoperative Blood Transfusion, *n* (%)	206 (9.1%)	41 (18.2%)	< 0.001
Intraoperative Hypotension (IH), *n* (%)	209 (9.3%)	2 (0.1%)	< 0.001
Surgical Position	0.8 [0-3]	0.4 [0-3]	< 0.001
Surgical Dressing, *n* (%)	1,876 (83.3%)	1,433 (63.6%)	< 0.001
Dressing Site, *n* (%)	376 (16.7%)	95 (4.2%)	< 0.001
Anesthesia Grade	2.8 [1-5]	3.2 [2-5]	< 0.001
Anesthesia Method, *n* (%)	2,143 (95.2%)	1,660 (73.7%)	< 0.001
Oxygen Saturation (SpO2), %	97 [71-100]	97 [65-100]	0.002
Self-Care Ability Grade	2.9 [1-3]	2.9 [2-3]	< 0.001
Blood Glucose (BG)	7.9 [1.4-23.0]	8.0 [3.8-17.6]	0.007

Table 4. Supplementary data-enhanced dataset - GAN

Variables	Non-PI (*n*=2,252)	PI (*n*=2,252)	*p* value
Sex, *n* (%)	1,290 (59.0%)	821 (36.5%)	0.715
Age, years	55 [0-98]	74 [0-94]	< 0.001
Hypertension (HTN), *n* (%)	329 (14.6%)	778 (34.5%)	0.077
Hyperlipidemia, *n* (%)	11 (0.5%)	0 (0%)	0.396
Diabetes Mellitus (DM), *n* (%)	91 (4.0%)	39 (1.7%)	0.023
Cardiovascular Disease (CVD), *n* (%)	4 (0.2%)	0 (0%)	0.500
History of Malignant Tumor, *n* (%)	127 (5.6%)	22 (1.0%)	0.302
Smoking Status, *n* (%)	230 (10.2%)	32 (1.4%)	0.173
Alcohol Consumption Status, *n* (%)	2 (0.1%)	0 (0%)	0.166
Body Temperature, °C	36.6 [35.3-40.7]	36.5 [36.0-40.3]	0.687
Pulse Rate (PR), bpm	88 [18-198]	83 [52-165]	< 0.001
Respiratory Rate (RR), bpm	20 [10-78]	18 [12-33]	0.199
Diastolic Blood Pressure (DBP), mmHg	75 [24-152]	81 [56-134]	0.412
Systolic Blood Pressure (SBP), mmHg	124 [53-277]	135 [50-195]	0.322
Body Mass Index (BMI), kg/m²	21.6 [7.2-49]	22.7 [12.8-41.6]	< 0.001
Serum Albumin, g/L	36.8 [16.5-53.9]	33.2 [18.4-45.7]	0.649
Operating Time, h	2.89 [0.15-6.84]	2.76 [0.25-10.02]	0.026
Intraoperative Blood Transfusion, *n* (%)	222 (9.9%)	389 (17.3%)	0.919
Intraoperative Hypotension (IH), *n* (%)	198 (8.8%)	38 (1.7%)	0.371
Surgical Position	0.8 [0-3]	0.4 [0-3]	0.121
Surgical Dressing, *n* (%)	1,888 (83.8%)	1,315 (58.4%)	0.187
Dressing Site, *n* (%)	364 (16.2%)	159 (7.1%)	0.056
Anesthesia Grade	2.8 [1-5]	3.2 [2-5]	< 0.001
Anesthesia Method, *n* (%)	2,145 (95.2%)	1,519 (67.5%)	< 0.001
Oxygen Saturation (SpO2), %	97 [76-100]	97 [65-100]	< 0.001
Self-Care Ability Grade	2.9 [1-3]	2.9 [2-3]	0.002
Blood Glucose (BG)	7.9 [1.4-28.0]	7.9 [3.8-13.4]	0.447

Table 5. Supplementary data-enhanced dataset - SMOTE + ENN

Variables	Non-PI (*n*=2,252)	PI (*n*=2,252)	*p* value
Sex, *n* (%)	1,329 (59.0%)	809 (35.9%)	< 0.001
Age, years	53 [0-98]	74 [0-94]	< 0.001
Hypertension (HTN), *n* (%)	329 (14.6%)	89 (4.0%)	< 0.001
Hyperlipidemia, *n* (%)	11 (0.5%)	0 (0%)	0.489
Diabetes Mellitus (DM), *n* (%)	91 (4.0%)	30 (1.3%)	0.103
Cardiovascular Disease (CVD), *n* (%)	4 (0.2%)	0 (0%)	0.360
History of Malignant Tumor, *n* (%)	127 (5.6%)	31 (1.4%)	< 0.001
Smoking Status, *n* (%)	230 (10.2%)	32 (1.4%)	< 0.001
Alcohol Consumption Status, *n* (%)	2 (0.1%)	0 (0%)	0.267
Body Temperature, °C	36.6 [35.3-40.7]	36.5 [36.0-40.3]	0.586
Pulse Rate (PR), bpm	88 [18-198]	83 [52-165]	< 0.001
Respiratory Rate (RR), bpm	20 [10-78]	18 [12-33]	< 0.001
Diastolic Blood Pressure (DBP), mmHg	75 [24-152]	81 [56-134]	0.003
Systolic Blood Pressure (SBP), mmHg	124 [53-277]	134 [50-195]	0.979
Body Mass Index (BMI), kg/m²	21.6 [7.2-49]	22.7 [12.8-41.6]	< 0.001
Serum Albumin, g/L	36.8 [16.5-53.9]	33.2 [18.4-45.7]	0.451
Operating Time, h	2.89 [0.15-6.84]	2.76 [0.25-10.02]	< 0.001
Intraoperative Blood Transfusion, *n* (%)	222 (9.9%)	16 (0.7%)	< 0.001
Intraoperative Hypotension (IH), *n* (%)	198 (8.8%)	1 (0.1%)	< 0.001
Surgical Position	0.8 [0-3]	0.3 [0-3]	< 0.001
Surgical Dressing, *n* (%)	1,888 (83.8%)	1,335 (59.3%)	< 0.001
Dressing Site, *n* (%)	364 (16.2%)	150 (6.7%)	< 0.001
Anesthesia Grade	2.8 [1-5]	3.2 [2-5]	< 0.001
Anesthesia Method, *n* (%)	2,145 (95.2%)	1,555 (69.0%)	< 0.001
Oxygen Saturation (SpO2), %	97 [76-100]	97 [65-100]	< 0.001
Self-Care Ability Grade	2.9 [1-3]	2.9 [2-3]	< 0.001
Blood Glucose (BG)	7.9 [1.4-28.0]	7.9 [3.8-13.4]	0.845

Table 6. Supplementary data-enhanced dataset - Borderline-SMOTE

Variables	Non-PI (*n*=2,252)	PI (*n*=2,252)	*p* value
Sex, *n* (%)	1,329 (59.0%)	701 (31.1%)	< 0.001
Age, years	53 [0-98]	79 [0-94]	< 0.001
Hypertension (HTN), *n* (%)	329 (14.6%)	773 (34.3%)	< 0.001
Hyperlipidemia, *n* (%)	11 (0.5%)	0 (0%)	0.014
Diabetes Mellitus (DM), *n* (%)	91 (4.0%)	53 (2.4%)	0.001
Cardiovascular Disease (CVD), *n* (%)	4 (0.2%)	0 (0%)	0.061
History of Malignant Tumor, *n* (%)	127 (5.6%)	11 (0.5%)	< 0.001
Smoking Status, *n* (%)	230 (10.2%)	12 (0.5%)	< 0.001
Alcohol Consumption Status, *n* (%)	2 (0.1%)	0 (0%)	0.985
Body Temperature, °C	36.6 [35.3-40.7]	36.6 [36.0-40.3]	0.718
Pulse Rate (PR), bpm	88 [18-198]	84 [52-165]	< 0.001
Respiratory Rate (RR), bpm	20 [10-78]	18 [12-33]	< 0.001
Diastolic Blood Pressure (DBP), mmHg	75 [24-152]	83 [56-134]	< 0.001
Systolic Blood Pressure (SBP), mmHg	124 [53-277]	138 [50-195]	0.005
Body Mass Index (BMI), kg/m²	21.6 [7.2-49]	22.7 [12.8-41.6]	< 0.001
Serum Albumin, g/L	36.8 [16.5-53.9]	33.2 [18.4-45.7]	0.034
Operating Time, h	2.89 [0.15-6.84]	2.49 [0.25-10.02]	0.137
Intraoperative Blood Transfusion, *n* (%)	222 (9.9%)	372 (16.5%)	< 0.001
Intraoperative Hypotension (IH), *n* (%)	198 (8.8%)	39 (1.7%)	0.468
Surgical Position	0.8 [0-3]	0.3 [0-3]	< 0.001
Surgical Dressing, *n* (%)	1,888 (83.8%)	1,168 (51.9%)	< 0.001
Dressing Site, *n* (%)	364 (16.2%)	172 (7.6%)	< 0.001
Anesthesia Grade	2.8 [1-5]	3.2 [2-5]	< 0.001
Anesthesia Method, *n* (%)	2,145 (95.2%)	1,466 (65.1%)	< 0.001
Oxygen Saturation (SpO2), %	97 [76-100]	97 [65-100]	0.151
Self-Care Ability Grade	2.9 [1-3]	2.9 [2-3]	0.001
Blood Glucose (BG)	7.9 [1.4-28.0]	7.9 [3.8-13.4]	0.969

Table 7. Confusion matrix for each model (TN/FP/FN/TP)

Model	RAW Data	SMOTE	Borderline-SMOTE
SVM	2814/0/158/0	1477/775/461/1791	1683/569/250/2002
LR	2802/12/151/7	1938/314/367/1885	1995/257/292/1960
RF	2812/2/157/1	2184/68/57/2195	2203/49/80/2172
ET	2813/1/157/1	2182/70/56/2196	2199/53/76/2176
KNN	2787/27/153/5	1767/485/28/2224	1909/343/38/2214
DT	2655/159/123/35	2059/193/114/2138	2067/185/113/2139
XGBoost	2780/34/147/11	2195/57/83/2169	2189/63/86/2166
Model	SMOTE+ENN	ADASYN	GAN
SVM	1477/775/461/1791	1415/837/421/1831	2252/0/341/1911
LR	1938/314/367/1885	1916/336/369/1883	2237/15/302/1950
RF	2184/68/57/2195	2168/84/76/2176	2242/10/83/2169
ET	2182/70/56/2196	2147/105/89/2163	2238/14/87/2165
KNN	1767/485/28/2224	1734/518/21/2231	2218/34/75/2177
DT	2059/193/114/2138	2019/233/149/2103	2102/150/124/2128
XGBoost	2195/57/83/2169	2181/71/84/2168	2192/60/89/2163

all ML models improvedsignificantly after using SMOTE and its variants (*e.g.*, SMOTE+ENN, Borderline-SMOTE). For example, XGBoost improved its AUC value from 0.800 to 0.996 after using SMOTE, showing that methods of data enhancement play an important role in improving the model's ability to recognize a small number of classes.

Figure 4 shows the ROC analyses of the ML models under different conditions: (A) based on raw data, (B) based on SMOTE, (C) based on SMOTE+ENN, (D) based on Borderline-SMOTE, (E) based on GAN, and (F) based on ADASYN. These curves demonstrate the models' performance under various thresholds, with higher AUC values indicating better classification performance. The results clearly indicate that the models using SMOTE and its variants hadsignificantly higher AUC values compared to those using raw data, highlighting the effectiveness of these techniques in addressing a data imbalance.

When dealing with imbalanced data, SMOTE generates minority class samples by interpolation, which effectively increases sample diversity but may introduce

Table 8. Model performance comparison

Models	Precision	Recall	F1-score	Accuracy	AUC
Raw data					
SVM	0.000	0.000	0.000	0.947	0.565
LR	0.350	0.025	0.046	0.947	0.781
RF	0.200	0.012	0.023	0.946	0.795
ET	0.100	0.006	0.012	0.947	0.786
KNN	0.065	0.026	0.037	0.939	0.649
DT	0.151	0.165	0.152	0.898	0.573
XGBoost	0.290	0.082	0.124	0.938	0.800
SMOTE					
SVM	0.695	0.794	0.741	0.723	0.805
LR	0.846	0.816	0.829	0.834	0.919
RF	0.975	0.982	0.977	0.978	0.998
ET	0.974	0.974	0.973	0.974	0.996
KNN	0.832	0.991	0.904	0.895	0.961
DT	0.918	0.95	0.932	0.932	0.930
XGBoost	0.975	0.964	0.966	0.969	0.996
SMOTE+ENN					
SVM	0.701	0.786	0.741	0.725	0.802
LR	0.858	0.835	0.844	0.849	0.925
RF	0.974	0.978	0.975	0.976	0.997
ET	0.972	0.974	0.972	0.973	0.996
KNN	0.826	0.986	0.899	0.889	0.957
DT	0.917	0.946	0.930	0.930	0.932
XGBoost	0.982	0.961	0.967	0.971	0.993
Borderline-SMOTE					
SVM	0.782	0.887	0.831	0.820	0.879
LR	0.884	0.871	0.875	0.879	0.949
RF	0.978	0.964	0.968	0.971	0.992
ET	0.984	0.971	0.975	0.977	0.992
KNN	0.872	0.985	0.925	0.921	0.969
DT	0.927	0.955	0.938	0.940	0.938
XGBoost	0.974	0.960	0.963	0.968	0.989
ADASYN					
SVM	0.690	0.783	0.729	0.715	0.791
LR	0.834	0.818	0.824	0.827	0.930
RF	0.957	0.924	0.939	0.941	0.995
ET	0.949	0.907	0.927	0.930	0.993
KNN	0.813	0.934	0.869	0.859	0.957
DT	0.881	0.876	0.878	0.879	0.921
XGBoost	0.958	0.944	0.949	0.951	0.994
GAN					
SVM	1.000	0.930	0.958	0.965	0.933
LR	0.993	0.930	0.954	0.962	0.932
RF	0.995	0.930	0.955	0.963	0.939
ET	0.994	0.930	0.955	0.962	0.939
KNN	0.985	0.930	0.950	0.957	0.957
DT	0.933	0.930	0.922	0.928	0.927
XGBoost	0.975	0.930	0.945	0.952	0.951

noise near boundary samples. Nevertheless, SMOTE performs well in most models, and especially in XGBoost and Random Forest (RF), with an AUC value close to 1, indicating strong classification ability.SMOTE+ENN combines the undersampling techniques of SMOTE and ENN, which aim to remove noisy samples and further optimize the quality of the minority class samples. Although its performance is slightly inferior to SMOTE in most models, its performance is close in some models (*e.g.*, KNN), suggesting that it is effective in removing noise but may have sacrificed some of the sample diversity.Borderline-SMOTE focuses on generating samples near the category boundaries, which helps to improve the model's ability to discriminate between the boundary regions, but has limited performance improvement in most models and with high-dimensional data, the definition of boundary samples may not be clear enough, limiting its effectiveness.ADASYN is similar to SMOTE, but focuses more on the hard-to-learn regions of the minority class samples and improves the model performance through adaptive sampling.ADASYN performs well in some models (*e.g.*, XGBoost), but the overall performance is slightly lower than that of SMOTE, probably because the way it generates samples relies more on the local distribution of the minority class samples.GAN, as a state-of-the-art generative adversarial

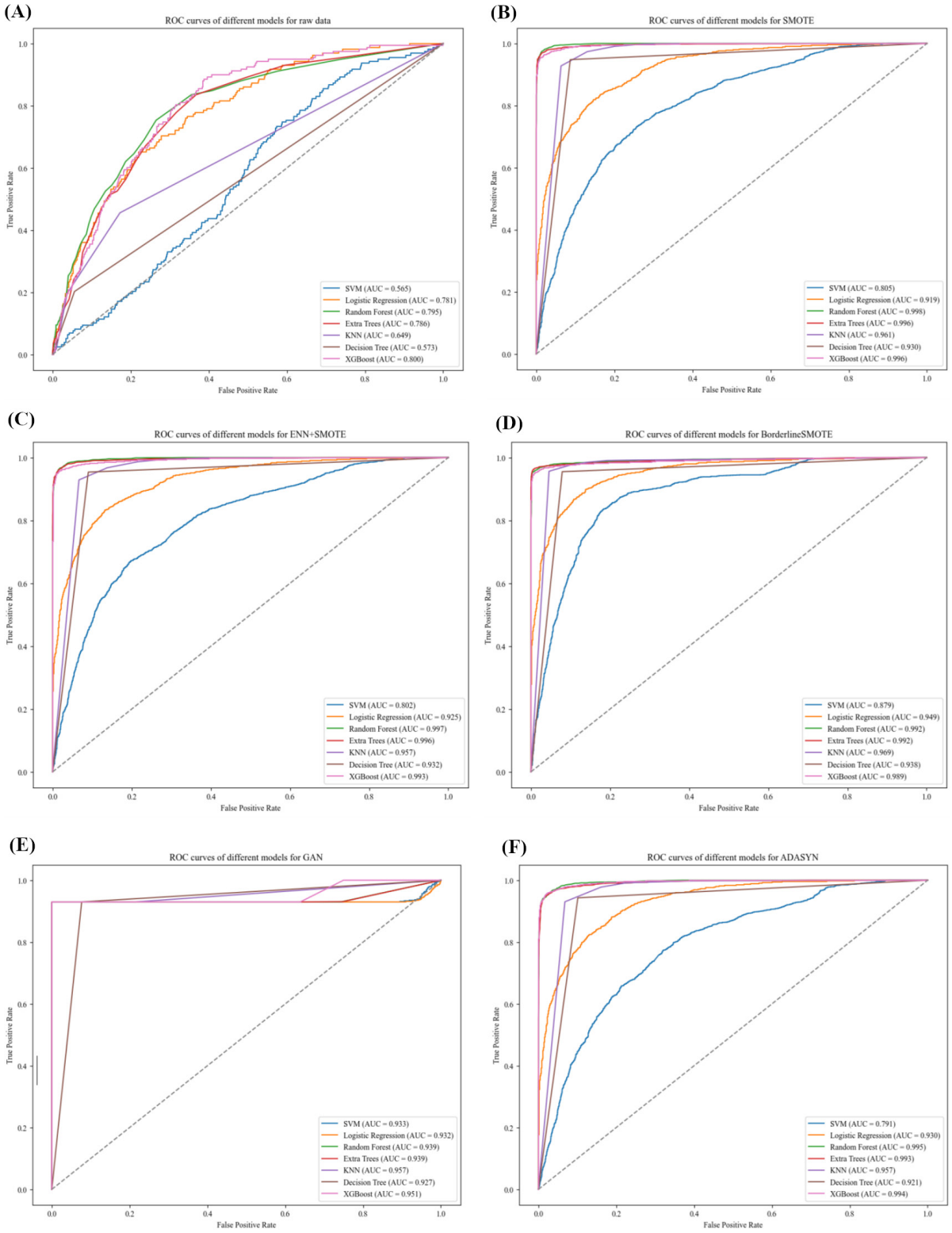


Figure 4. ROC analyses of applied machine learning models.(A) based on raw data, **(B)** based on SMOTE, **(C)** based on SMOTE+ENN, **(D)** based on Borderline-SMOTE, **(E)** based on GAN, **(F)** based on ADASYN.

network, generates high-quality minority class samples through the adversarial training of generators and discriminators. It performs well in some models, but its computational cost is high and it may face the problem of unstable training with high-dimensional data. In addition, the samples generated by GANs may be too close to the original samples, increasing the risk of overfitting.

In conclusion, methods of data enhancement, and especially SMOTE and its variants, have significant effects on improving the performance of models. In practical use, the most appropriate methods of data enhancement can be selected dependingto the specific

problems and models. These methods effectively improve the performance of the model withimbalanced datasets by increasing the number and diversity of samples, which improves the precision, recall, and overall classification performance of the model.

3.3. Validation and interpretability

Table 9 shows the five-fold cross-validation of SMOTE-based processed data.After comparing the performance of the model in 5-fold cross-validation and the original dataset, RF and XGBoost displayed the great stability and consistency in both methods of evaluation, with an AUC value close to 1, indicating its excellent generalization ability across different datasets.

In order to prevent possible overfitting after SMOTE processing and to test the generalization ability of the model, external validation of the constructed modelwas

Table 9. Fifty-fold cross-validation of the models

Models	Precision	Recall	F1-score	Accuracy	AUC
SVM	0.695 (0.664 - 0.721)	0.794 (0.772 - 0.823)	0.741 (0.729 - 0.763)	0.723 (0.700 - 0.744)	0.805 (0.788 - 0.831)
LR	0.846 (0.777 - 0.854)	0.816 (0.584 - 0.894)	0.829 (0.666 - 0.872)	0.834 (0.709 - 0.856)	0.919 (0.802 - 0.952)
RF	0.975 (0.925 - 0.990)	0.982 (0.857 - 1.000)	0.977 (0.958 - 0.987)	0.978 (0.922 - 0.992)	0.998 (0.997 - 1.000)
ET	0.974 (0.929 - 0.990)	0.974 (0.798 - 1.000)	0.973 (0.879 - 0.990)	0.974 (0.892 - 0.990)	0.996 (0.968 - 1.000)
KNN	0.832 (0.724 - 0.888)	0.991 (0.860 - 1.000)	0.904 (0.729 - 0.940)	0.895 (0.819 - 0.914)	0.961 (0.932 - 0.971)
DT	0.918 (0.868 - 0.949)	0.950 (0.766 - 0.970)	0.932 (0.848 - 0.949)	0.932 (0.863 - 0.949)	0.932 (0.863 - 0.949)
XGBoost	0.975 (0.938 - 0.985)	0.964 (0.660 - 1.000)	0.966 (0.790 - 0.987)	0.969 (0.825 - 0.995)	0.997 (0.975 - 1.000)

Table 10. External validation dataset distribution

Variables	Non-PI (*n*=277)	PI (*n*=277)	*p* value
Sex, *n* (%)	160 (57.7%)	192 (68.6%)	0.004
Age, years	53 [0-91]	73 [2-94]	< 0.001
Hypertension (HTN), *n* (%)	244 (88.1%)	173 (62.4%)	< 0.001
Hyperlipidemia, *n* (%)	276 (99.6%)	277 (100.0%)	0.751
Diabetes Mellitus (DM), *n* (%)	13 (4.7%)	8 (2.9%)	0.037
Cardiovascular Disease (CVD), *n* (%)	1 (0.4%)	0 (0%)	0.266
History of Malignant Tumor, *n* (%)	11 (4.0%)	2 (0.7%)	0.084
Smoking Status, *n* (%)	28 (10.1%)	4 (1.4%)	0.067
Alcohol Consumption Status, *n* (%)	1 (0.4%)	0 (0%)	0.481
Body Temperature, °C	36.6 [35.9-38.4]	36.5 [36.0-39.7]	0.599
Pulse Rate (PR), bpm	88 [38-170]	82 [53-150]	0.044
Respiratory Rate (RR), bpm	20 [11-70]	18 [12-32]	0.058
Diastolic Blood Pressure (DBP), mmHg	75 [30-113]	81 [56-120]	0.012
Systolic Blood Pressure (SBP), mmHg	123 [57-190]	133 [86-194]	0.133
Body Mass Index (BMI), kg/m²	21.5 [11.1-49]	22.7 [13.9-29.9]	0.090
Serum Albumin, g/L	37.1 [18.6-34.3]	33.1 [20.0-42.8]	0.401
Operating Time, h	2.89 [0.17-6.43]	2.78 [0.37-9.71]	< 0.001
Intraoperative Blood Transfusion, *n* (%)	256 (92.4%)	230 (83.0%)	0.004
Intraoperative Hypotension (IH), *n* (%)	248 (89.5%)	270 (97.5%)	0.423
Surgical Position	0.8 [0-3]	0.3 [0-3]	< 0.001
Surgical Dressing, *n* (%)	233 (84.1%)	156 (56.3%)	< 0.001
Dressing Site, *n* (%)	44 (15.9%)	15 (5.4%)	< 0.001
Anesthesia Grade	2.8 [1-5]	3.1 [2-5]	0.541
Anesthesia Method, *n* (%)	263 (94.9%)	195 (70.4%)	0.001
Oxygen Saturation (SpO2), %	97 [76-100]	96 [67-100]	0.442
Self-Care Ability Grade	2.9 [2-3]	2.9 [2-3]	0.206
Blood Glucose (BG)	7.9 [2.6-23.0]	7.9 [3.8-12.9]	0.655

Table 11. Performance of each model under external validation

Models	Precision	Recall	F1-score	Accuracy	AUC
SVM	0.731	0.733	0.731	0.731	0.731
LR	0.825	0.829	0.825	0.824	0.825
RF	0.939	0.945	0.939	0.938	0.939
ET	0.940	0.947	0.940	0.940	0.940
KNN	0.872	0.872	0.872	0.872	0.872
DT	0.942	0.948	0.942	0.942	0.942
XGBoost	0.977	0.978	0.977	0.977	0.977

attempted, but since the original amount of data was very small and had already been internally validated, cutting the data inside for external validation would have affected the performance of the original model, so the missing values that had been excluded from the original set of 9-14 data strips were used on a 1:1 basis. Positive and negative sampleswere selected in the order of missing values, and the missing values were added according to the data filling method in the previous section to serve as the external validation set. The data distribution of the external validation set is shown in Table 10, and the external validation results are shown in Table 11.

Based onthe external validation data, the precision, recall, F1 score, accuracy and AUC values of all models are very close to each other, indicatinga high degree of consistency in the performance of the models on the new dataset. The XGBoost model performed best in external validation, with a precision, recall, F1 score, accuracy and AUC value of 0.977, which is close to perfect, indicating excellent generalizability and prediction performance.

Combining the results of internal cross-validation and external validation, XGBoost, Random Forest (RF), and Extra Trees (ET) performed the best in terms of performance and stability. These models not only displayed low variability and high stability in internal cross-validation but also exhibited extremely high accuracy and AUC values in external validation, indicating their excellent generalizability. Especially, XGBoost, with its near-perfect external validation results, is the best choice among all models.

In response to the pervasive black-box problem of ML, SHAP (SHapley Additive exPlanations)has been introduced to increase the interpretability of the model. The scatterplot of SHAP values reveals the extent to which different features contribute to the predicted results of a ML model. Each point in the graph represents the SHAP value of a sample, which measures the contribution of a particular feature to the model output. The color gradient goes from blue to red, representing low to high feature values, respectively. Figure 5 shows that Anesthesia Grade has a significant effect on the model output. A high Anesthesia Grade (red points) is generally located on the right side of the graph, which indicates that it tends to increase the predictive value of the model when the Anesthesia Grade is high. Conversely, low anesthesia levels (blue points) tend to decrease the predicted value of the model, and most of these points are located on the left side of the graph. Age is also a key factor that has a broad impact on model predictions. Older people (red dots) tend to have

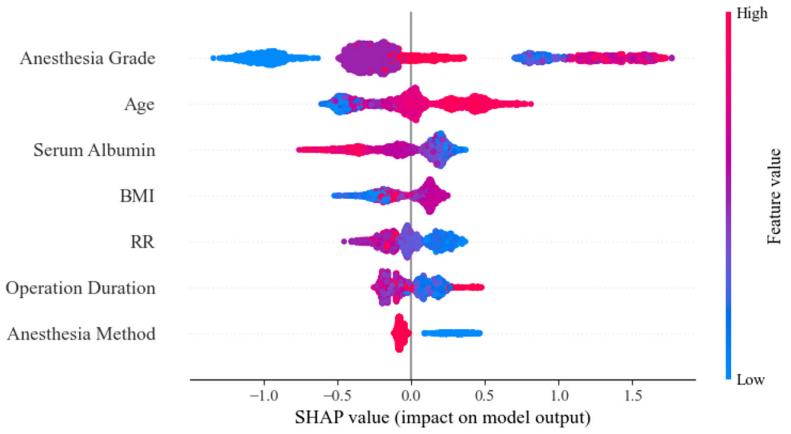


Figure 5. SHAP Summary plot of key factors.

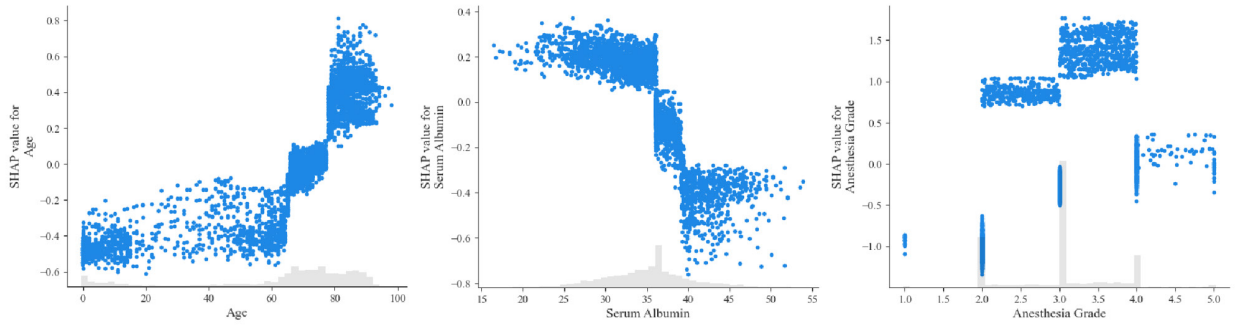


Figure 6. SHAP dependency plot.

positive SHAP values, implying that an increase in age may improve the model's predictions. In contrast, SHAP values for younger people (blue dots) tend to be negative, suggesting that a younger age may decrease model predictions. Serum albumin levels also had a significant impact on model predictions. Samples with high serum albumin levels (red dots) tend to have negative SHAP values in the graph, which could mean that higher serum albumin levels are associated with lower predictions in the model. Low serum albumin levels (blue dots), in contrast, are associated with positive SHAP values, suggesting that lower serum albumin levels may improve the predictive value of the model. The corresponding SHAP values when the variable of interest is a particular value are shown in Figure 6.

4. Discussion

The SMOTE algorithm yielded significant results whendealing with a data imbalance, but there are some potential limitations and risks. First, SMOTE may introduce noise, and especially when noise or outliers are present in a few class samples, and the synthesized samples may also contain that noise, affecting the model performance. Second, SMOTE is sensitive to the choice of parameter k (number of nearest neighbors), and improper values for k may lead to overfitting or the introduction of excessively noisy data. In addition, SMOTE is computationally expensive, and especially when dealing with large-scale datasets, and calculating the k nearest neighbors can be very time-consuming. More importantly, SMOTE increases the number of samples in a few classes by synthesizing samples that may be too close to the original samples, increasing the risk of overfitting and reducing the generalizability of the model. Finally, SMOTE may introduce noise or produce unrealistic data points when generating new samples near the boundary samples, affecting the classification effectiveness of the model.

SMOTEhas been widely used to address the issue of imbalanced class distribution in various ML applications. Sáez *et al.* introduced SMOTE-IPF, a re-sampling method with filtering, to tackle the problem of noisy and borderline examples in imbalanced classification(*21*). Rastogi *et al.* focused on implementing SMOTE in a distributed environment under Spark, highlighting the importance of applyingSMOTE to big data classification(*22*). Bao *et al.*integrated SMOTE with KNN and long short-term memory networks (LSTMs) to detect anomalies in high-dimensional and imbalanced data(*23*). Hemalatha *et al.* proposed FG-SMOTE, a fuzzy-based Gaussian synthetic minority oversampling algorithm, to handle imbalanced data and improve classifier performance(*24*). However, that study identified limitations such as the need to apply FG-SMOTE to multiclass imbalanced datasets and to evaluate theproblem of imbalancein a distributed environment. Mukherjee *et al.* introduced SMOTE-ENC, a novel SMOTE-based method for generating synthetic data with both nominal and continuous features(*25*). That study found that SMOTE-ENC outperformed SMOTE-NC in datasets with a substantial number of nominal features and associations between categorical features and the target class. Xia *et al.* proposed GBSMOTE, a sampling method based on granular-ball computing and SMOTE, to address the limitations of SMOTE such as noisy generated samples and boundary blurring(*26*). In the context of specific applications, Ismail *etal.* combined oversampling and undersampling techniques in SMOTE-RUS to classify imbalanced autism spectrum disorder datasets effectively(*27*). Nazarudin *et al.*usedsynthetic data generation techniques, including SMOTE and GAN-SMOTE, to train ML models to predictTenaga Nasional Berhad stock price movements(*28*). Overall, SMOTE has been a valuable tool in addressing a class imbalance, but studies have identified its limitations such as noisy samples, boundary blurring, and challenges in handling multiclass datasets and distributed environments. Future research may focus on enhancing SMOTE algorithms to overcome these limitations and improve their effectiveness in various applications.

To address these limitations and risks, future work can explore several directions. First, improved versions of SMOTE, such as Borderline-SMOTE or ADASYN, can be investigated and developed to improve the performance and stability of the algorithm through different strategies ofselecting the original samples used for generating new samples or adjusting the way in which new samples are generated. Second, the SMOTE algorithm can be used in conjunction with other techniques (*e.g.*, undersampling and integrated learning) to further improve the performance of the model. For example, the SMOTE algorithm can be used to oversample a small number of classes first, and then integrated learning methods can be used to train multiple models and obtain the final prediction results by voting or averaging. In addition, suitable evaluation metrics need to be used to assess the performance of the models, and especially withimbalanced datasets, where metrics such as recall and F1 scores often reflect the actual performance of the models better than accuracy. New learning algorithms designed specifically for imbalanced data can also be developedto improve the recognition of minority classes by adjusting sample weights or other mechanisms without increasing the number of samples. Finally, with the advent of the big data era, the useof SMOTEin big data environments can beexploredto address the challenges posed by the expanded size of data, such as computational efficiency and storage issues, is also an important direction for future work. Through these efforts, we can address the problem of a data imbalance more effectively and improve the predictive performance and generalization ability of the model.

This study had several limitations.First,the total

number of samples is still somewhat small relative to ML, so the model performance after SMOTE is bound to have overfitting to a certain extent. With the subsequent supplementation of the external validation set, there is also a certain amount of contamination of the training data. Second, based on data from only one hospital, the population is affected by the geographic area and may not necessarily be generalizable to other geographic areas.Further research will be conducted based on these issues in conjunction with multiple hospitals.

5. Conclusion

This study underscores the significance of usingML models to address the challenge of data imbalances in the prediction of perioperative PIs. The integration of synthetic minority oversampling techniques, and particularly SMOTE, with ML algorithms has been found to markedly enhance predictive accuracy, and especially in scenarios with few positive samples. The useof SMOTE and its variants, such as SMOTE+ENN and Borderline-SMOTE, has been shown to bolster the model's capacity to recognize minority classes, leading to more nuanced predictive modeling for PIs in emergency patient populations.

Among the seven ML models assessed, the combination of XGBoost with SMOTE emerged as the most effective, withan internally validated AUC of 0.996 and an externally validated AUC of 0.977. This result underscores the superior discriminative power of the XGBoost model when combined with SMOTE, outperforming other models across various metrics including precision, recall, F1 score, and accuracy. This study not only highlights the clinical utility of ML models augmented with SMOTE technology in predicting PIs but also underscores the importance of controlling a data imbalance toenhance the predictive value of ML models in healthcare settings. The findings suggest that the synergy of SMOTE with ML algorithms presents a viable strategy for mitigating the limitations of conventional risk assessment tools and dealing with the inherent data imbalances present in healthcare data. Future research is warranted to refine SMOTE techniques, explore their integration with other methodologies, and develop novel algorithms tailored for imbalanced datasets, thereby improving the reliability and accuracy of ML models in healthcare.

Acknowledgements

The authorswould like to sincerely thank the hospital's clinical research center for their invaluable support and assistance in database maintenance and data collection.

Funding: This work was financially supported by Shanghai Science and Technology Innovation Action Plan (23410761400), Xinhua Hospital CRU (21XHDB03), the Ministry of Education of China Humanities and Social Sciences Youth Fund Project (22YJC790189), Shanghai Key Laboratory of Urban Design and Urban Science Open Topic Grants of NYU Shanghai (Grant No.2023YWZhou_LOUD), and the Cultivation Project of School of Intelligent Emergency Management of University of Shanghai for Science and Technology.

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Bajer D, Zonć B, Dudjak M, Martinović G. Performance analysis of SMOTE-based oversampling techniques when dealing with data imbalance. 2019 International Conference on Systems, Signals and Image Processing (IWSSIP). 2019; 265-271. https://doi.org/10.1109/IWSSIP.2019.8787306.
2. Kumar V, Singh Lalotra G, Sasikala P, Singh Rajput D, Kaluri R, Lakshmanna K, Shorfuzzaman M, Alsufyani A, Uddin M. Addressing binary classification over class imbalanced clinical datasets using computationally intelligent techniques. Healthcare (Basel). 2022; 10:1293.
3. Ishaq A, Sadiq S, Umer M, Ullah S, Mirjalili S, Rupapara V, Nappi M. Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. IEEE Access. 2021; 9:39707-39716.
4. van den Goorbergh R, van Smeden M, Timmerman D, Van Calster B. The harm of class imbalance corrections for risk prediction models: Illustration and simulation using logistic regression. J Am Med Inform Assoc. 2022; 29:1525-1534.
5. Ridwan AM, Setyawan GD. Perbandingan berbagai model machine learning untuk mendeteksi diabetes. Teknokon. 2023; 6:127-122.
6. Javid SA, Feghhi MM. Early diagnosis of Alzheimer's disease from MRI images with deep learning model. 2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP). 2024; 1-5. https://ieeexplore.ieee.org/document/10475240.
7. Kottner J, Cuddigan J, Carville K, Balzer K, Berlowitz D, Law S, Litchford M, Mitchell P, Moore Z, Pittman J, Sigaudo-Roussel D, Yee CY, Haesler E. Prevention and treatment of pressure ulcers/injuries: The protocol for the second update of the international Clinical Practice Guideline 2019. J Tissue Viability. 2019; 28:51-58.
8. Anderson C, Bekele Z, Qiu Y, Tschannen D, Dinov ID. Modeling and prediction of pressure injury in hospitalized patients using artificial intelligence. Med Inform Decis Mak. 2021; 21:253.
9. Vera-Salmerón E, Domínguez-Nogueira C, Sáez JA, Romero-Béjar JL, Mota-Romero E. Differentiating pressure ulcer risk levels through interpretable classification models based on readily measurable indicators. Healthcare (Basel). 2024; 12:913.
10. Chumachenko D, Yakovlev S. Artificial intelligence algorithms for healthcare. Algorithms. 2024; 17:105.
11. Chawla LV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Art Intel Res. 2002; 16:321-357.
12. He H, Bai Y, Garcia EA, Li S. ADASYN: Adaptive

synthetic sampling approach for imbalanced learning. IEEE International Joint Conference on Neural Networks. 2008; 1322-1328. https://ieeexplore.ieee.org/document/4633969.
13. Nizam-Ozogur H, Orman Z. A heuristic-based hybrid sampling method using a combination of SMOTE and ENN for imbalanced health data. Expert Systems. 2024; e13596.
14. Ma H, Song M, Zhu Y. Improved Borderline-SMOTE oversampling method for boundary classification. J Nanjing Univ (Natural Science). 2023; 59:1003-1012. (in Chinese)
15. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. AdvanNeural Info Process Syst. 2014; 2:2672-2680.
16. Araveeporn A, Wanitjirattikal W. Comparison of machine learning methods for binary classification of multicollinearity data. 2024 International Conference on Machine Learning and Cybernetics (ICMLC). IEEE. 2024; 44-49. https://doi.org/10.1145/3686592.3686600.
17. Kyriazos T, Poga M. Application of machine learning models in social sciences: Managing nonlinear relationships. Encyclopedia. 2024; 4:1790-1805.
18. Mujahid M, Kına E, Rustam F, Gracia Villar M, Silva Alvarado E, De La Torre Diez I, Ashraf I. Data oversampling and imbalanced datasets: An investigation of performance for machine learning and feature engineering. J Big Data. 2024; 11:87.
19. Halder RK, Uddin MN, Uddin MA, Aryal S, Khraisat A. Enhancing K-nearest neighbor algorithm: A comprehensive review and performance analysis of modifications. J Big Data. 2024; 11:113.
20. Azevedo BF, Rocha AMAC, Pereira AI. Hybrid approaches to optimization and machine learning methods: a systematic literature review. Mach Learn. 2024; 113:4055-4097.
21. Sáez JA, Luengo J, Stefanowski J, Herrera F. SMOTE-IPF: Addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering. InfoSci. 2015; 291:184-203.
22. Rastogi, A. K., Narang, N., & Siddiqui, Z. A. Imbalanced big data classification: A distributed implementation of SMOTE. Proceedings of the Workshop Program of the 19th International Conference on Distributed Computing and Networking. 2018; 14:1-6. https://doi.org/10.1145/3170521.3170535.
23. Bao F, Wu Y, Li Z, Li Y, Liu L, Chen G. Effect improved for high-dimensional and imbalanced data anomaly detection model based on KNN-SMOTE-LSTM. Complexity. 2020; 2020:1-17.
24. Hemalatha P, Amalanathan GM. FG-SMOTE: Fuzzy-based Gaussian synthetic minority oversampling with deep belief networks classifier for skewed class distribution. Interntl J Intel Computing Cybernet. 2021; 14:270-287.
25. Mukherjee M, Khushi M. SMOTE-ENC: A novel SMOTE-based method to generate synthetic data for nominal and continuous features. Appl Syst Innov. 2021; 4:18.
26. Xia T, Shao Y, Xia S, Xiong Y, Lian X, Ling W. GBSMOTE: A robust sampling method based on Granular-ball Computing and SMOTE for class imbalance. Proceedings of the 2023 8th International Conference on Pattern Recognition and Artificial Intelligence. 2023: 19-24. https://dl.acm.org/doi/10.1145/3594300.3594304.
27. Ismail E, Gad WK, Hashem M. SMOTE-RUS: Combined oversampling and undersampling technique to classify the imbalanced Autism Spectrum Disorder dataset. Interntl J Intel Computing Cybernet. 2023; 23:83-94.
28. Mohd Nazarudin NAS, Mohd Ariffin NH, Maskat R. Leveraging on synthetic data generation techniques to train machine learning models for Tenaga Nasional Berhad stock price movement prediction. Internatl Arab J Info Tech. 2024; 21:483-494.

Received January 11, 2025; Revised February 25, 2025; Accepted March 26, 2025.

*Address correspondence to:
Liping Jiang and Ping Liu, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, No. 1665 Kongjiang Road, Yangpu District, Shanghai 200092, China.
E-mail: jiangliping@xinhuamed.com.cn (LJ), 13868311990@163.com (PL)

Released online in J-STAGE as advance publication April 15, 2025.

Original Article

DOI: 10.5582/bst.2025.01065

Exosomes derived from olfactory mucosa mesenchymal stem cells attenuate cognitive impairment in a mouse model of Alzheimer's disease

Xiqi Hu, Ya-nan Ma, Jun Peng, Zijie Wang, Yuchang Liang, Ying Xia*

Department of Neurosurgery, Haikou Affiliated Hospital of Central South University Xiangya School of Medicine, Haikou, China.

SUMMARY: Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by cognitive decline, neuroinflammation, and endoplasmic reticulum (ER) stress. In recent years, exosomes have garnered significant attention as a potential therapeutic tool for neurodegenerative diseases. This study, for the first time, investigates the neuroprotective effects of exosomes derived from olfactory mucosa mesenchymal stem cells (OM-MSCs-Exos) in AD and further explore the potential role of low-density lipoprotein receptor-related protein 1 (LRP1) in this process. Using an Aβ1-42-induced AD mouse model, we observed that OM-MSCs-Exos significantly improved cognitive function in behavioral tests, reduced neuroinflammatory responses, alleviated ER stress, and decreased neuronal apoptosis. Further analysis revealed that OM-MSCs-Exos exert neuroprotective effects by modulating the activation of microglia and astrocytes and influencing the ER stress response, a process that may involve LRP1. Although these findings support the potential neuroprotective effects of OM-MSCs-Exos, further studies are required to explore their long-term stability, dose dependency, and immunogenicity to assess their feasibility for clinical applications.

Keywords: cellular stress regulation, microglial activation, astrocytic response, pro-inflammatory cytokines, neuroprotection, cognitive improvement

1. Introduction

With aging of the global population, Alzheimer's disease (AD) has emerged as a major public health challenge and a key focus of neurodegenerative disease research (*1*). The pathological hallmarks of AD primarily include neurofibrillary tangles (NFTs) formed by hyperphosphorylated tau protein and amyloid plaques resulting from the deposition of insoluble β-amyloid protein (Aβ) (*2*). These abnormal proteins progressively accumulate, triggering neuroinflammatory responses, activating microglia and astrocytes, and exacerbating neuronal damage and neurodegenerative changes (*2*).

The pathogenesis of AD is complex, involving multiple cellular and molecular abnormalities, such as endoplasmic reticulum (ER) stress, mitochondrial dysfunction, oxidative stress, and chronic inflammation (*2*). Studies have shown that several ER stress markers, including phosphorylated PERK, phosphorylated IRE1α, phosphorylated eIF2α, XBP1, and CHOP, are significantly upregulated in the brain tissue of patients with AD (*3-5*). Moreover, ER stress enhances γ-secretase activity, promoting Aβ secretion and intensifying the accumulation of abnormal proteins, thereby accelerating the progression of AD (*6,7*).

Despite advances in AD research, the development of effective treatments to slow its progression remains elusive due to its heterogeneity and multifactorial pathogenesis (*8*). Consequently, the development of targeted and innovative therapeutic approaches has become a critical priority in AD research. In recent years, stem cell therapy has gained considerable attention due to its pluripotency and immunomodulatory properties, emerging as a promising avenue in neurodegenerative disease studies (*9,10*). Specifically, research has demonstrated that transplantation of bone marrow-derived mesenchymal stem cells (BMSCs), adipose-derived mesenchymal stem cells (ADMSCs), and neural stem cells (NSCs) can potentially have neuroprotective and reparative effects on neurodegenerative diseases (*8*).

Olfactory mucosa mesenchymal stem cells (OM-MSCs), a unique type of stem cell derived from the ectoderm, retain the immunomodulatory and tissue repair capabilities of traditional mesenchymal stem cells (MSCs) while exhibiting enhanced neurogenic potential. OM-MSCs have displayed distinct advantages in neurodegenerative disease research, particularly due to their minimally invasive isolation from the nasal cavity,

which reduces ethical concerns. Despite the remarkable therapeutic potential of stem cell therapy, however, direct transplantation faces certain limitations, including difficulties in crossing the blood-brain barrier (BBB) and potential tumorigenic risks. As a result, recent studies have increasingly shifted focus toward the paracrine effects of stem cells, and particularly the extracellular vesicles (EVs) and exosomes they secrete.

Exosomes are lipid nanoparticles with diameters ranging from approximately 30 to 150 nanometers, that are capable of carrying a variety of bioactive molecules, such as proteins, nucleic acids, and lipids. Due to their favorable biocompatibility and low immunogenicity, exosomes can effectively cross the BBB, positioning them as an emerging therapeutic strategy for neurodegenerative diseases (*11*). For instance, a study has shown that BMSCs-derived exosomes can significantly ameliorate cognitive dysfunction in AD-like mouse models (*9*).

Nevertheless, the potential role and underlying molecular mechanisms of OM-MSCs-Exos in AD remain largely unexplored. To address this gap, the current study aims to investigate the neuroprotective effects and mechanisms of OM-MSCs-Exos in AD treatment using a mouse model of Aβ1-42-induced AD and an SH-SY5Y cell model. Specifically, this study examines whether OM-MSCs-Exos have an effect by modulating neuroinflammation and ER stress responses. The findings of this study are expected to provide theoretical support for stem cell-based therapeutic strategies in AD and lay the groundwork for future clinical translational research.

2. Materials and Methods

2.1. Isolation and culture of OM-MSCs

Olfactory epithelial tissue was isolated from the nasal cavity of C57BL/6 mice, cut into small pieces, and cultured in DMEM/F-12 medium (Gibco, USA) supplemented with 15% fetal bovine serum (FBS) for 7 days (*12*). Non-adherent cells were removed, and the remaining cells were digested with trypsin and expanded until passage 3.

Surface markers of OM-MSCs were analyzed using flow cytometry. Specifically, 1×10^6 cells (100 μL) were placed in a 1.5 mL EP tube and incubated with antibodies against CD29, CD90, CD44, CD34, CD45, and CD11b (eBioscience, USA). After incubation at room temperature in the dark for 30 minutes, cells were washed with 1 mL PBS and centrifuged at 350 g for 5 minutes. The supernatant was discarded, and the cells were resuspended in 350 μL of PBS for flow cytometry analysis.

Osteogenic and adipogenic differentiation of OM-MSCs was induced under specific culture conditions. For osteogenic differentiation, OM-MSCs were cultured in an osteogenic induction medium (Abiowell, China) for 3 weeks and stained with Alizarin Red to assess differentiation. For adipogenic differentiation, cells were cultured in an adipogenic induction medium (Abiowell, China) for 14 days and stained with Oil Red O to evaluate differentiation.

2.2. Isolation and characterization of OM-MSCs-Exos

OM-MSCs from passages 3–5 were cultured to 90% confluence, washed three times with PBS, and then incubated in medium containing 10% exosome-depleted FBS for 48 hours. The conditioned medium was collected and stored at -80°C.

The collected supernatant was centrifuged at 1500 rpm for 5 minutes, followed by 3000 rpm for 30 minutes to remove cellular debris. The supernatant was then filtered through a 0.22-μm membrane and concentrated *via* ultrafiltration. After centrifugation at 3000 rpm for 10 minutes, exosomes were pelleted by ultracentrifugation at 100,000 g for 2 hours.

The size distribution of exosomes was measured using nanoparticle tracking analysis (NTA, Nanosight NS300, Malvern, UK). The morphology and size of exosomes were observed using transmission electron microscopy (TEM, HITACHI, Japan).

2.3. Animal model and experimental design

Eight-week-old male C57BL/6 mice weighing 23–25 g were purchased from Hunan SJA Laboratory Animal Co., Ltd. All animals met specific pathogen-free standards and were housed under controlled conditions (temperature: 25±1°C; humidity: 60±5%; 12-hour light/dark cycle) with free access to food and water. The experiments were approved by the Ethics Committee of Haikou Hospital Affiliated with Xiangya Medical College, Central South University, and conducted in accordance with the Guide for the Care and Use of Laboratory Animals published by the National Institutes of Health (*13*).

The mouse model of AD was established as previously described (*14*). Mice were randomly divided into five groups: a sham group, an AD group, an AD+OM-MSCs-Exos group, an AD+si-NC-Exos group, and an AD+si-LRP1-Exos group, with 6 mice per group. To induce AD, mice were anesthetized with sodium pentobarbital, and Aβ1-42 (6 μg) was injected bilaterally into the hippocampus (anterior-posterior: -2.0 mm; medial-lateral: ±1.6 mm; dorsal-ventral: 1.5 mm from Bregma) using a stereotaxic apparatus. The control group received an equal volume of saline.

Subsequently, each mouse received tail vein injections of 100 μL PBS, OM-MSCs-Exos, si-NC-Exos, or si-LRP1-Exos (1 mg/mL) twice weekly for 4 weeks. Following the final injection, cognitive function was assessed using the Morris water maze (MWM) test (*15*) and the novel object recognition test (NORT) (*9*).

2.4. Cell experiments

The human neuroblastoma cell line SH-SY5Y (AW-CCH335, Abiowell, China) was cultured in MEM/F-12 medium supplemented with 10% FBS and 1% penicillin/streptomycin at 37°C in a 5% CO_2 incubator. To establish an *in vitro* model of AD, SH-SY5Y cells were treated with 20 μM Aβ1-42 for 24 hours (*15*). For the treatment groups, 40 μg/mL of OM-MSCs-Exos, si-NC-Exos, or si-LRP1-Exos was added to the culture and incubated for 12 hours (*16*).

2.5. TUNEL fluorescence assay

Cell apoptosis was detected using a TUNEL apoptosis detection kit (FITC). Tissue sections were deparaffinized with xylene, dehydrated through a graded ethanol series, and processed using a TUNEL kit (Shanghai Yeasen Biotech, China) according to the manufacturer's instructions. Sections were incubated with 100 μL of proteinase K working solution at 37°C for 20 minutes, followed by 100 μL 1× equilibration buffer at room temperature for 10–30 minutes. Subsequently, 50 μL of TdT incubation buffer was added, and sections were incubated at 37°C in the dark for 60 minutes. Nuclei were stained with DAPI working solution at 37°C in the dark for 10 minutes. After mounting, sections were observed under a fluorescence microscope.

2.6. Nissl staining

Tissue sections were deparaffinized with xylene and dehydrated through a graded ethanol series before Nissl staining. Differentiation was performed using a specific differentiation solution. Sections were mounted with glycerol and observed under a microscope.

2.7. Immunohistochemistry (IHC)

Brain tissue was dehydrated through a graded ethanol series, embedded in paraffin, and sectioned. Sections were deparaffinized, rehydrated, and subjected to antigen retrieval by heating. They were then incubated at 4°C with primary antibodies against GFAP (PTG, USA) and IBA1 (Abiowell, China) overnight. After they were washed three times with PBS, sections were incubated with HRP-conjugated secondary antibodies. Color development was achieved using a DAB substrate, followed by counterstaining with hematoxylin. Sections were subsequently observed under a microscope.

2.8. RT-qPCR

Total RNA was extracted using a Trizol reagent (Thermo, USA) and reverse-transcribed into cDNA. RT-qPCR was performed using the UltraSYBR Mixture (Beijing CWBio, China). Primers were synthesized by Beijing Tsingke Biotech Co., Ltd. The RT-qPCR protocol included denaturation at 95°C for 10 minutes, followed by 40 amplification cycles (95°C for 15 seconds, 60°C for 30 seconds) (Supplemental Table S1, *https://www.biosciencetrends.com/action/getSupplementalData.php?ID=248*),

2.9. ELISA

Levels of IL-1β, IL-6, TNF-α, and Aβ1-42 expression in brain tissue and cell culture supernatant were measured using ELISA kits (Wuhan Fine Biotech, China) according to the protocol provided.

2.10. Western blot (WB)

Protein concentrations were quantified using a BCA protein assay kit (Abiowell, China). Equal amounts of protein were loaded onto SDS-PAGE gels and transferred to PVDF membranes for immunoblotting. Primary antibodies included Calnexin (Abiowell, China), CD9 (Abiowell, China), CD63 (Proteintech, USA), BCL-2 (Abiowell, China), BAX (Abiowell, China), Cleaved-caspase3 (PTG, USA), CHOP (Abiowell, China), GRP78 (Abiowell, China), ATF6 (Abcam, UK), LRP1 (Abcam, UK), and β-actin (Abiowell, China). Membranes were incubated with primary antibodies at 4°C overnight, followed by HRP-conjugated secondary antibodies. Chemiluminescence was visualized using an ECL detection kit, and band intensities were quantified with the software ImageJ.

2.11. Statistical analysis

Data were analyzed using the software GraphPad Prism 8.0 (GraphPad Software, USA). Normally distributed data are expressed as the mean ± standard deviation, while non-normally distributed data are expressed as the median. Between-group comparisons were evaluated using the Student's *t*-test. For comparisons involving three or more groups, one-way analysis of variance (ANOVA) was used. Qualitative data were analyzed using the chi-square test. A P-value < 0.05 was considered statistically significant.

3. Results

3.1. Isolation and characterization of OM-MSCs and OM-MSCs-Exos

OM-MSCs and OM-MSCs-Exos were successfully isolated and characterized. OM-MSCs were extracted from the olfactory mucosa of mice and passaged to the third generation (P3). Flow cytometry analysis revealed that P3 OM-MSCs expressed typical MSC surface markers, with positive staining for CD29, CD90, and CD44, and negative staining for CD34,

CD45, and CD11b (Figure 1A). These results align with established MSC identification standards. Moreover, the multilineage differentiation potential of OM-MSCs was confirmed through Alizarin Red S staining (to assess osteogenic differentiation) and Oil Red O staining (to assess adipogenic differentiation) (Figure 1B, C).

OM-MSCs-Exos were isolated from the conditioned medium of P3–P5 OM-MSCs. NTA showed that the particle size of OM-MSCs-Exos ranged from 80–180 nm, with a peak at 136 nm (Figure 1D). TEM images revealed a characteristic double-membrane spherical structure, consistent with exosome morphology (Figure 1E). Additionally, WB analysis confirmed positive expression of exosome markers CD9 and CD63 in OM-MSCs-Exos, while these markers were undetectable in OM-MSCs. Calnexin, an endoplasmic reticulum marker, was not detected in OM-MSCs-Exos (Figure 1F). Collectively, these results validated the successful isolation and characterization of OM-MSCs and OM-MSCs-Exos, meeting the established criteria for MSCs and exosomes.

3.2. OM-MSCs-Exos improve spatial learning and memory *in vivo*

To investigate the effects of OM-MSCs-Exos on cognitive function in mice with AD, behavioral experiments were conducted one month after OM-MSCs-Exos transplantation. These included the MWM and NORT to assess spatial learning and memory (Figure 2). In the NORT (Figure 2A), mice with AD that were treated with Aβ1-42 spent significantly less time exploring novel objects, indicating memory impairment induced by Aβ1-42. In contrast, OM-MSCs-Exos treatment enhanced object recognition ability compared to the AD group.

In the MWM experiment, mice with AD displayed longer escape latencies on days 4 and 5 compared to the sham and OM-MSCs-Exos groups (Figure 2B), suggesting impaired spatial learning ability in the AD group, which improved following OM-MSCs-Exos treatment. In a probe test conducted on day 7 of the spatial learning experiment, after removing the hidden platform, the number of crossings over the original platform location and the distance traveled within 90 seconds were used as indicators of spatial memory. Results showed that the OM-MSCs-Exos group traveled less distance than the AD group (Figure 2C), while the OM-MSCs-Exos group made more platform crossings compared to the AD group (Figure 2D), further substantiating the contention that OM-MSCs-Exos treatment enhances spatial memory.

Moreover, the time spent in the target quadrant (Figure 2E) was significantly longer in the OM-MSCs-Exos group than in the AD group, substantiating the cognitive improvement induced by OM-MSCs-Exos. To explore the underlying mechanism for that, ELISA was used to measure Aβ1-42 levels in hippocampal tissue (Figure

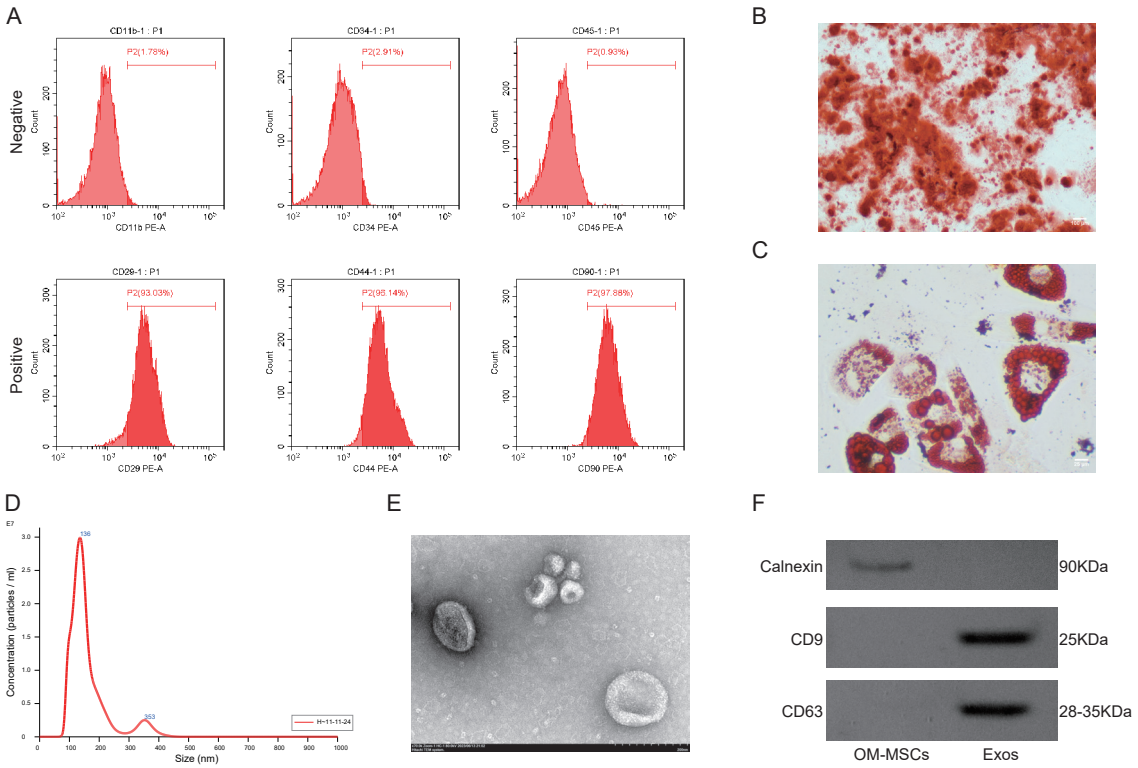


Figure 1. Characterization of OM-MSCs and OM-MSCs-Exos. (A) Flow cytometry analysis of surface markers on OM-MSCs. **(B)** Alizarin Red S staining. **(C)** Oil Red O staining. **(D)** Nanoparticle tracking analysis. **(E)** Transmission electron microscopy. **(F)** Western blot analysis of exosome marker expression.

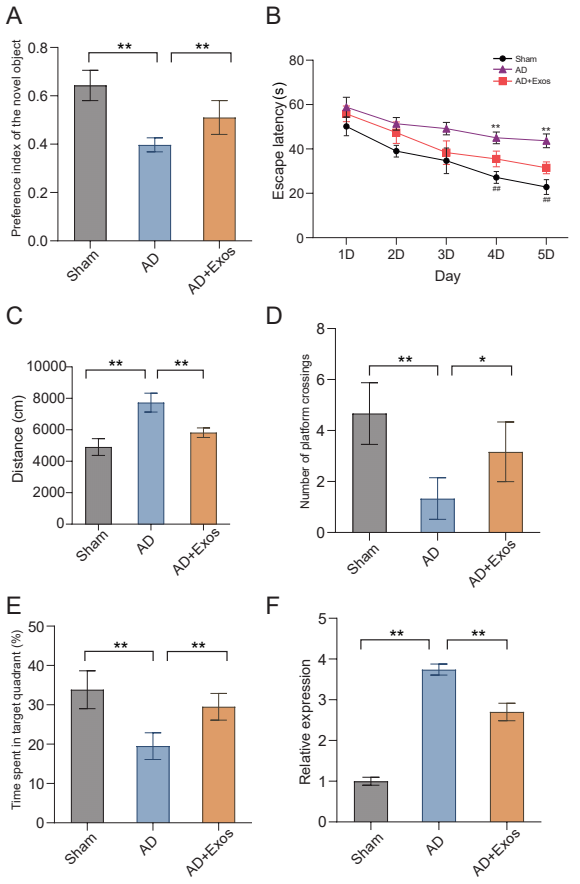


Figure 2. OM-MSCs-Exos improve spatial learning and memory in mice with AD. (A) Novel object recognition test. **(B)** Escape latency in the Morris Water Maze (MWM) (*: Sham *vs* AD; #: AD *vs* AD+Exos). **(C)** Distance traveled in the platform quadrant during the spatial exploration test. **(D)** Number of platform crossings during the spatial exploration test. **(E)** Time spent in the platform quadrant during the spatial exploration test. **(F)** ELISA analysis of changes in Aβ1-42 levels in the hippocampal tissue of mice with AD. $^{*}P < 0.05$, $^{**}P < 0.01$ and $^{\#\#}P < 0.01$.

2F). Results indicated that Aβ1-42 levels were lower in the OM-MSCs-Exos group compared to the AD group, suggesting that OM-MSCs-Exos treatment reduces Aβ1-42 deposition.

In summary, compared to mice with AD that were treated with PBS, OM-MSCs-Exos transplantation significantly improved spatial learning and memory, an effect likely associated with reduced Aβ1-42 accumulation.

3.3. OM-MSCs-Exos suppress neuroinflammation, ER stress, and neuronal loss

To examine the effects of OM-MSCs-Exos on neuroinflammation, ER stress, and neuronal loss in mice with AD, the morphology and number of Nissl bodies in brain tissue was first assessed using Nissl staining. Results revealed a significant reduction in the number and size of Nissl bodies in mice with AD. However, in mice with AD that were treated with OM-MSCs-Exos, the number of Nissl bodies markedly increased, indicating that exosomes effectively improve neuronal health (Figure 3A).

Activated microglia and astrocytes are key markers of neuroinflammation. IHC was used to detect the levels of expression of GFAP (an astrocyte marker) and Iba1 (a microglial marker) in the hippocampal region of mice with AD. Findings showed a significant increase in GFAP- and Iba1-positive cells in mice with AD. Following OM-MSCs-Exos treatment, however, the activation of these cells decreased substantially (Figure 3B-E). This suggests that OM-MSCs-Exos may mitigate neuroinflammation and protect neurons by suppressing astrocyte and microglial activation.

Next, TUNEL staining was used to detect apoptosis in mouse brain tissue. Results indicated that OM-MSCs-Exos treatment significantly reduced the number of apoptotic cells in the brains of mice with AD (Figure 3F, G). WB analysis further revealed that OM-MSCs-Exos treatment upregulated the expression of the anti-apoptotic protein Bcl-2 while downregulating the pro-apoptotic proteins Bax and cleaved caspase-3 (Figure 3H-J). These findings indicate that OM-MSCs-Exos may inhibit apoptosis *via* the Bcl-2/Bax signaling pathway, thereby reducing neuronal damage in the mouse model of AD.

To further validate the anti-inflammatory effects of OM-MSCs-Exos, ELISA was used to measure the levels of pro-inflammatory cytokines TNF-α, IL-1β, and IL-6 in the brain tissue of mice with AD. Results showed that OM-MSCs-Exos treatment significantly reduced the concentrations of these cytokines (Figure 3K), suggesting that OM-MSCs-Exos may alleviate neuroinflammation by suppressing inflammatory cytokine release.

ER stress plays a critical role in the progression of AD. WB analysis was used to assess the levels of expression of the key ER stress-related proteins CHOP, GRP78, and ATF6. Results indicated that OM-MSCs-Exos treatment downregulated the expression of GRP78, ATF6, and CHOP (Figure 3L, M). This appropriate downregulation of GRP78, ATF6, and CHOP likely contributes to alleviating neuronal damage caused by excessive ER stress.

3.4. OM-MSCs-Exos ameliorate Aβ1-42-induced ER stress, neuroinflammation, and apoptosis *in vitro*

Aβ1-42 is a key pathogenic factor in AD, and its aggregation triggers cellular stress responses, including ER stress, neuroinflammation, and apoptosis. To validate the bioactivity of OM-MSCs-Exos, their protective effects on cell viability, apoptosis, inflammation, and ER stress was further evaluated in a cellular model of AD.

First, the uptake of OM-MSCs-Exos by SH-SY5Y cells was observed using fluorescence microscopy. After co-incubating PKH67-labeled OM-MSCs-Exos with SH-SY5Y cells for 6 hours, clear green fluorescence signals

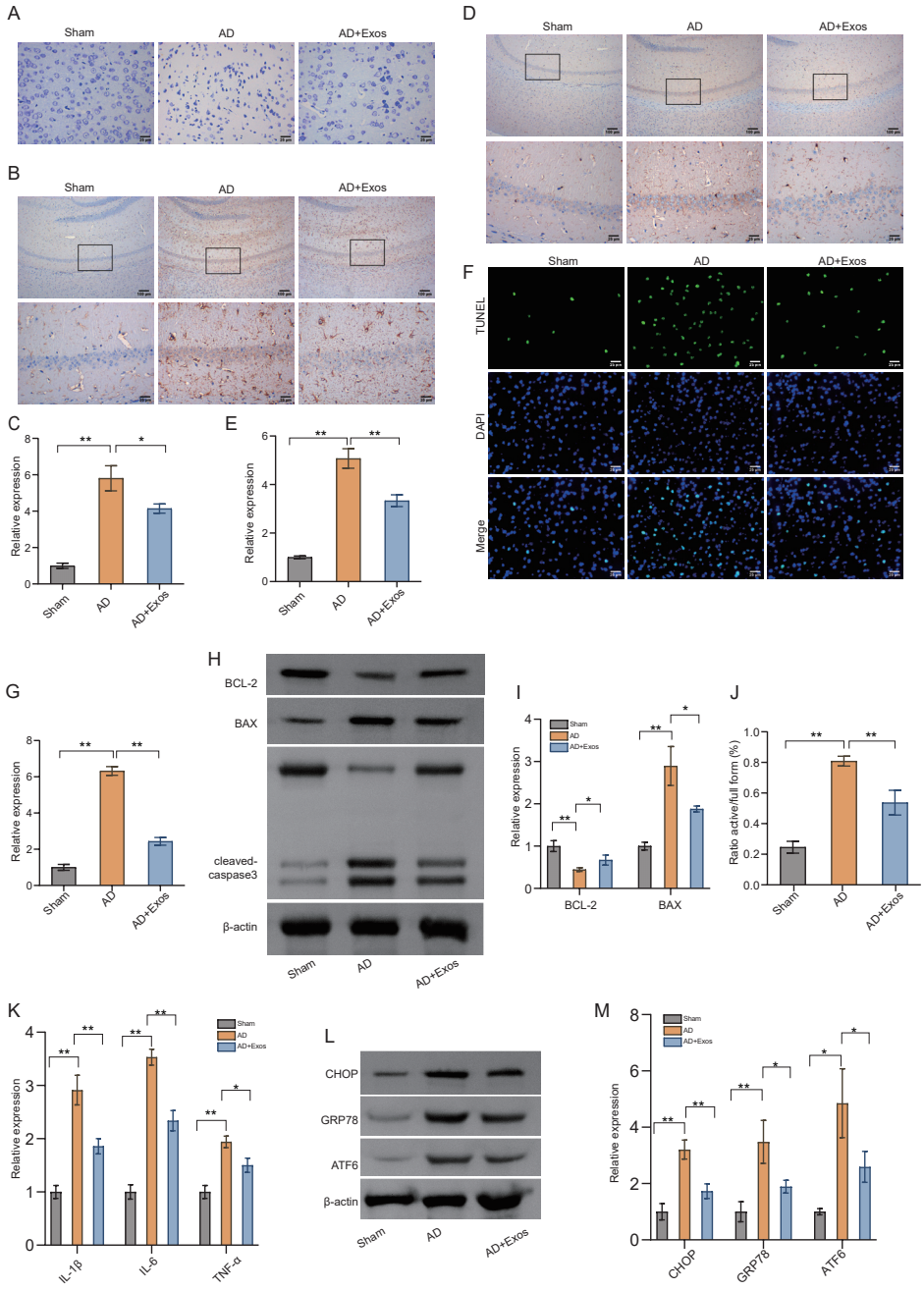


Figure 3. Neuroprotective effects of OM-MSCs-Exos in mice with AD. (A) Nissl staining of mouse brain tissue. **(B, C)** Immunohistochemical (IHC) staining and analysis of GFAP expression in the hippocampal region of mice. **(D, E)** IHC staining and analysis of Iba1 expression in the hippocampal region of mice. **(F, G)** TUNEL staining of mouse brain tissue. **(H-J)** Western blot analysis of levels of Bcl-2, Bax, and cleaved caspase-3 expression in mouse brain tissue. **(K)** ELISA analysis of TNF-α, IL-1β, and IL-6 levels in mouse brain tissue. **(L, M)** Western blot analysis of levels of CHOP, GRP78, and ATF6 expression in mouse brain tissue. $*P < 0.05$ and $**P < 0.01$.

were detected around the cell nuclei, confirming effective internalization of the exosomes (Figure 4A). Cell viability and proliferation were then assessed using the CCK-8 assay. Compared to the Aβ1-42-treated group, the OM-MSCs-Exos-treated group had significantly higher SH-SY5Y cell survival rates (Figure 4B), indicating that OM-MSCs-Exos effectively mitigate Aβ1-42-induced cell damage and promote cell survival.

To further evaluate apoptosis, Annexin V/PI dual staining flow cytometry was performed. The results showed that the total apoptosis rate in the OM-MSCs-Exos-treated group was lower than that in the Aβ1-42 group, with a notable reduction in early apoptotic cells (Figure 4C, D). WB analysis of levels of Bcl-2, Bax, and cleaved caspase-3 expression revealed that OM-MSCs-Exos treatment upregulated Bcl-2 while downregulating Bax and cleaved caspase-3 (Figure 4E-G). These findings further support the hypothesis that OM-MSCs-Exos inhibit apoptosis *via* the Bcl-2/Bax signaling pathway, exerting a neuroprotective effect.

Additionally, to investigate the suppression of Aβ1-42-induced neuroinflammatory cytokine release by OM-

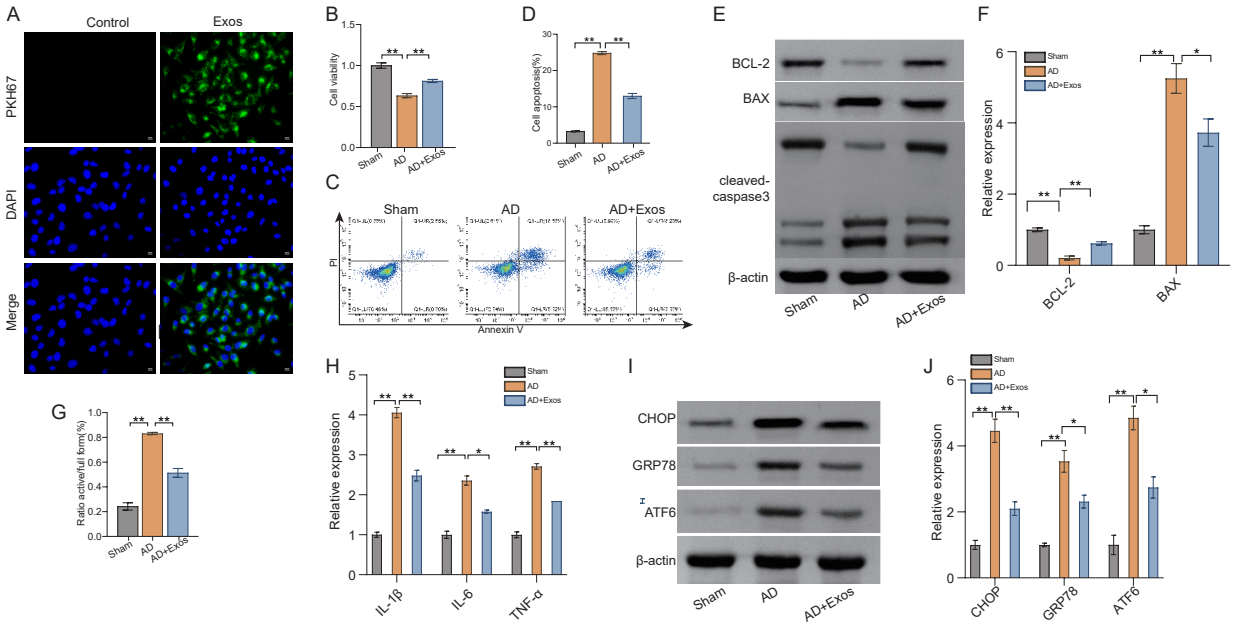


Figure 4. Neuroprotective effects of OM-MSCs-Exos in an Aβ1-42-induced SH-SY5Y cell model. (A) Fluorescence images of PKH67-labeled OM-MSCs-Exos co-cultured with SH-SY5Y cells. **(B)** CCK-8 assay to examine cell viability. **(C, D)** Annexin V/PI double staining flow cytometry to detect apoptosis. **(E-G)** Western blot analysis of Bcl-2, Bax, and cleaved caspase-3 levels. **(H)** ELISA assessment of TNF-α, IL-1β, and IL-6 levels in cell supernatant. **(I, J)** Western blot analysis of levels of CHOP, GRP78, and ATF6 expression. **P* < 0.05 and ***P* < 0.01.

MSCs-Exos, ELISA was used to measure TNF-α, IL-1β, and IL-6 levels in SH-SY5Y cell culture medium. Results indicated that OM-MSCs-Exos treatment inhibited the release of these inflammatory cytokines compared to the Aβ1-42 group (Figure 4H), indicating that OM-MSCs-Exos suppress Aβ1-42-induced neuroinflammation.

To further assess the regulation of ER stress by OM-MSCs-Exos, WB analysis was used to measure the levels of expression of the key ER stress-related proteins GRP78, ATF6, and CHOP. Results showed that OM-MSCs-Exos treatment downregulated GRP78, ATF6, and CHOP expression (Figure 4I, J). These findings suggest that OM-MSCs-Exos may alleviate Aβ1-42-induced ER stress by modulating the ER stress response, thereby protecting cells from stress-induced damage.

3.5. OM-MSCs-Exos mitigate Aβ1-42-induced ER stress, neuroinflammation, and apoptosis *via* LRP1

Although the above experiments confirmed the neuroprotective effects of OM-MSCs-Exos in models of AD, the precise molecular mechanisms remain unclear. This study further explored the role of key proteins in OM-MSCs-Exos in neuronal repair and cognitive recovery. Proteomic analysis revealed that OM-MSCs-Exos are enriched in exosome-related proteins. Xun *et al.* (*17*) identified 304 proteins secreted by OM-MSCs that are closely associated with neurotrophy, cell growth, differentiation, apoptosis, inflammation, and neuronal repair. Notably, OM-MSCs-Exos express LRP1. Previous studies have shown that LRP1 can suppress neuroinflammation and ER stress (*18-20*). Thus, the hypothesis was that OM-MSCs-Exos have a neuroprotective effect in the cellular model of AD *via* LRP1.

First, WB analysis revealed that LRP1 expression was downregulated in the cellular model of AD, while OM-MSCs-Exos treatment partially restored LRP1 expression (Figure 5A-B). These findings suggest that LRP1 may be involved in the neuroprotective effects mediated by OM-MSCs-Exos. To further validate this, LRP1 expression in OM-MSCs was silenced using siRNA, which reduced LRP1 mRNA expression in OM-MSCs-Exos (Figure 5C), confirming the efficacy of siRNA silencing. In the cellular model of AD, OM-MSCs-Exos treatment increased cell viability (Figure 5D), whereas the si-LRP1 intervention group exhibited reduced cell viability, indicating that LRP1 is linked to the neuroprotective effects of OM-MSCs-Exos. Furthermore, Annexin V/PI flow cytometry analysis showed that OM-MSCs-Exos treatment reduced apoptosis, while the si-LRP1 intervention group exhibited a significantly higher apoptosis rate (Figure 5E, F), corroborating LRP1's role in apoptosis regulation.

Additional WB analysis demonstrated that OM-MSCs-Exos treatment upregulated the anti-apoptotic protein Bcl-2 and downregulated the pro-apoptotic proteins Bax and cleaved caspase-3. However, si-LRP1 intervention reversed these protective effects, as evinced by increased Bax and cleaved caspase-3 expression and decreased Bcl-2 expression (Figure 5G-I). Moreover, ELISA results showed that OM-MSCs-Exos treatment significantly inhibited the release of the Aβ1-42-induced pro-inflammatory cytokines TNF-α, IL-1β, and IL-

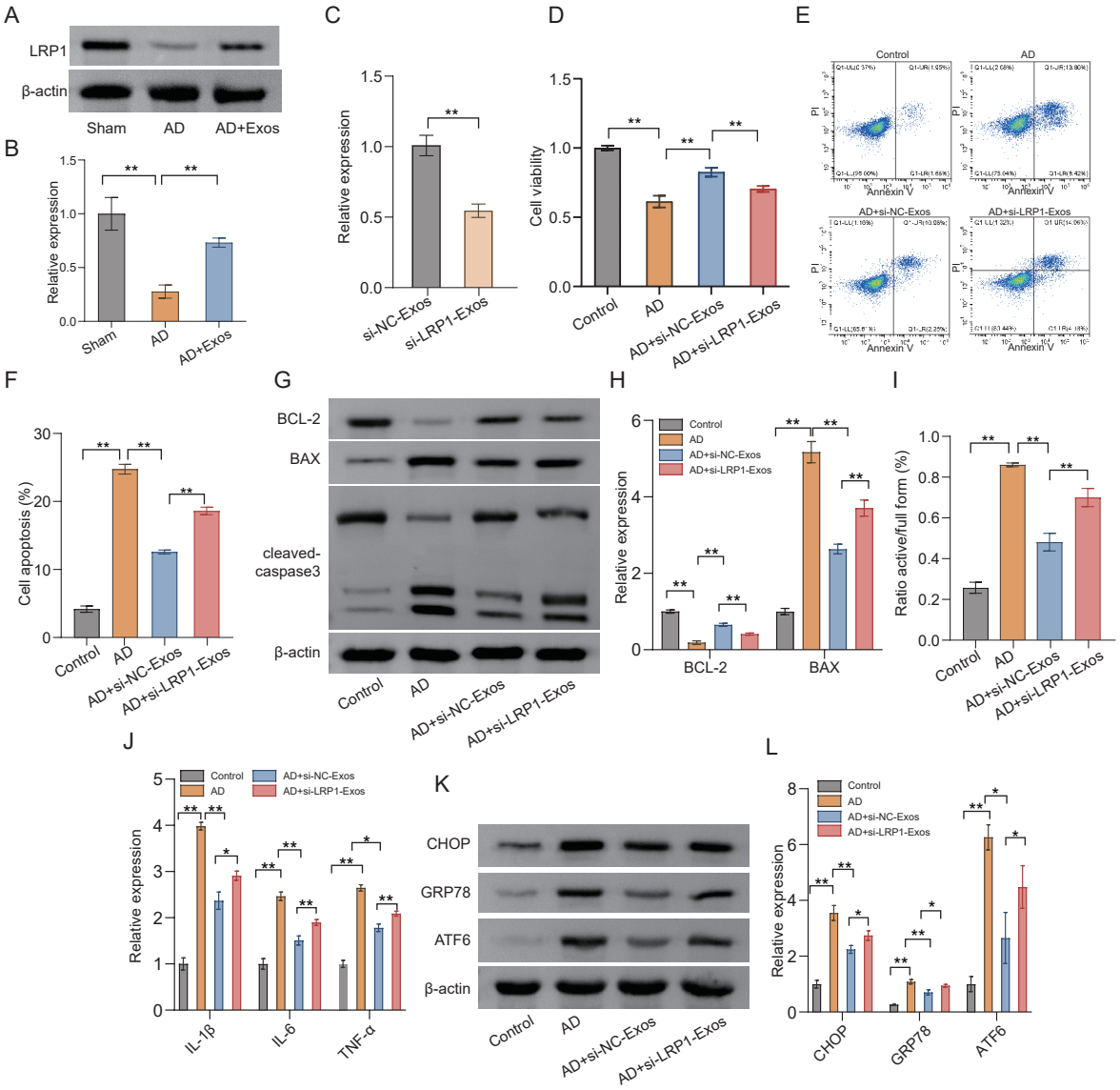


Figure 5. OM-MSCs-Exos alleviate Aβ1-42-induced SH-SY5Y cell damage by regulating LRP1 expression. (A, B) Western blot analysis of changes in LRP1 expression. **(C)** qPCR quantification of LRP1 mRNA expression. **(D)** CCK-8 assay for cell viability. **(E, F)** Annexin V/PI flow cytometry analysis of cell apoptosis. **(G-I)** Western blot analysis of levels of Bcl-2, Bax, and cleaved caspase-3 expression. **(J)** ELISA assessment of TNF-α, IL-1β, and IL-6 levels. **(K, L)** Western blot analysis of GRP78, ATF6, and CHOP expression. *$P < 0.05$ and **$P < 0.01$.

6, whereas the si-LRP1 intervention group exhibited increased cytokine release (Figure 5J).

Finally, to explore LRP1's role in ER stress, the levels of expression of the ER stress-related proteins CHOP, GRP78, and ATF6 were examined. OM-MSCs-Exos treatment downregulated GRP78, ATF6, and CHOP expression, indicating that OM-MSCs-Exos mitigate cellular damage by regulating ER stress. In contrast, the si-LRP1 intervention group showed an increased ER stress response compared to the OM-MSCs-Exos-treated group (Figure 5K, L).

3.6. Downregulation of LRP1 attenuates OM-MSCs-Exos-mediated improvements in spatial learning and memory in mice with AD

To further investigate whether LRP1 is a key molecule in OM-MSCs-Exos-mediated cognitive improvement *in vivo*, C57BL/6 mice were randomly divided into five groups: a sham group, an AD group, an AD+OM-MSCs-Exos group, an AD+si-NC-Exos group, and an AD+si-LRP1-Exos group. After 4 weeks of treatment, cognitive behavioral tests, including the NORT and MWM, were performed (Figure 6A-E).

In the NORT, mice with AD spent less time exploring novel objects, indicating memory impairment due to Aβ1-42 treatment. OM-MSCs-Exos treatment enhanced object recognition ability compared to the AD group. However, the si-LRP1-Exos group spent less time exploring, suggesting that LRP1 suppression diminished the memory-enhancing effects of OM-MSCs-Exos (Figure 6A).

In the MWM test, differences in escape latency were observed on days 4 and 5. Compared to the AD

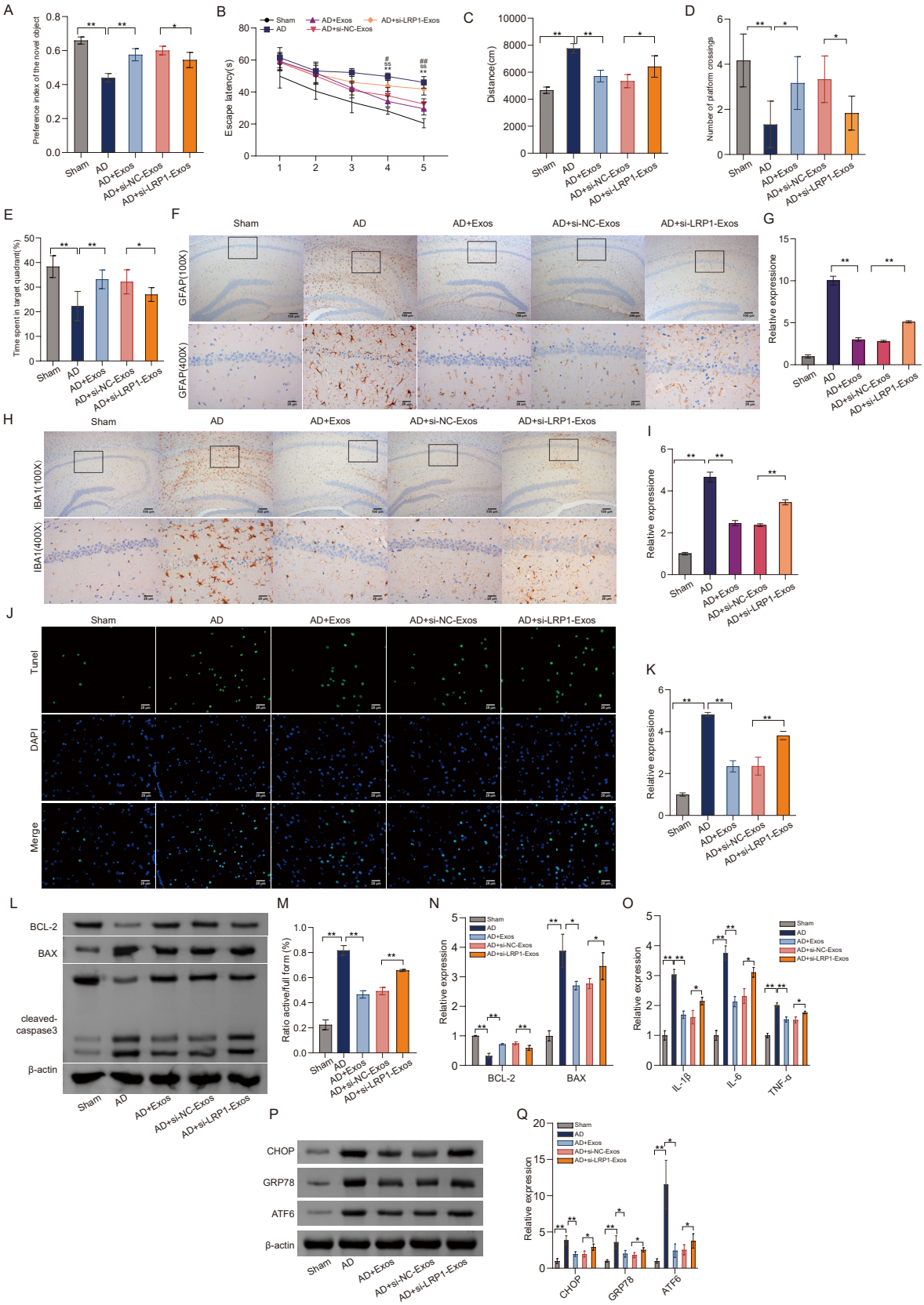


Figure 6. Downregulation of LRP1 attenuates the cognitive enhancement brought about by OM-MSCs-Exos in mice with AD. (A) Novel object recognition test. **(B)** Escape latency in the MWM (*: Sham *vs* AD; #: AD *vs* AD+Exos; s:AD+si-NC-Exos *vs* AD+si-LRP1-Exos). **(C)** Distance traveled in the platform quadrant during the spatial exploration test. **(D)** Number of platform crossings during the spatial exploration test. **(E)** Time spent in the platform quadrant during the spatial exploration test. **(F, G)** IHC analysis of GFAP expression in the hippocampal region of mice. **(H, I)** IHC analysis of Iba1 expression in the hippocampal region of mice. **(J, K)** TUNEL detection of apoptosis in brain tissue. **(L-N)** Western blot analysis of Bcl-2, Bax, and cleaved caspase-3 expression. **(O)** ELISA analysis of TNF-α, IL-1β, and IL-6 levels in mouse brain tissue. **(P-Q)** Western blot analysis of GRP78, ATF6, and CHOP expression in mouse brain tissue. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{\#}P < 0.01$, $^{\#\#}P < 0.05$, and $^{ss}P < 0.01$.

group, OM-MSCs-Exos-treated mice had shorter escape latencies, indicating improved spatial learning ability. However, the si-LRP1-Exos group had slightly longer escape latencies, suggesting that LRP1 suppression diminished the ability of OM-MSCs-Exos to enhance spatial learning (Figure 6B). The number of crossings over the original platform location (Figure 6D) and distance traveled (Figure 6C) within 90 seconds revealed that the OM-MSCs-Exos group crossed the platform more frequently and traveled less distance than the AD group, confirming the efficacy of OM-MSCs-Exos in enhancing spatial memory. In contrast, the si-LRP1-Exos group made fewer crossings and traveled a longer distance compared to the si-NC-Exos group, indicating that LRP1 suppression attenuated the spatial memory improvements mediated by OM-MSCs-Exos.

Furthermore, OM-MSCs-Exos-treated mice spent more time in the target quadrant than the AD group, reflecting enhanced spatial memory and target recognition ability. Compared to the si-NC-Exos group, the si-LRP1-Exos group spent less time in the target quadrant, further corroborating the critical role of LRP1 in OM-MSCs-Exos-mediated cognitive protection (Figure 6E).

3.7. Downregulation of LRP1 attenuates the effects of OM-MSCs-Exos on cognitive improvement and amelioration of ER stress, neuroinflammation, and apoptosis *in vivo*

To investigate the neuroprotective role of LRP1 in AD pathogenesis, IHC was used to analyze the levels of expression of the microglial marker Iba1 and the astrocytic marker GFAP in the hippocampal tissue of mice with AD (Figure 6F-I). Results showed that the OM-MSCs-Exos group had significantly reduced Iba1 and GFAP expression compared to the AD group, indicating that OM-MSCs-Exos effectively suppress neuroinflammatory responses. However, Iba1 and GFAP expression increased in the si-LRP1-Exos group compared to the si-NC-Exos group, suggesting that LRP1 suppression diminished the anti-inflammatory effects of OM-MSCs-Exos.

TUNEL staining was used to detect apoptosis in mouse brain tissue (Figure 6J, K). Results revealed that the OM-MSCs-Exos group had significantly fewer apoptotic cells than the AD group, indicating that OM-MSCs-Exos effectively inhibit neuronal apoptosis in mice with AD. However, the number of apoptotic cells increased significantly in the si-LRP1-Exos group compared to the si-NC-Exos group, further demonstrating that LRP1 suppression diminished the neuroprotective effects of OM-MSCs-Exos.

To further evaluate changes in the apoptosis signaling pathway, WB analysis was performed to assess the levels of Bcl-2, Bax, and cleaved caspase-3 expression. Compared to the AD group, the OM-MSCs-Exos group displayed upregulated Bcl-2 expression and downregulated Bax and cleaved caspase-3 expression, indicating that OM-MSCs-Exos effectively suppress the activation of the apoptosis signaling pathway (Figure 6L-N). However, levels of Bax and cleaved caspase-3 expression significantly increased in the si-LRP1-Exos group while Bcl-2 expression decreased. These findings further suggest that LRP1 suppression diminishes the inhibitory effect of OM-MSCs-Exos on the apoptosis signaling pathway.

ELISA was used to measure the levels of pro-inflammatory cytokines TNF-α, IL-1β, and IL-6 in mouse brain tissue. Results showed that the OM-MSCs-Exos group had significantly lower levels of these cytokines compared to the AD group (Figure 6O), indicating that OM-MSCs-Exos effectively suppress inflammation. However, the levels of these inflammatory cytokines increased in the si-LRP1-Exos group compared to the OM-MSCs-Exos group, further demonstrating that LRP1 suppression diminished the anti-inflammatory effects of exosomes.

Finally, to investigate changes in ER stress, WB analysis was used to assess the levels of expression of the ER stress-related proteins CHOP, GRP78, and ATF6. The OM-MSCs-Exos group had significantly reduced CHOP, GRP78, and ATF6 expression compared to the AD group (Figure 6P, Q), indicating that OM-MSCs-Exos effectively alleviate ER stress in mice with AD. However, levels of CHOP, GRP78, and ATF6 expression increased again in the si-LRP1-Exos group, suggesting that LRP1 suppression diminished the protective effects of OM-MSCs-Exos on ER stress alleviation.

4. Discussion

AD is a complex neurodegenerative disorder typically characterized by a progressive decline in cognitive function and neuronal dysfunction (*21*). The pathological progression of AD is driven by multiple factors, with neuroinflammation and ER stress being considered key contributors (*22,23*). The hallmark pathological features of AD are the deposition of Aβ and the abnormal phosphorylation of tau protein. These aberrant changes not only directly impair neurons but also exacerbate disease progression by triggering neuroinflammatory responses (*24*). In the brains of patients with AD, persistent low-level neuroinflammation is commonly observed, with the activation of microglia and astrocytes being considered the primary sources of inflammation (*25*). Elevated levels of pro-inflammatory cytokines, such as TNF-α, IL-6, and IL-1β, contribute to neuronal damage and dysfunction, thereby promoting neurodegeneration (*26*).

ER stress also plays an important role in the pathological development of AD (*27*). When misfolded proteins accumulate in the ER, the unfolded protein response (UPR) is activated to restore ER homeostasis.

However, prolonged or severe ER stress can trigger pro-apoptotic signaling pathways, particularly in neurons, making it a significant contributor to cell death in AD (*28*). Aβ deposition is recognized as a key inducer of ER stress, disrupting ER function and thereby amplifying neuronal damage (*29*). In the current study, levels of the pro-inflammatory cytokines IL-1β, IL-6, and TNF-α were significantly elevated in a mouse model of AD, and ER-associated proteins such as CHOP, GRP78, and ATF6 were markedly upregulated. Immunofluorescence staining further revealed a significant increase in the activation of Iba1-positive microglia and GFAP-positive astrocytes in the hippocampal region. These findings suggest that suppressing neuroinflammation and excessive ER stress may offer potential benefits in terms of alleviating AD-related neurological deficits.

In recent years, MSCs have shown considerable promise in the treatment of various neurodegenerative diseases (*30*). Traditionally, MSCs are derived from mesodermal tissues such as bone marrow and adipose tissue. These MSCs display some neuroprotective effects in terms of neural repair, but there are several with their clinical use, including limited neurogenic differentiation potential, ethical concerns, and difficulties in crossing the BBB (*31*). In contrast, OM-MSCs, which originate from ectodermal tissue, demonstrate greater neurogenic differentiation potential, positioning them as a more promising candidate for neural repair. Moreover, OM-MSCs can be isolated from human nasal mucosa *via* minimally invasive procedures, offering excellent biosafety and circumventing the ethical controversies associated with other stem cell sources. However, there are obstacles to the direct transplantation of OM-MSCs, such as challenges in crossing the BBB and potential tumorigenic risks. Unlike stem cells, exosomes derived from stem cells cannot self-replicate, eliminating the risk of tumor formation associated with stem cell transplantation (*21*). Consequently, exosomes have emerged as a safer and more effective therapeutic strategy. Exosomes are vesicles enclosed by a lipid bilayer, distinguished by the presence of tetraspanins (CD9, CD81, and CD63), ALG-2-interacting protein X (Alix), and tumor susceptibility gene 101 protein (TSG101) on their membrane surface (*32*). They can carry a variety of bioactive molecules, including proteins, nucleic acids, and lipids, and are capable of crossing the BBB to deliver therapeutic agents, making them a focal point in research on neurodegenerative diseases such as AD, stroke, and traumatic brain injury (*33,34*).

The current study successfully isolated and characterized OM-MSCs-Exos. TEM, NTA, and WB confirmed that these exosomes exhibit a typical lipid bilayer structure and express exosome markers. Further immunofluorescence staining demonstrated that PKH67-labeled OM-MSCs-Exos were effectively internalized by cells. These results indicate that OM-MSCs-Exos significantly inhibited the activation of microglia and astrocytes in the hippocampus, reduced the release of pro-inflammatory cytokines in brain tissue, and lowered the levels of expression of ER stress-related proteins. These effects were closely associated with significant improvements in cognitive function and reduced neuronal apoptosis in mice with AD. Conventional ultracentrifugation was used during the isolation of OM-MSCs-Exos. However, the purity and yield of exosomes may be influenced by factors such as cell culture conditions and centrifugation parameters, which could potentially interfere with subsequent results (*35,36*). Moreover, fluorescence labeling experiments have demonstrated that OM-MSCs-Exos can be taken up by cells, but their specific sites of action and underlying mechanisms within the cells remain unclear. To address this, more in-depth studies, possibly utilizing techniques such as confocal microscopy, may need to be conducted.

To further elucidate the neuroprotective mechanisms of OM-MSCs-Exos, the proteomic profile of OM-MSCs-Exos as reported by Xun *et al.* (*17*) was analyzed, and enrichment of LRP1 in these exosomes was noted. Previous studies have demonstrated that LRP1 plays a critical role in regulating the activation of microglia and astrocytes as well as modulating inflammatory responses (*18,37,38*). Additionally, LRP1 influences ER stress-related signaling pathways, impacting cell survival and function (*39*). The current findings revealed that LRP1 expression decreased significantly in a mouse model of AD, while ER stress and pro-inflammatory cytokine levels increased markedly. Following OM-MSCs-Exos treatment, however, LRP1 expression in mouse brain tissue increased, accompanied by an alleviation of both ER stress and inflammatory responses. Further experiments that used siRNA to silence LRP1 expression in OM-MSCs demonstrated that the neuroprotective effects of OM-MSCs-Exos in suppressing ER stress and neuroinflammation significantly diminished, underscoring the central role of LRP1 in the neuroprotective mechanisms of OM-MSCs-Exos treatment. However, the pathogenesis of AD is highly complex, involving the interplay of multiple cell types and signaling pathways (*40*). In addition to the known mechanisms, OM-MSCs-Exos may have an effect through other as yet unidentified pathways.

From a clinical translation perspective, the long-term stability, dose-dependency, and immunogenicity of exosomes warrant further evaluation. Although this study has demonstrated that OM-MSCs-Exos improve cognitive function in AD, their therapeutic efficacy across different stages of AD remains unclear. Future research could utilize multi-omics approaches, such as single-cell transcriptomics and protein interactomics, to further clarify the mechanisms of OM-MSCs-Exos and optimize their therapeutic potential through pharmacological enhancement.

5. Conclusion

This study provides the first evidence that OM-MSCs-Exos can significantly enhance cognitive function in mice with AD. The neuroprotective effects of OM-MSCs-Exos appear to be mediated through the suppression of neuroinflammatory responses, attenuation of microglial and astrocytic activation, and reduction in the expression of pro-inflammatory cytokines and ER stress markers, thereby mitigating neuronal damage. Furthermore, LRP1 may play a key role in these protective mechanisms. These findings provide novel insights into the molecular pathways underlying the therapeutic potential of OM-MSCs-Exos in the treatment of AD. However, despite these promising results, further research is required to evaluate the long-term stability, dose dependency, and immunogenicity of OM-MSCs-Exos, as well as to validate their clinical applicability and safety.

Acknowledgements

The authors sincerely thank Professor Wei Tang (University of Tokyo, Japan) and Professor Peipei Song (National Center for Global Health and Medicine, Japan) for their invaluable guidance and insightful discussions, which have significantly contributed to the study.

Funding: This work was supported by grants from National Natural Science Foundation of China (82460268), the Hainan Province Clinical Medical Research Center (No. LCYX202410, LCYX202309), and the Hainan Province Postdoctoral Research Project (403254).

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Scheltens P, De Strooper B, Kivipelto M, Holstege H, Chetelat G, Teunissen CE, Cummings J, van der Flier WM. Alzheimer's disease. Lancet. 2021; 397:1577-1590.
2. Botella Lucena P, Heneka MT. Inflammatory aspects of Alzheimer's disease. Acta Neuropathol. 2024; 148:31.
3. Lee JH, Won SM, Suh J, Son SJ, Moon GJ, Park UJ, Gwag BJ. Induction of the unfolded protein response and cell death pathway in Alzheimer's disease, but not in aged Tg2576 mice. Exp Mol Med. 2010; 42:386-394.
4. Unterberger U, Hoftberger R, Gelpi E, Flicker H, Budka H, Voigtlander T. Endoplasmic reticulum stress features are prominent in Alzheimer disease but not in prion diseases *in vivo*. J Neuropathol Exp Neurol. 2006; 65:348-357.
5. Hoozemans JJ, van Haastert ES, Nijholt DA, Rozemuller AJ, Eikelenboom P, Scheper W. The unfolded protein response is activated in pretangle neurons in Alzheimer's disease hippocampus. Am J Pathol. 2009; 174:1241-1251.
6. Ajoolabady A, Lindholm D, Ren J, Pratico D. ER stress and UPR in Alzheimer's disease: Mechanisms, pathogenesis, treatments. Cell Death Dis. 2022; 13:706.
7. Chiti F, Dobson CM. Protein misfolding, Amyloid formation, and human disease: A summary of progress over the last decade. Annu Rev Biochem. 2017; 86:27-68.
8. Andrzejewska A, Dabrowska S, Lukomska B, Janowski M. Mesenchymal stem cells for neurological disorders. Adv Sci (Weinh). 2021; 8:2002944.
9. Liu S, Fan M, Xu JX, Yang LJ, Qi CC, Xia QR, Ge JF. Exosomes derived from bone-marrow mesenchymal stem cells alleviate cognitive decline in AD-like mice by improving BDNF-related neuropathology. J Neuroinflammation. 2022; 19:35.
10. Lin F, Chen W, Zhou J, Zhu J, Yao Q, Feng B, Feng X, Shi X, Pan Q, Yu J, Li L, Cao H. Mesenchymal stem cells protect against ferroptosis *via* exosome-mediated stabilization of SLC7A11 in acute liver injury. Cell Death Dis. 2022; 13:271.
11. Guo M, Yin Z, Chen F, Lei P. Mesenchymal stem cell-derived exosome: A promising alternative in the therapy of Alzheimer's disease. Alzheimers Res Ther. 2020; 12:109.
12. Tian J, Zhu Q, Zhang Y, Bian Q, Hong Y, Shen Z, Xu H, Rui K, Yin K, Wang S. Olfactory ecto-mesenchymal stem cell-derived exosomes ameliorate experimental colitis *via* modulating Th1/Th17 and Treg cell responses. Front Immunol. 2020; 11:598322.
13. Care IoLARCo, Animals UoL. Guide for the care and use of laboratory animals. US Department of Health and Human Services, Public Health Service, National Institutes of Health, 1986.
14. Yu L, Wang S, Chen X, Yang H, Li X, Xu Y, Zhu X. Orientin alleviates cognitive deficits and oxidative stress in Abeta1-42-induced mouse model of Alzheimer's disease. Life Sci. 2015; 121:104-109.
15. Xu M, Huang H, Mo X, Zhu Y, Chen X, Li X, Peng X, Xu Z, Chen L, Rong S, Yang W, Liu S, Liu L. Quercetin-3-O-Glucuronide alleviates cognitive deficit and toxicity in Abeta(1-42)-induced AD-like mice and SH-SY5Y cells. Mol Nutr Food Res. 2021; 65:e2000660.
16. Chen HX, Liang FC, Gu P, Xu BL, Xu HJ, Wang WT, Hou JY, Xie DX, Chai XQ, An SJ. Exosomes derived from mesenchymal stem cells repair a Parkinson's disease model by inducing autophagy. Cell Death Dis. 2020; 11:288.
17. Xun C, Ge L, Tang F, Wang L, Zhuo Y, Long L, Qi J, Hu L, Duan D, Chen P, Lu M. Insight into the proteomic profiling of exosomes secreted by human OM-MSCs reveals a new potential therapy. Biomed Pharmacother. 2020; 131:110584.
18. Yang L, Liu CC, Zheng H, Kanekiyo T, Atagi Y, Jia L, Wang D, N'Songo A, Can D, Xu H, Chen XF, Bu G. LRP1 modulates the microglial immune response *via* regulation of JNK and NF-kappaB signaling pathways. J Neuroinflammation. 2016; 13:304.
19. Mantuano E, Henry K, Yamauchi T, Hiramatsu N, Yamauchi K, Orita S, Takahashi K, Lin JH, Gonias SL, Campana WM. The unfolded protein response is a major mechanism by which LRP1 regulates Schwann cell survival after injury. J Neurosci. 2011; 31:13376-13385.
20. Hamlin AN, Basford JE, Jaeschke A, Hui DY. LRP1 protein deficiency exacerbates palmitate-induced steatosis and toxicity in hepatocytes. J Biol Chem. 2016; 291:16610-16619.
21. Ma YN, Hu X, Karako K, Song P, Tang W, Xia Y. Exploring the multiple therapeutic mechanisms and challenges of mesenchymal stem cell-derived exosomes in Alzheimer's disease. Biosci Trends. 2024; 18:413-430.

22. Nagar P, Sharma P, Dhapola R, Kumari S, Medhi B, HariKrishnaReddy D. Endoplasmic reticulum stress in Alzheimer's disease: Molecular mechanisms and therapeutic prospects. Life Sci. 2023; 330:121983.
23. Santos LE, Ferreira ST. Crosstalk between endoplasmic reticulum stress and brain inflammation in Alzheimer's disease. Neuropharmacol. 2018; 136:350-360.
24. Leng F, Edison P. Neuroinflammation and microglial activation in Alzheimer disease: Where do we go from here? Nat Rev Neurol. 2021; 17:157-172.
25. Hu X, Ma YN, Xia Y. Association between abnormal lipid metabolism and Alzheimer's disease: New research has revealed significant findings on the APOE4 genotype in microglia. Biosci Trends. 2024; 18:195-197.
26. Firdous SM, Khan SA, Maity A. Oxidative stress-mediated neuroinflammation in Alzheimer's disease. Naunyn Schmiedebergs Arch Pharmacol. 2024; 397:8189-8209.
27. Singh R, Kaur N, Choubey V, Dhingra N, Kaur T. Endoplasmic reticulum stress and its role in various neurodegenerative diseases. Brain Res. 2024; 1826:148742.
28. Lim D, Tapella L, Dematteis G, Genazzani AA, Corazzari M, Verkhratsky A. The endoplasmic reticulum stress and unfolded protein response in Alzheimer's disease: A calcium dyshomeostasis perspective. Ageing Res Rev. 2023; 87:101914.
29. Ekundayo BE, Obafemi TO, Adewale OB, Obafemi BA, Oyinloye BE, Ekundayo SK. Oxidative stress, endoplasmic reticulum stress and apoptosis in the pathology of Alzheimer's disease. Cell Biochem Biophys. 2024; 82:457-477.
30. Giovannelli L, Bari E, Jommi C, Tartara F, Armocida D, Garbossa D, Cofano F, Torre ML, Segale L. Mesenchymal stem cell secretome and extracellular vesicles for neurodegenerative diseases: Risk-benefit profile and next steps for the market access. Bioact Mater. 2023; 29:16-35.
31. Rahimi Darehbagh R, Seyedoshohadaei SA, Ramezani R, Rezaei N. Stem cell therapies for neurological disorders: Current progress, challenges, and future perspectives. Eur J Med Res. 2024; 29:386.
32. Kalluri R, LeBleu VS. The biology, function, and biomedical applications of exosomes. science. 2020; 367:eaau6977.
33. Mot YY, Moses EJ, Mohd Yusoff N, Ling KH, Yong YK, Tan JJ. Mesenchymal stromal cells-derived exosome and the roles in the treatment of traumatic brain injury. Cell Mol Neurobiol. 2023; 43:469-489.
34. Pratiwi DIN, Alhajlah S, Alawadi A, Hjazi A, Alawsi T, Almalki SG, Alsalamy A, Kumar A. Mesenchymal stem cells and their extracellular vesicles as emerging therapeutic tools in the treatment of ischemic stroke. Tissue Cell. 2024; 87:102320.
35. Xu K, Jin Y, Li Y, Huang Y, Zhao R. Recent progress of exosome isolation and peptide recognition-guided strategies for exosome research. Front Chem. 2022; 10:844124.
36. Yang D, Zhang W, Zhang H, Zhang F, Chen L, Ma L, Larcher LM, Chen S, Liu N, Zhao Q, Tran PHL, Chen C, Veedu RN, Wang T. Progress, opportunity, and perspective on exosome isolation - Efforts for efficient exosome-based theranostics. Theranostics. 2020; 10:3684-3707.
37. He Y, Ruganzu JB, Jin H, Peng X, Ji S, Ma Y, Zheng L, Yang W. LRP1 knockdown aggravates Abeta(1-42)-stimulated microglial and astrocytic neuroinflammatory responses by modulating TLR4/NF-kappaB/MAPKs signaling pathways. Exp Cell Res. 2020; 394:112166.
38. He Y, Ruganzu JB, Zheng Q, Wu X, Jin H, Peng X, Ding B, Lin C, Ji S, Ma Y, Yang W. Silencing of LRP1 exacerbates inflammatory response *via* TLR4/NF-kappaB/MAPKs signaling pathways in APP/PS1 transgenic mice. Mol Neurobiol. 2020; 57:3727-3743.
39. Wan C, Liu XQ, Chen M, Ma HH, Wu GL, Qiao LJ, Cai YF, Zhang SJ. Tanshinone IIA ameliorates Abeta transendothelial transportation through SIRT1-mediated endoplasmic reticulum stress. J Transl Med. 2023; 21:34.
40. Monteiro AR, Barbosa DJ, Remiao F, Silva R. Alzheimer's disease: Insights and new prospects in disease pathophysiology, biomarkers and disease-modifying drugs. Biochem Pharmacol. 2023; 211:115522.

Received February 8, 2025; Revised March 3, 2025; Accepted March 16, 2025.

Address correspondence to:
Ying Xia, Department of Neurosurgery, Haikou Affiliated Hospital of Central South University Xiangya School of Medicine, Haikou 570208, China.
E-mail: xiaying008@163.com

Released online in J-STAGE as advance publication March 18, 2025.

Original Article

DOI: 10.5582/bst.2024.01351

Development and validation of a nomogram model for predicting immune-mediated hepatitis in cancer patients treated with immune checkpoint inhibitors

Qianjie Xu[1,§], Xiaosheng Li[2,§], Yuliang Yuan[1,§], Zuhai Hu[3], Wei Zhang[1], Ying Wang[2], Ai Shen[4,*], Haike Lei[1,*]

[1] Chongqing Cancer Multi-omics Big Data Application Engineering Research Center, Chongqing University Cancer Hospital, Chongqing, China;
[2] Chongqing Key Laboratory of Translational Research for Cancer Metastasis and Individualized Treatment, Chongqing University Cancer Hospital, Chongqing, China;
[3] Department of Health Statistics, School of Public Health, Chongqing Medical University, Chongqing, China;
[4] Hepatobiliary Pancreatic Cancer Center, Chongqing University Cancer Hospital, Chongqing, China.

SUMMARY: Immune checkpoint inhibitors (ICIs) have been widely used in various types of cancer, but they have also led to a significant number of adverse events, including ICI-induced immune-mediated hepatitis (IMH). This study aimed to explore the risk factors for IMH in patients treated with ICIs and to develop and validate a new nomogram model to predict the risk of IMH. Detailed information was collected between January 1, 2020, and December 31, 2023. Univariate logistic regression analysis was used to assess the impact of each clinical variable on the occurrence of IMH, followed by stepwise multivariate logistic regression analysis to determine independent risk factors for IMH. A nomogram model was constructed based on the results of the multivariate analysis. The performance of the nomogram model was evaluated *via* the area under the receiver operating characteristic curve (AUC), calibration curves, decision curve analysis (DCA), and clinical impact curve (CIC) analysis. A total of 216 (8.82%) patients developed IMH. According to stepwise multivariate logistic analysis, hepatic metastasis, the TNM stage, the WBC count, LYM, ALT, TBIL, ALB, GLB, and ADA were identified as risk factors for IMH. The AUC for the nomogram model was 0.817 in the training set and 0.737 in the validation set. The calibration curves, DCA results, and CIC results indicated that the nomogram model had good predictive accuracy and clinical utility. The nomogram model is intuitive and straightforward, making it highly suitable for rapid assessment of the risk of IMH in patients receiving ICI therapy in clinical practice. Implementing this model enables early adoption of preventive and therapeutic strategies, ultimately reducing the likelihood of immune-related adverse events (IRAEs), and especially IMH.

Keywords: ICIs, IMH, influencing factors, risk model

1. Introduction

Cancer has become the second leading cause of death worldwide, resulting in approximately 9.6 million deaths and 182.8 million years of life lost (*1*). Immune checkpoint inhibitors (ICIs) have revolutionized the treatment of malignancies and have been used to treat many different types of cancer. ICIs enhance the body's immune response to cancer cells by blocking negative regulatory factors expressed on immune cells or tumor cells through a unique mechanism. ICIs mainly consist of cytotoxic T-lymphocyte antigen-4 (CTLA-4) inhibitors, programmed cell death protein (PD)-1 inhibitors (PD-1), and PD-ligand 1 inhibitors (PD-L1) (*2,3*). However, the expanded indications for ICIs and their increased use has led to the discovery of a large number of adverse events associated with ICIs, termed immune-related adverse events (IRAEs), in clinical settings.

Studies have shown that IRAEs are caused by an overactive immune response, primarily in the skin, endocrine, hepatic, and pulmonary systems (*4,5*). Owing to its unique immune characteristics, the liver is one of the organs most susceptible to the effects of tumor immunotherapy. Hepatitis caused by ICI treatment is commonly referred to as ICI-induced immune-mediated hepatitis (IMH). Research indicates that IMH is the third most common IRAE, with an incidence ranging from 5% to 10% (*6*), followed by skin toxicity (44%-68%)

and gastrointestinal adverse reactions (35%-50%) (*7*). Although most cases of IMH are asymptomatic and can be appropriately controlled with supportive therapy and corticosteroids (*8*), improper diagnosis or management can lead to immunotherapy failure, acute liver failure, and death, especially in patients with chronic liver disease (*9,10*). Previous studies have suggested that IMH accounts for a high proportion of fatal IRAEs. According to data from a global database on fatal IRAEs, 124 of the 613 reported deaths were associated with IMH (*11*). Similarly, a study by Wang *et al.* found that among 21 melanoma patients who died from IRAEs, 5 deaths (23.8%) were caused by IMH (*12*).

The mechanisms by which ICIs cause IMH have yet to be fully elucidated, and data on the clinical risk factors for IMH are very limited. Most importantly, there is no clinical model with which to accurately assess the risk of IMH in patients. This makes the prevention and management of IMH in patients receiving ICI therapy particularly challenging in clinical practice. Therefore, identifying the risk factors associated with IMH and predicting the risk of IMH in patients receiving ICI therapy is highly clinically important. This information will help clinicians quickly identify high-risk IMH patients and manage them individually, ultimately reducing the incidence of IMH at its source. In current clinical research, nomogram models are widely used to explore risk factors and predict risk (*13*). Su *et al.* recruited 2,281 consecutive patients with hepatitis B-related hepatocellular carcinoma from four tertiary hospitals in China from April 2011 to March 2022 (*14*). They utilized multivariate Cox regression to establish a nomogram risk prediction model, which accurately predicted the mortality risk of patients and effectively identified high-risk patients.

Therefore, the current study aimed to investigate the risk factors for IMH in patients receiving ICI therapy and to develop and validate a new nomogram model to predict the risk of IMH. Ultimately, this model will guide personalized strategies to prevent IMH.

2. Materials and Methods

2.1. Subjects and inclusion and exclusion criteria

This study collected relevant information from 2,663 cancer patients who received ICI therapy at Chongqing University Cancer Hospital from January 1, 2020, to December 31, 2023. The collected data include basic patient information such as sex, age, and body mass index (BMI); tumor-related data such as liver metastasis, TNM stage, and Karnofsky performance status (KPS); and biomarker data such as lymphocyte (LYM), white blood cell (WBC), and platelet (PLT) counts and alanine transaminase (ALT), aspartate transaminase (AST), albumin (ALB), globulin (GLB), total bilirubin (TBIL), alkaline phosphatase (AKP), adenosine deaminase (ADA), C-reactive protein (CRP), and β2-microglobulin (β2-MG) levels. The definition of IMH in this study was based on the Guidelines for the Diagnosis and Treatment of Autoimmune Hepatitis (2021) (*15,16*). The diagnostic criteria include elevated serum aminotransferase levels, positive serum autoantibodies, elevated IgG levels, and characteristic histological changes in the liver, while excluding other potential causes. All blood tests were conducted in the laboratory of Chongqing University Cancer Hospital. Informed consent was obtained from each patient. This study was conducted in accordance with the guidelines outlined in the Declaration of Helsinki and received ethical approval from the Ethics Committee of Chongqing University Cancer Hospital.

The inclusion criteria for this study were as follows: *i*) age ≥ 18 years; *ii*) hospitalized at least once; and *iii*) received ICI therapy with any of three inhibitors: CTLA-4, PD-1, or PD-L1. The exclusion criteria were as follows: *i*) missing critical pathological data such as ALT, AST, PLT, ALB, GLB, and ADA; *ii*) death within 48 hours of admission; *iii*) chronic hepatitis due to other causes, such as viral hepatitis, alcoholic hepatitis, nonalcoholic fatty liver disease, drug-induced liver disease, schistosomiasis, and other parasitic infections causing liver disease; *iv*) concurrent autoimmune liver diseases, such as primary biliary cirrhosis, primary sclerosing cholangitis, and overlap syndromes; *v*) primary liver cancer; and *vi*) combined use of two or more inhibitors. After applying the inclusion and exclusion criteria, 2,448 patients were included in the model, as shown in Figure 1.

2.2. Model construction and validation

Patients meeting the inclusion and exclusion criteria were randomly divided into a training cohort (n = 1,714) and a validation cohort (n = 734) at a 7:3 ratio. This process was implemented *via* the "caret" package in R software, with a fixed random seed number used throughout the study. In the training cohort, univariate logistic regression analysis was used to assess the impact of each clinical variable on the occurrence of IMH in patients. Variables with a p value < 0.2 in the results were then included in stepwise multivariate logistic regression analysis to identify independent factors influencing the development of IMH. A nomogram model was constructed on the basis of these results. The performance of the nomogram model was validated in the validation cohort. The discriminative ability of the nomogram was assessed *via* the area under the receiver operating characteristic curve (AUC). Calibration curves were generated *via* the bootstrap method with 1,000 resamples to validate the predictive accuracy of the nomogram in both the training and validation sets. The Hosmer-Lemeshow test was used to evaluate the goodness of fit of the nomogram model. Decision curve analysis (DCA) and clinical impact curve (CIC) analysis were performed *via* the "rmda" package

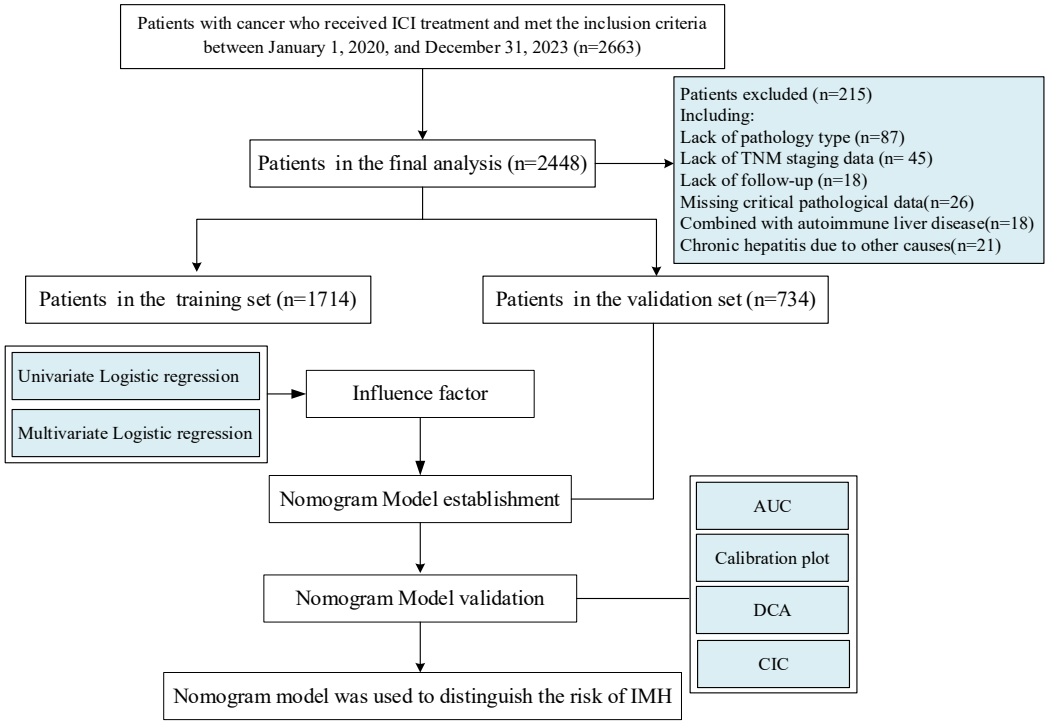


Figure 1. Flow chart for patients enrolled in the final study cohorts.

to evaluate the practical value of the nomogram model in clinical settings.

2.3. Statistical analysis

For normally distributed data, the mean ± SD was used for description, and a *t*-test was used for comparison. For nonnormally distributed data, the median (M), P25, and P75 were used for description, and nonparametric tests were used for comparison. Categorical data are expressed as frequencies and percentages, and comparisons were made *via* the chi-square test. Missing data were filled in with the "mice" package. All of the statistical analyses were performed using R version 4.1.2, and statistical significance was defined as a two-tailed p value < 0.05.

3. Results

3.1. Clinical characteristics of subjects

After applying the inclusion and exclusion criteria, 2,448 study subjects were retained, 216 (8.82%) of whom developed IMH. The median age of the included subjects was 59.00 years, and 75.82% were male. Additionally, more than half of the patients had a BMI of 18.5-23.9, were in TNM stage IV, and did not have hepatic metastases, with proportions of 57.92%, 66.79%, and 87.95%, respectively. Significant differences were observed between patients with and without IMH in terms of age, Karnofsky Performance Scale (KPS) score, and all hematological indices except for C-reactive protein (CRP) (p values for all < 0.05). Specifically, patients without IMH were older and had a higher WBC and PLT count, whereas those with IMH had higher levels of LYM, ALT, AST, ALB, GLB, TBIL, AKP, ADA, and β2-MG. Details are shown in Table 1.

3.2. Characteristics of the training and validation cohorts

This study used random sampling to allocate the 2,448 patients into the training and validation cohorts, with 1,714 patients in the training cohort and 734 patients in the validation cohort, while maintaining a 7:3 split ratio. As shown in Table 2, there were no significant differences between the training and validation cohorts (p values for all > 0.05).

3.3. Factors influencing the development of IMH

Univariate and stepwise multivariate logistic regression analyses were performed with the training cohort to investigate the factors affecting the occurrence of IMH in patients receiving ICI therapy. The detailed results are shown in Table 3. According to stepwise multivariate logistic analysis, several factors were found to increase the likelihood of developing IMH to varying degrees: hepatic metastasis, TNM stage IV, WBC, LYM, ALT, TBIL, ALB, GLB, and ADA. Hepatic metastasis and TNM stage IV disease in particular were associated with the greatest increase in IMH risk, with a 63% and 61% greater likelihood than in patients without hepatic metastasis or those with TNM stage III disease. Interestingly, age was a protective factor according to univariate and stepwise multivariate logistic analyses.

Table 1. Demographic and clinical characteristics of patients with or without IMH

Variables	Total (*n* = 2,448)	No-IMH (*n* = 2,232)	IMH (*n* = 216)	*p* value
Age (years)	59.05 ± 10.89	59.47 ± 10.79	54.69 ± 11.04	< 0.001
Sex				0.649
Female	592 (24.18)	543 (91.72)	49 (8.28)	
Male	1,856 (75.82)	1,689 (91.00)	167 (9.00)	
BMI (%)				0.284
18.5-23.9	1,418 (57.92)	1,292 (91.11)	126 (8.98)	
24-27.9	685 (27.98)	628 (91.68)	57 (8.32)	
≥ 28	129 (5.27)	112 (86.82)	17 (13.18)	
< 18.5	216 (8.82)	200 (92.59)	16 (7.41)	
TNM (%)				0.037
III	813 (33.21)	727 (89.42)	86 (10.58)	
IV	1,635 (66.79)	1,505 (92.05)	130 (7.95)	
KPS	82.40 ± 7.63	82.27 ± 7.62	83.75 ± 7.55	0.006
Hepatic metastases (%)				0.440
No	2,153 (87.95)	1,959 (90.99)	194 (9.01)	
Yes	295 (12.05)	273 (92.54)	22 (7.46)	
WBC (10^9/L)	7.53 ± 4.22	7.68 ± 4.30	5.96 ± 2.88	< 0.001
PLT (10^9/L)	216.61 ± 90.08	217.72 ± 89.50	205.21 ± 95.38	0.051
LYM (10^9/L)	1.05 ± 0.53	1.03 ± 0.52	1.21 ± 0.56	< 0.001
ALT (U/L)*	23.00 [15.00, 39.00]	22.00 [15.00, 36.00]	36.75 [20.00, 73.00]	< 0.001
AST (U/L)*	24.00 [18.00, 35.50]	23.25 [18.00, 34.00]	40.00 [23.00, 63.78]	< 0.001
ALB (g/L)	37.10 ± 5.41	37.03 ± 5.43	37.85 ± 5.19	0.032
GLB (g/L)	30.70 ± 6.60	30.43 ± 6.48	33.47 ± 7.22	< 0.001
TBIL (μmol/L)*	9.01 [6.58, 12.59]	8.84 [6.52, 12.15]	11.09 [7.81, 16.41]	< 0.001
AKP (U/L)*	88.00 [71.00, 113.62]	86.00 [70.00, 110.05]	110.00 [82.75, 165.25]	< 0.001
ADA (U/L)	10.75 ± 5.70	10.32 ± 5.28	15.15 ± 7.69	< 0.001
CRP (mg/L)*	7.54 [2.66, 37.88]	7.56 [2.66, 36.36]	7.20 [2.53, 43.79]	0.762
β2-MG (mg/L)	3.03 ± 1.28	3.00 ± 1.26	3.29 ± 1.48	0.002

Note: *Expressed as the median (M) [P25, P75].

Table 2. Clinical characteristics of the training and validation cohorts

Variables	Training cohort (*n* = 1,714)	Validation cohort (*n* = 734)	*p* value
Age (years)	58.97 ± 10.88	59.23 ± 10.91	0.583
Sex			0.471
Female	407 (23.75)	185 (25.20)	
Male	1,307 (76.25)	549 (74.80)	
BMI (%)			0.540
18.5-23.9	1,003 (58.52)	415 (56.54)	
24-27.9	465 (27.13)	220 (29.97)	
≥ 28	91 (5.31)	38 (5.18)	
< 18.5	155 (9.04)	61 (8.31)	
TNM (%)			0.591
III	563 (32.85)	250 (34.06)	
IV	1,151 (67.15)	484 (65.94)	
KPS	82.23 ± 7.70	82.79 ± 7.44	0.094
Hepatic metastases (%)			0.689
No	1,504 (87.75)	649 (88.42)	
Yes	210 (12.25)	85 (11.58)	
WBC (10^9/L)	7.57 ± 4.14	7.44 ± 4.40	0.498
PLT (10^9/L)	217.82 ± 90.54	213.80 ± 88.99	0.312
LYM (10^9/L)	1.07 ± 0.54	1.01 ± 0.51	0.130
ALT (U/L)*	23.00 [15.00, 38.85]	23.00 [15.00, 41.54]	0.697
AST (U/L)*	24.00 [18.00, 35.00]	24.60 [18.00, 38.75]	0.127
ALB (g/L)	36.99 ± 5.45	37.37 ± 5.30	0.109
GLB (g/L)	30.62 ± 6.51	30.88 ± 6.82	0.386
TBIL (μmol/L)*	8.98 [6.54, 12.38]	9.05 [6.66, 12.88]	0.231
AKP (U/L)*	88.00 [71.00, 114.00]	86.40 [71.00, 111.00]	0.219
ADA (U/L)	10.70 ± 5.63	10.87 ± 5.86	0.492
CRP (mg/L)*	7.53 [2.58, 35.75]	7.59 [2.86, 39.64]	0.635
β2-MG (mg/L)	3.02 ± 1.27	3.05 ± 1.31	0.619

Note: *Expressed as the median (M) [P25, P75].

For each 1-year increase in age, the likelihood of developing IMH decreased by 4%.

3.4. Construction and evaluation of the nomogram model

Based on the results of stepwise multivariate logistic regression analysis, a nomogram model was constructed to predict the risk of IMH in patients receiving ICI therapy, as shown in Figure 2A. The total score is obtained by adding the scores for each factor and then locating the corresponding IMH risk level on the scale. According to the nomogram, TBIL has the greatest impact on predicting IMH risk, followed by GLB, WBC, and age. LYM, ALT, ALB, and ADA have a moderate impact on predicting IMH risk in patients with breast cancer undergoing chemotherapy.

The nomogram model had an AUC of 0.817 (95% CI: 0.782-0.852) in the training set and 0.737 (95% CI: 0.664-0.811) in the validation set, indicating good performance. The model effectively identified risk levels, with ROC curve results (Figure 2B) showing strong generalizability and effective risk identification for IMH in ICI patients. Similarly, the calibration curves (Figures 3A and 3B) revealed that all the points were close to the

Table 3. Logistic regression analysis of the risk factors for IMH in the training cohort

Variable	OR (Univariable)	OR (Stepwise - multivariable)
Age (years)	0.96 (0.94-0.97, $p < 0.001$)	0.96 (0.95-0.98, $p < 0.001$)
Sex		
Female		
Male	1.23 (0.82-1.83, $p = 0.314$)	
KPS	1.02 (0.99-1.04, $p = 0.110$)	
Hepatic metastases		
No		
Yes	2.41 (1.61-3.60, $p < 0.001$)	1.63 (1.02-2.60, $p = 0.040$)
TNM		
III		
IV	1.58 (1.08-2.31, $p = 0.018$)	1.61 (1.05-2.46, $p = 0.030$)
BMI		
18.5-23.9		
24-27.9	1.04 (0.71-1.51, $p = 0.844$)	
≥ 28	1.42 (0.75-2.70, $p = 0.284$)	
< 18.5	0.87 (0.46-1.63, $p = 0.662$)	
WBC (10^9/L)	1.04 (1.01-1.08, $p = 0.007$)	1.05 (1.01-1.09, $p = 0.008$)
PLT (10^9/L)	1.00 (0.99-1.00, $p = 0.066$)	
LYM (10^9/L)	1.63 (1.24-2.15, $p < 0.001$)	1.50 (1.06-2.11, $p = 0.021$)
ALT (U/L)*	1.01 (1.01-1.01, $p < 0.001$)	1.01 (1.01-1.01, $p < 0.001$)
AST (U/L)*	1.01 (1.01-1.01, $p < 0.001$)	
ALB (g/L)	1.03 (1.02-1.05, $p < 0.001$)	1.02 (1.01-1.03, $p = 0.006$)
GLB (g/L)	1.00 (0.99-1.00, $p < 0.001$)	
TBIL (μmol/L)*	1.03 (1.00-1.06, $p = 0.029$)	1.03 (1.01-1.07, $p = 0.041$)
AKP (U/L)*	1.07 (1.04-1.09, $p < 0.001$)	1.04 (1.01-1.06, $p = 0.012$)
ADA (U/L)	1.11 (1.08-1.14, $p < 0.001$)	1.07 (1.04-1.10, $p < 0.001$)
CRP (mg/L)*	1.00 (0.99-1.01, $p = 0.110$)	
β2-MG (mg/L)	1.13 (1.02-1.26, $p = 0.023$)	

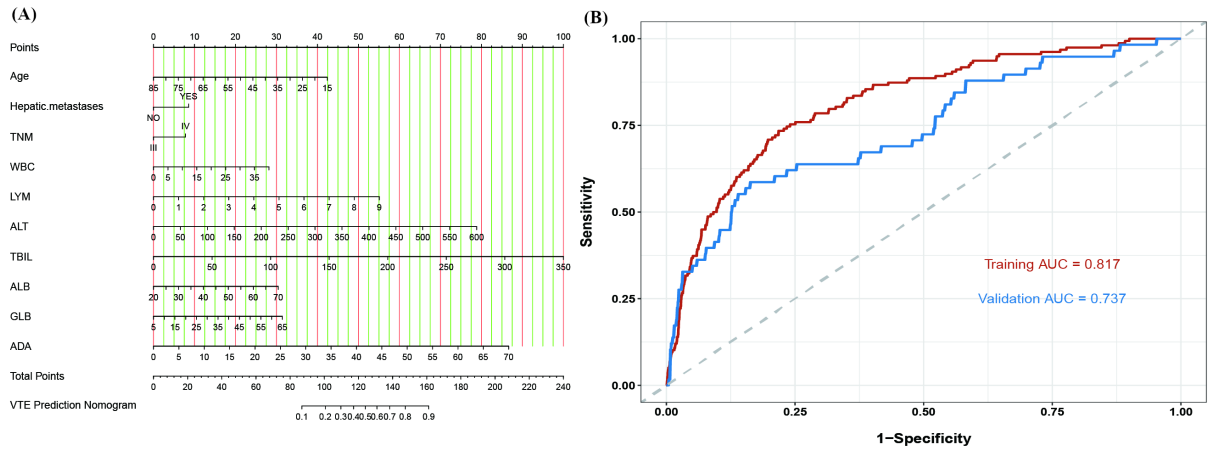


Figure 2. (A) Nomogram model for predicting IMH risk in ICI patients; (B) The ROC curve for the nomogram model.

diagonal line. The Hosmer-Lemeshow test showed that the *p* values were 0.270 and 0.857 for the training set and the validation set, respectively, indicating that the model fit well. These findings indicate that the nomogram model accurately predicts IMH risk in both the training and validation cohorts and performs excellently.

To evaluate the clinical benefit of the model, DCA was used, and the results are shown in Figures 4A and 4B. In the training cohort, the model indicated greater net benefit than the "all" and "none" lines at threshold probabilities between 1% and 39%, indicating clinical value. Similarly, the model indicated clinical applicability in the validation cohort at threshold probabilities between 1% and 35%. CIC (Figure 4C and 4D) revealed that the nomogram model can be used to indicate clinical benefits for any ICI patients.

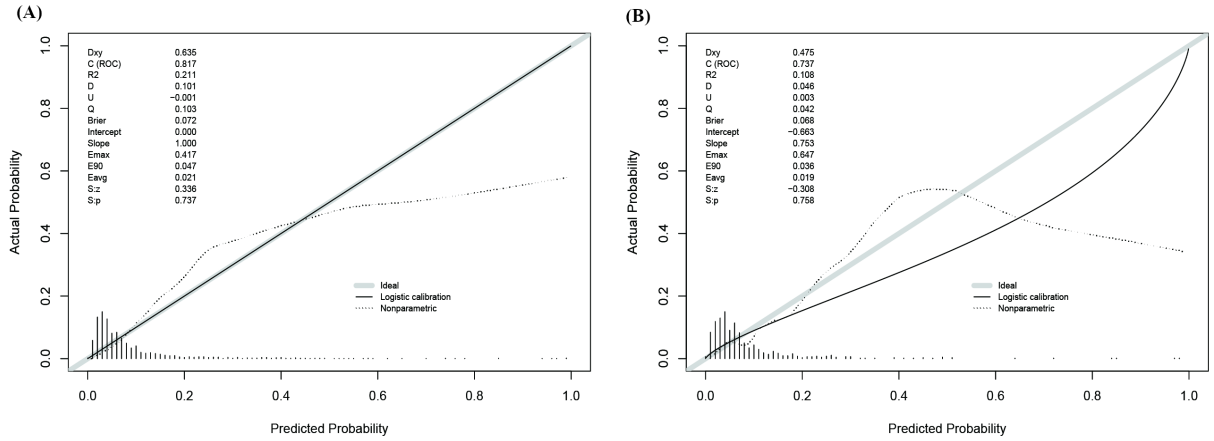


Figure 3. The calibration curves for the nomogram model. (A) training cohort; **(B)** validation cohort.

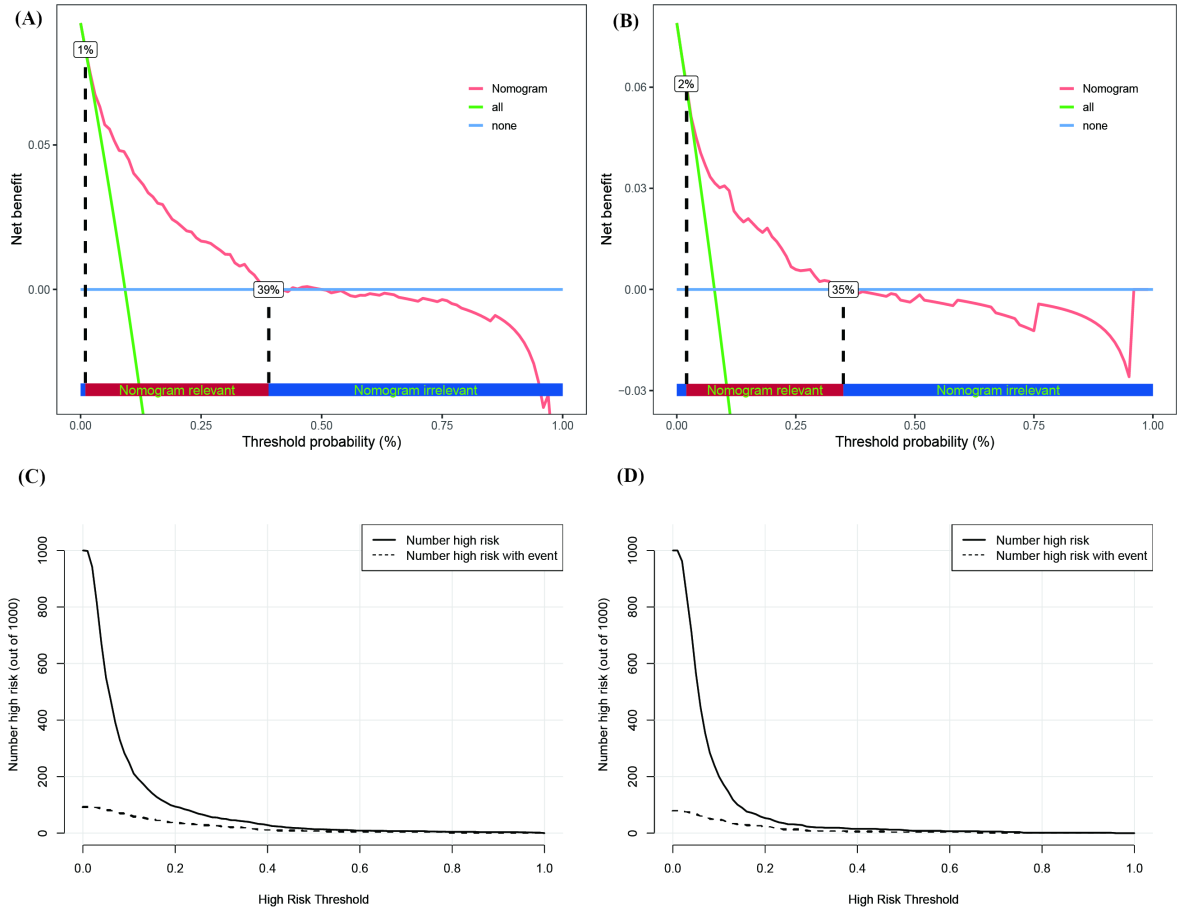


Figure 4. The DCA curves for the nomogram model. (A) training cohort; **(B)** validation cohort and CIC curves for the nomogram model; **(C)** training cohort; **(D)** validation cohort.

4. Discussion

Understanding the risk factors for developing IMH and constructing a model to predict the risk of IMH are crucial to guiding treatment interventions and improving patient outcomes. In this study, liver function data, demographic data, and relevant hematological indices of cancer patients treated with ICIs were integrated to investigate the risk factors for IMH. Based on these findings, a new risk prediction model was constructed using a nomogram to identify the risk of IMH in these patients. The nomogram model can quickly and accurately identify the risk levels of IMH without the need for invasive procedures such as liver biopsies. It is readable and practical, making it more suitable for clinical practice. This model can assist in personalized medical treatment and optimize the safety of ICIs in clinical practice.

Nomogram models play crucial roles in predicting the risk of liver-related diseases. They have been used to develop risk prediction tools for various liver diseases. For example, a nomogram model was constructed for hepatocellular carcinoma (HCC) patients treated with ICIs on the basis of clinical characteristics and the serum alpha-fetoprotein response to predict patient mortality risk (*17*). Similarly, nomogram models have been developed for patients with autoimmune hepatitis (AIH) to identify predictors of poor treatment response and advanced liver fibrosis and even to predict the risk of AIH without requiring a liver biopsy (*18*). However, no studies have proposed the use of a nomogram model to predict the risk of IMH. The current study is the first to utilize real-world data from hospitals to construct a nomogram model to predict the risk of IMH in patients receiving ICI treatment. Several studies have shown that nomogram models have greater predictive accuracy than other hepatitis risk assessment tools do. For example, Zhao *et al*. developed a nomogram model to predict acute liver failure (ALF) in patients with spontaneous rupture of hepatocellular carcinoma (SRHCC) with a high level of accuracy, achieving a C-index of 0.91 that was superior to those of the Child–Pugh and ALBI models (*19*). Similarly, Yang *et al.* constructed a nomogram model to predict 90-day mortality risk in patients with hepatitis B virus-related acute–chronic liver failure (HBV-ACLF) (*20*). This model outperformed the MELD score, Age-Bilirubin-International Normalized Ratio-Creatinine (ABIC) score, and Albumin-Bilirubin (ALBI) score in terms of prediction accuracy.

In the current study, patients with liver metastasis had a significantly increased risk of IMH, with a 1.52-fold greater risk than those without liver metastasis. However, the relationship between IMH and liver metastasis is complex. A retrospective case–control study by Storm *et al*. revealed that while liver metastasis was initially associated with an increased likelihood of IMH, this association was not significant after adjusting for covariates (*21*). Similarly, a systematic review and meta-analysis by Pan *et al*. reported that the association between liver metastasis and IMH was not statistically significant (OR: 1.47, 95% CI: 0.99-2.18; $p = 0.056$) (*11*). Therefore, the hypothesis is that liver metastasis may play a role in the occurrence of IMH. However, other factors, such as liver function and cancer staging, appear to have a greater impact on the risk of IMH in patients receiving ICI treatment. Patients with TNM stage IV disease have more severe cancer progression and often receive more intensive and frequent ICI treatment (*22*). This increases their risk of developing IMH compared to patients in other stages. Older patients tend to have reduced bodily activity and liver function compared to younger patients (*23,24*), which manifests as a lower risk of IMH in older patients in this study. Consequently, age emerged as a protective factor against IMH in the univariate and multivariate analyses.

The mechanism of IMH involves T-cell overactivation (*25*). Thus, WBCs play a crucial role in the development of IMH. Studies have shown that a small number of intrahepatic virus-specific cytotoxic T lymphocytes (CTLs) and recruited monocytes/macrophages can lead to chronic liver inflammation, increasing the risk of IMH (*26*). Additionally, T-cell-mediated immune mechanisms are related to hepatitis B virus (HBV) infection, and immunosuppressants can impair T-cell function, leading to immune-mediated hepatocyte lysis and reduced viral clearance, further increasing the risk of IMH (*27*). That said, Johnson *et al.* examined a mouse model of T-cell-mediated hepatitis induced by lymphocytic choriomeningitis virus (LCMV) infection and they reported that the severity of hepatitis was associated with the activity of cytotoxic T cells in the liver and spleen (*28*). These findings emphasize the role of T cells in liver injury and indicate that WBC dysfunction can exacerbate immune-mediated liver damage, increasing the risk of IMH.

Extensive research has shown that lymphocytes play a crucial role in IMH by mediating liver injury and disease progression (*29*). Platelets coordinate liver inflammation and damage through signaling factors such as TPL2 in iNKT cells, influencing immune-driven liver diseases and thereby increasing the risk of IMH (*30*). This finding is similar to the current study's findings. AST and TBIL are common markers of liver function and injury, and their elevation is a key feature of IMH, typically manifesting as elevated transaminases and other liver function abnormalities (*21*). Abnormal liver function often increases the risk of IMH.

Additionally, Zhang *et al*. reported that IMH is often accompanied by increased AST and TBIL levels (*31*). Owing to the unique immunological characteristics of the liver, the occurrence of IMH is often accompanied by elevated levels of ALB and GLB (*32*). In a study on the impact of ICIs on liver enzymes and attenuation, Park

et al. reported that patients treated with ICIs had higher ALB levels than those at the baseline did, indirectly indicating that IMH is accompanied by elevated ALB levels (*33*). ADA levels in body fluids reflect the activity of cellular immune responses. When IMH occurs, the liver's cellular immune response intensifies, leading to increased ADA levels (*34*). In the current study, this was evident in an increase in ADA levels of one unit, which increased the risk of IMH by 2%.

The current study had several innovative aspects. First, stringent inclusion and exclusion criteria were applied to exclude all unsuitable patients, and comprehensive characteristic data were thoroughly collected from patients in all age groups, ensuring the validity of data. Second, the direction and extent of the impact of each predictor on the occurrence of VTE in patients was investigated in the nomogram model, providing theoretical guidance for preventing VTE in clinical practice.

That said, this study had several limitations. First, this was a single-center study, with all patient data collected from one hospital. Therefore, the generalizability of the nomogram model is debatable. Future studies could involve multicenter collaboration to validate the model's performance using patient data from other centers. Second, this study was retrospective, so it has inherent limitations such as recall bias and recording bias. Finally, the impact of patients' imaging data or liver biopsy results on the risk of IMH occurrence was not considered, and these factors were not included in the model as predictors. Future research could incorporate detailed patient characteristics, such as imaging and liver biopsy data. This would increase the initial cost of the study, but it would undoubtedly enhance the model's performance and quality.

In conclusion, a model was developed to estimate the risk of IMH in cancer patients receiving ICI treatment. Based on the nomogram algorithm, this model is intuitive and straightforward, making it well-suited for assessment of the risk of patients developing IMH after ICI therapy in clinical practice. This nomogram model enables the prompt formulation of preventive and therapeutic strategies, ultimately reducing the likelihood of IRAEs, and particularly IMH. The practical use of this model in clinical settings could potentially enhance the quality of life of cancer patients.

Acknowledgements

The authors wish to thank all the participants, as no meaningful research could have been conducted without them.

Funding: None.

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Frick C, Rumgay H, Vignat J, Ginsburg O, Nolte E, Bray F, Soerjomataram I. Quantitative estimates of preventable and treatable deaths from 36 cancers worldwide: a population-based study. Lancet Glob Health. 2023; 11:e1700-e1712.
2. Patrinely JR Jr, McGuigan B, Chandra S, *et al*. A multicenter characterization of hepatitis associated with immune checkpoint inhibitors. Oncoimmunology. 2021; 10:1875639.
3. Kotanides H, Li Y, Malabunga M, *et al*. Bispecific targeting of PD-1 and PD-L1 enhances T-cell activation and antitumor immunity. Cancer Immunol Res. 2020; 8:1300-1310.
4. Pham JP, Joshua AM, da Silva IP, Dummer R, Goldinger SM. Chemotherapy in cutaneous melanoma: Is there still a role? Curr Oncol Rep. 2023; 25:609-621.
5. Regev A, Avigan MI, Kiazand A, *et al*. Best practices for detection, assessment and management of suspected immune-mediated liver injury caused by immune checkpoint inhibitors during drug development. J Autoimmun. 2020; 114:102514.
6. Kröner PT, Mody K, Farraye FA. Immune checkpoint inhibitor-related luminal GI adverse events. Gastrointest Endosc. 2019; 90:881-892.
7. Liu Z, Zhu Y, Xie H, Zou Z. Immune-mediated hepatitis induced by immune checkpoint inhibitors: Current updates and future perspectives. Front Pharmacol. 2023; 13:1077468.
8. Xu C, Chen YP, Du XJ, *et al.* Comparative safety of immune checkpoint inhibitors in cancer: Systematic review and network meta-analysis. BMJ. 2018; 363:k4226.
9. Vozy A, De Martin E, Johnson DB, Lebrun-Vignes B, Moslehi JJ, Salem JE. Increased reporting of fatal hepatitis associated with immune checkpoint inhibitors. Eur J Cancer. 2019; 123:112-115.
10. Yamamoto A, Yano Y, Ueda Y, Yasutomi E, Hatazawa Y, Hayashi H, Yoshida R, Asaji N, Shiomi Y, Tobimatsu K, Sakai A, Kodama Y. Clinical features of immune-mediated hepatotoxicity induced by immune checkpoint inhibitors in patients with cancers. J Cancer Res Clin Oncol. 2021; 147:1747-1756.
11. Pan J, Liu Y, Guo X, Bai Z, Levi Sandri GB, Méndez-Sánchez N, Qi X. Risk factors for immune-mediated hepatotoxicity in patients with cancer treated with immune checkpoint inhibitors: A systematic review and meta-analysis. Expert Opin Drug Saf. 2022; 21:1275-1287.
12. Wang DY, Salem JE, Cohen JV, *et al.* Fatal toxic effects associated with immune checkpoint inhibitors: A systematic review and meta-analysis. JAMA Oncol. 2018; 4:1721-1728.
13. Wang X, Liu H, Wang P, Wang Y, Yi Y, Li X. A nomogram for analyzing risk factors of poor treatment response in patients with autoimmune hepatitis. Eur J Gastroenterol Hepatol. 2024; 36:113-119.
14. Su K, Shen Q, Tong J, *et al.* Construction and validation of a nomogram for HBV-related hepatocellular carcinoma: A large, multicenter study. Ann Hepatol. 2023; 28:101109.
15. Chinese Society of Hepatology, Chinese Medical Association. Guidelines on the diagnosis and management of autoimmune hepatitis (2021). Zhonghua Nei Ke Za Zhi. 2021; 60:1038-1049. (in Chinese)

16. Muratori L, Lohse AW, Lenzi M. Diagnosis and management of autoimmune hepatitis. BMJ. 2023; 380:e070201.
17. Zhang Y, Shen H, Zheng R, Sun Y, Xie X, Lu M-D, Liu B, Huang G. Development and assessment of nomogram based on AFP response for patients with unresectable hepatocellular carcinoma treated with immune checkpoint inhibitors. Cancers (Basel). 2023; 15:5131.
18. Zhang Z, Wang J, Wang H, *et al.* An easy-to-use AIHF-nomogram to predict advanced liver fibrosis in patients with autoimmune hepatitis. Front Immunol. 2023; 14:1130362.
19. Zhao ZH, Jiang C, Wu QY, Lv GY, Wang M. Nomogram for estimation of acute liver failure risk in spontaneous ruptured hepatocellular carcinoma. J Hepatocell Carcinoma. 2023; 10:2223-2237.
20. Yang J, Xue R, Wu J, Jia L, Li J, Yu H, Zhu Y, Dong J, Meng Q. Development and validation of a nomogram for 90-day outcome in patients with hepatitis B virus-related acute-on-chronic liver failure. J Clin Transl Hepatol. 2022; 10:458-466.
21. Storm EM, Makraki D, Lin GI, Kennedy LC, Shah EE, Phipps AI, Liou IW, Hockenbery D, Grivas P, Khaki ARJCR. Immune related liver toxicity and potential risk factors: A case-control study. Cancer Research. 2022; 82:1981-1981.
22. Ulas EB, Dickhoff C, Schneiders FL, Senan S, Bahce I. Neoadjuvant immune checkpoint inhibitors in resectable non-small-cell lung cancer: A systematic review. ESMO Open. 2021; 6:100244.
23. Ninomiya K, Oze I, Kato Y, Kubo T, Ichihara E, Rai K, Ohashi K, Kozuki T, Tabata M, Maeda Y, Kiura K, Hotta K. Influence of age on the efficacy of immune checkpoint inhibitors in advanced cancers: A systematic review and meta-analysis. Acta Oncol. 2020; 59:249-256.
24. Yan X, Tian X, Wu Z, Han W. Impact of age on the efficacy of immune checkpoint inhibitor-based combination therapy for non-small-cell lung cancer: A systematic review and meta-analysis. Front Oncol. 2020; 10:1671.
25. Kitagataya T, Suda G, Nagashima K, *et al.* Prevalence, clinical course, and predictive factors of immune checkpoint inhibitor monotherapy-associated hepatitis in Japan. J Gastroenterol Hepatol. 2020; 35:1782-1788.
26. Russo FP, Zanetto A, Pinto E, Battistella S, Penzo B, Burra P, Farinati F. Hepatocellular carcinoma in chronic viral hepatitis: Where do we stand? Int J Mol Sci. 2022; 23:500
27. Shi Y, Zheng M. Hepatitis B virus persistence and reactivation. BMJ. 2020; 370:m2200.
28. Johnson DM, Khakhum N, Wang M, Warner NL, Jokinen JD, Comer JE, Lukashevich IS. Pathogenic and apathogenic strains of lymphocytic choriomeningitis virus have distinct entry and innate immune activation pathways. Viruses. 2024; 16:635.
29. Hora S, Wuestefeld T. Liver injury and regeneration: Current understanding, new approaches, and future perspectives. Cells. 2023; 12:2129.
30. Gu X, Chu Q, Ma X, Wang J, Chen C, Guan J, Ren Y, Wu S, Zhu H. New insights into iNKT cells and their roles in liver diseases. Front Immunol. 2022; 13:1035950.
31. Zhang T, Zhang Y, Zhang Y. Immune checkpoint inhibitor-induced immune-mediated hepatitis in a lung cancer patient undergoing long-term immunotherapy: A case report. Clinical Cancer Bulletin. 2022; 1:176-181.
32. Efe C, Kulkarni AV, Beretta-Piccoli B, *et al.* Liver injury after SARS-CoV-2 vaccination: Features of immune-mediated hepatitis, role of corticosteroid therapy and outcome. Hepatology. 2022; 76:1576-1586.
33. Park BC, Lee AXT, Ye F, Turker I, Johnson DB. Immune checkpoint inhibitors and their impact on liver enzymes and attenuation. BMC Cancer. 2022; 22:998.
34. Jeon JH, Thoudam T, Choi EJ, Kim MJ, Harris RA, Lee IK. Loss of metabolic flexibility as a result of overexpression of pyruvate dehydrogenase kinases in muscle, liver and the immune system: Therapeutic targets in metabolic diseases. J Diabetes Investig. 2021; 12:21-31.

Received October 30, 2024; Revised December 17, 2024; Accepted January 29, 2025.

[§]These authors contributed equally to this work.
**Address correspondence to:*
Haike Lei, Chongqing Cancer Multi-omics Big Data Application Engineering Research Center, Chongqing University Cancer Hospital, Chongqing 400030, China.
E-mail: tohaike@163.com

Ai Shen, Hepatobiliary Pancreatic Cancer Center, Chongqing University Cancer Hospital, Chongqing 400030, China.
E-mail: shenai200808@163.com

Released online in J-STAGE as advance publication February 1, 2025.

Original Article

DOI: 10.5582/bst.2024.01387

Risk stratification model for predicting distant metastasis after hepatectomy for hepatocellular carcinoma: A multi-institutional analysis

Mingda Wang[1,2,§,*], Shaodong Lv[1,§], Yongkang Diao[1,2,§], Jiahao Xu[1,2], Fujie Chen[3], Yuchen Li[3], Weimin Gu[4], Hong Wang[5], Yuze Yang[6], Yongyi Zeng[7], Yahao Zhou[8], Xianming Wang[9], Jie Li[10], Tinghao Chen[11], Yingjian Liang[12], Lanqing Yao[1,2], Lihui Gu[1,2], Han Wu[1,2], Xinfei Xu[1,2], Chao Li[1,2], Feng Shen[1,2], Tian Yang[1,2,3,*]

[1] Department of Hepatobiliary Surgery, Eastern Hepatobiliary Surgery Hospital, Second Military Medical University (Naval Medical University), Shanghai, China;
[2] Eastern Hepatobiliary Clinical Research Institute, Eastern Hepatobiliary Surgery Hospital, Second Military Medical University (Naval Medical University), Shanghai, China;
[3] Department of Graduate, Bengbu Medical College, Bengbu, China;
[4] The First Department of General Surgery, the Fourth Hospital of Harbin, Heilongjiang, China;
[5] Department of General Surgery, Liuyang People's Hospital, Hunan, China;
[6] Department of Hepatobiliary and Pancreatic Surgery, General Surgery Centre, First Hospital of Jilin University, Changchun, China;
[7] Department of Hepatobiliary Surgery, Mengchao Hepatobiliary Hospital, Fujian Medical University, Fujian, China;
[8] Department of Hepatobiliary Surgery, Pu'er People's Hospital, Yunnan, China;
[9] Department of General Surgery, First Affiliated Hospital of Shandong First Medical University & Shandong Provincial Qianfoshan Hospital, Shandong, China;
[10] Department of Hepatobiliary Surgery, Fuyang People's Hospital, Anhui, China;
[11] Department of General Surgery, Ziyang First People's Hospital, Sichuan, China;
[12] Department of Hepatobiliary Surgery, First Affiliated Hospital of Harbin Medical University, Heilongjiang, China.

SUMMARY: Distant metastasis after hepatectomy for hepatocellular carcinoma (HCC) significantly impairs long-term outcome. This study aimed to identify patterns, risk factors, and develop a prediction model for distant metastasis at first recurrence following HCC resection. This multi-center retrospective study included patients undergoing curative hepatectomy for HCC. Risk factors for distant metastasis were identified using Cox regression. A nomogram was constructed and validated using the concordance index (C-index) and calibration curves. Among 2,705 patients, 1,507 experienced recurrence, with 342 (22.7 per cent) developing distant metastasis. Common metastatic sites included extrahepatic vessels (36.2 per cent), lungs (26.0 per cent), and lymph nodes (20.8 per cent). Patients with distant metastasis had significantly worse 5-year overall survival compared to those with intrahepatic recurrence (9.1 versus 41.1 per cent, $p < 0.001$). Independent risk factors included preoperative tumor rupture, tumor size over 5.0 cm, multiple tumors, satellite nodules, macro- and microvascular invasion, narrow resection margin, and intraoperative blood transfusion. The nomogram demonstrated excellent discrimination (C-index > 0.85) and accurately stratified patients into three risk categories. In conclusion, distant metastasis at first recurrence following HCC resection was associated with poor prognosis. The proposed nomogram facilitates accurate prediction of distant metastasis, potentially informing personalized postoperative monitoring and interventions for high-risk patients.

Keywords: hepatocellular carcinoma, hepatectomy, distant metastasis, survival, recurrence

1. Introduction

Hepatocellular carcinoma (HCC) remains a formidable global health challenge, ranking as the 6th most common cancer and the 3rd leading cause of cancer-related mortality worldwide (*1,2*). Despite advancements in treatment modalities, hepatectomy persists as the primary curative option for a significant proportion of HCC patients. However, postoperative recurrence poses a substantial obstacle to long-term survival, with up to 70% of patients experiencing recurrence within 5 years (*3-8*). This high recurrence rate underscores critical need for a comprehensive understanding of recurrence patterns and associated risk factors to guide

postoperative surveillance and adjuvant treatment strategies (*9,10*).

Recurrence patterns in HCC can be broadly categorized as intrahepatic or extrahepatic (distant metastasis) (*11,12*). While intrahepatic recurrence is more common and often attributed to residual microscopic lesions or *de novo* tumors, distant metastasis represents a particularly aggressive form of disease progression (*13,14*). Distant metastasis not only signifies a more malignant tumor phenotype but also portends a markedly adverse prognosis, especially when occurring at the initial diagnosis of recurrence (*15*). Patients with distant metastasis face limited effective therapeutic options compared to those with isolated intrahepatic recurrence, who may benefit from curative treatments such as repeat hepatectomy or local ablation (*11*). Recent studies have identified various risk factors for HCC recurrence, including tumor-related factors (*e.g.*, tumor size, tumor number, microsatellite nodules, vascular invasion) and treatment-related factors (*e.g.*, narrow surgical margin, intraoperative blood transfusion) (*4,6,7,15-22*). However, the comprehensive landscape of risk factors specifically for distant metastasis at first recurrence remains sparsely documented. Moreover, existing prediction models for HCC recurrence often lack specificity for distant metastasis and have not been widely validated across diverse patient populations (*23-26*).

Recent advancements in understanding distant metastasis have led to potential long-term survival benefits through re-resection for oligometastases or systemic treatments for unresectable multiple metastases (*27,28*). However, standardized protocols and optimized treatment strategies for managing distant metastasis at first recurrence after HCC resection are still lacking, necessitating further in-depth exploration. Given the significant impact of distant metastasis on patient outcomes and potential for targeted interventions in high-risk individuals, there is an urgent need for accurate risk stratification tools. Early identification of patients at elevated risk for distant metastasis could inform more intensive surveillance protocols and guide application of adjuvant therapies, potentially improving long-term outcomes (*29,30*).

As such, this large-scale, multi-institutional study aimed at elucidating patterns and risk factors of distant metastasis at first recurrence following curative-intent hepatic resection for HCC. Furthermore, this study sought to develop and validate a predictive nomogram to stratify patients according to their risk of distant metastasis. This tool could potentially enable clinicians to tailor postoperative management strategies and optimize allocation of healthcare resources during postoperative follow-up for HCC patients.

2. Patients and Methods

2.1. Study design and patient population

This retrospective cohort study encompassed patients who underwent curative-intent hepatic resection for initially diagnosed HCC from January 2013 to December 2020 at 11 tertiary hospitals across China (Eastern Hepatobiliary Surgery Hospital, Fourth Hospital of Harbin, Liuyang People's Hospital, First Hospital of Jilin University, Mengchao Hepatobiliary Hospital, Pu'er People's Hospital, Shandong Provincial Qianfoshan Hospital, Fuyang People's Hospital, Ziyang First People's Hospital, First Affiliated Hospital of Harbin Medical University, and Affiliated Hospital of Nantong University). Each participating center contributed more than 100 cases (ranging from 118 to 1,053 cases). The study protocol adhered to the ethical guidelines of the 1975 Declaration of Helsinki (as revised in Brazil 2013) and was approved by the Institutional Review Boards of all participating centers. Informed consent for data use in clinical research was obtained from all patients at the time of surgery. Inclusion criteria were: *i*) age ≥ 18 years, *ii*) histologically confirmed HCC, iii) no prior anti-tumor treatments, and *iv*) curative hepatectomy (R0 resection) performed. Exclusion criteria included: *i*) palliative hepatectomy (R1 or R2 resection), *ii*) early death within 90 days after surgery or loss to follow-up at 6 months after surgery, and *iii*) missing critical prognostic data.

2.2. Data collection

Detailed baseline information on clinicopathological characteristics and operative variables was obtained from prospectively maintained databases at each institution. Patient-related factors included age, sex, etiology of liver disease, Child-Pugh grade, preoperative serum alpha-fetoprotein (AFP) level, and presence of cirrhosis or portal hypertension. Tumor and surgery-related variables consisted of largest tumor size, tumor number, vascular invasion status (microscopic or macroscopic), satellite nodules, tumor differentiation, preoperative tumor rupture, width of resection margin, intraoperative blood loss and transfusion, surgical approach (open or laparoscopic), type (anatomical or non-anatomical) and extent (minor or major) of hepatectomy. Preoperative tumor rupture was documented based on clinical presentations and imaging findings. For patients with controlled rupture and no evidence of peritoneal seeding on intraoperative exploration, surgical resection was considered feasible since rupture typically occurs on tumor surface rather than affecting intrahepatic boundaries. Cirrhosis was confirmed by postoperative histopathological findings. Portal hypertension was determined based on endoscopic evidence of esophageal varices or splenomegaly with platelet count less than 100 $\times 10^9$/L. Major hepatectomy was defined as removal of ≥ 3 Couinaud liver segments (*31*). Anatomical hepatectomy was characterized as complete anatomical removal of hepatic segments based on Couinaud's classification according to the Brisbane 2000 system (*31*).

2.3. Follow-up and study endpoints

After hepatectomy, a relatively uniform and standardized surveillance strategy for recurrence was implemented across all participating centers. This protocol involved regular monitoring of serum AFP level, abdominal ultrasonography, and contrast-enhanced computed tomography (CT) or magnetic resonance imaging (MRI) every 2-3 months for the first two years, then every 3-6 months thereafter.

The diagnosis of HCC recurrence primarily relied on the identification of new lesions showing consistent radiological manifestations with the primary tumors, with or without a continuous increase of serum AFP level. Generally, intrahepatic recurrence was confirmed by abdominal CT or MRI scan, while suspicion of extrahepatic recurrence was confirmed by performing additional examinations, including brain and chest CT, bone scintigraphy, or positron emission tomography when necessary. A variety of treatment modalities ranging from curative to palliative approaches were implemented upon the confirmation of HCC recurrence, taking into consideration the type, location, and extent of the recurrent disease. Specifically, patients with only intrahepatic recurrence may undergo curative treatment options such as repeat hepatectomy, local ablation, or liver transplantation, as well as non-curative treatments including transarterial chemoembolization, systemic therapy, or best supportive care.

The primary outcome was the occurrence of distant metastasis at first recurrence following hepatectomy, which was defined as a recurrence site outside the liver, with or without concomitant intrahepatic lesions. Secondary outcomes included overall survival (OS), calculated from initial hepatectomy to the date of death or last follow-up, and post-recurrence survival (PRS), measured from the diagnosis of first recurrence to death or last follow-up. Early recurrence was defined as occurring within 12 months post-hepatectomy, while late recurrence occurred beyond 12 months. Detailed information regarding the patterns of recurrence, treatment modality, and post-recurrence follow-up data is documented.

2.4. Statistical analysis

Continuous variables were expressed as mean ± standard deviation or median (interquartile range), and categorical variables as frequencies (n) or percentages (%). Comparisons between groups were performed using the Student's *t*-test or Mann-Whitney *U* test for continuous variables and the χ^2 or Fisher's exact test for categorical variables.

Survival outcomes were estimated using the Kaplan-Meier method and compared using the log-rank test. Univariate and multivariate Cox proportional hazards regression analyses were conducted to identify risk factors associated with distant metastasis. Variables with a p-value < 0.10 in univariate analysis were included in the multivariate model. Results were presented as hazard ratios (HR) with 95% confidence intervals (CI). A nomogram for predicting distant metastasis was constructed based on the independent risk factors identified in the multivariate analysis. The model's discriminative capability was assessed using Harrell's concordance index (C-index) and the area under the receiver operating characteristic curve (AUC). Calibration was evaluated using calibration curves. The nomogram was subjected to internal validation using bootstrap resampling (1,000 resamples). Based on the nomogram scores, patients were stratified into low-, intermediate-, and high-risk groups using cut-off values setting at the 50th and 85th percentiles. Kaplan-Meier curves were used to compare distant metastasis-free survival among the risk groups. All statistical analyses were performed using R software version 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria) and SPSS version 26.0 (IBM Corp., Armonk, NY, USA). A two-tailed p-value < 0.05 was considered statistically significant.

3. Results

3.1. Patient characteristics and recurrence patterns

Of 2,902 HCC patients who underwent curative-intent hepatectomy, 2,705 met the inclusion criteria and constituted the final analytic cohort (Supplemental Figure S1, *https://www.biosciencetrends.com/action/getSupplementalData.php?ID=246*). At a median follow-up of 62.0 months, 1,507 patients (55.7%) experienced postoperative recurrence. Among these, 342 patients (22.7% of overall recurrences, 12.6% of the entire cohort) presented with distant metastasis at first recurrence.

Comparisons of clinicopathological characteristics among patients with intrahepatic recurrence only, distant metastasis at first recurrence, and patients without recurrence are summarized in Table 1. Compared to patients without recurrence or with only intrahepatic recurrence, those with distant metastasis were significantly younger, had higher rates of preoperative tumor rupture, larger tumor size, multiple tumors, satellite nodules, poor tumor differentiation, and increased likelihood of microvascular and macrovascular invasion (all p < 0.001). Additionally, patients with distant metastasis more frequently underwent major hepatectomy, had narrow surgical margin (< 1.0 cm), experienced massive intraoperative blood loss, and received intraoperative blood transfusion.

3.2. Patterns of distant metastasis

Table 2 details the clinical characteristics of patients with

Table 1. Clinicopathologic and operative characteristics of patients undergoing hepatectomy for hepatocellular carcinoma

n (%)	Total (*n* = 2,705)	No Recurrence (*n* = 1,198)	Only Intrahepatic Recurrence (*n* = 1,165)	Distant Metastasis (*n* = 342)	p^a	p^b	p^C
Age, years*	53 ± 12	54 ± 11	53 ± 12	50 ± 11	0.001	< 0.001	0.001
> 65 years	421 (15.6)	203 (16.9)	182 (15.6)	36 (10.5)	0.077	0.006	0.015
Male sex	2,364 (87.4)	1,033 (86.2)	1,030 (88.4)	301 (88.0)	0.103	0.729	0.260
ASA score > 2	351 (13.0)	183 (15.3)	140 (12.0)	28 (8.2)	0.002	0.004	0.001
HBV (+)	2,314 (85.5)	981 (81.9)	1,040 (89.3)	293 (85.7)	< 0.001	0.962	< 0.001
HCV (+)	71 (2.6)	23 (1.9)	40 (3.4)	8 (2.3)	0.052	0.857	0.067
Cirrhosis	1,918 (70.9)	807 (67.4)	879 (75.5)	232 (67.8)	< 0.001	0.169	< 0.001
Portal hypertension	625 (23.1)	260 (21.7)	286 (24.5)	79 (23.1)	0.123	0.914	0.260
Preoperative Child-Pugh grade B	224 (8.3)	73 (6.1)	120 (10.3)	31 (9.1)	< 0.001	0.530	0.001
Preoperative BCLC stage B/C	1,184 (48.3)	312 (26.0)	599 (51.4)	273 (79.8)	< 0.001	< 0.001	< 0.001
Preoperative AFP level > 400 ng/mL	768 (32.9)	231 (26.5)	379 (33.0)	158 (50.8)	< 0.001	< 0.001	< 0.001
Preoperative tumor rupture	100 (3.7)	20 (1.7)	40 (3.4)	40 (11.7)	< 0.001	< 0.001	< 0.001
Largest tumor size, cm*	5.9 ± 3.9	4.7 ± 3.1	6.2 ± 4.0	8.7 ± 4.3	< 0.001	< 0.001	< 0.001
> 5.0 cm	1,315 (48.6)	449 (37.5)	601 (51.5)	265 (77.7)	< 0.001	< 0.001	< 0.001
Multiple tumors	520 (19.2)	120 (10.0)	280 (24.1)	120 (35.2)	< 0.001	< 0.001	< 0.001
Macrovascular invasion	255 (9.4)	31 (2.6)	121 (10.4)	103 (30.1)	< 0.001	< 0.001	< 0.001
Microvascular invasion	1,195 (44.2)	343 (28.6)	607 (52.1)	245 (71.8)	< 0.001	< 0.001	< 0.001
Satellite nodules	516 (19.1)	103 (8.6)	276 (23.7)	137 (40.2)	< 0.001	< 0.001	< 0.001
Poor tumor differentiation	2,296 (84.9)	947 (79.0)	1,039 (89.2)	310 (90.6)	< 0.001	0.002	< 0.001
Laparoscopic approach	1,051 (38.9)	432 (36.1)	555 (47.6)	64 (18.7)	0.008	< 0.001	< 0.001
Narrow resection margin (< 1.0 cm)	1,428 (52.8)	518 (43.2)	698 (59.9)	212 (62.0)	< 0.001	< 0.001	< 0.001
Major hepatectomy	626 (23.1)	175 (14.6)	298 (25.6)	153 (44.7)	< 0.001	< 0.001	< 0.001
Intraoperative blood loss > 600 mL	485 (17.9)	157 (13.1)	208 (17.9)	120 (35.2)	< 0.001	< 0.001	< 0.001
Intraoperative blood transfusion	401 (14.8)	127 (10.6)	162 (13.9)	112 (32.8)	< 0.001	< 0.001	< 0.001

AFP, alpha-fetoprotein; ASA, American Society of Anesthesiologists; HBV, hepatitis B virus; HCV, hepatitis C virus; BCLC, Barcelona Clinic Liver Cancer. *Values are mean ± standard deviation or median (interquartile range). [a]Compared between patients with recurrence (distant metastasis with/without intrahepatic recurrence) *vs.* patients without recurrence. [b]Compared between patients with distant metastasis *vs.* patients without distant metastasis (only intrahepatic recurrence or no recurrence). [C]Compared among the three groups.

distant metastasis at first recurrence. The most common sites were extrahepatic gross vessels (*n* = 124, 36.2%), lungs (*n* = 89, 26.0%), lymph nodes (*n* = 71, 20.8%), peritoneal seeding (*n* = 48, 14.0%), adrenal glands (*n* = 26, 7.6%), bones (*n* = 10, 2.9%), and brain (*n* = 4, 1.2%). Notably, 230 patients (67.3%) experienced distant metastasis within the first year after surgery, and 255 (74.6%) had multiple metastatic lesions.

3.3. Survival outcomes

Table 2. Clinical characteristics of patients with distant metastasis at first recurrence after hepatectomy for hepatocellular carcinoma

n (%)	Distant Metastasis (*n* = 342)
Male sex	301 (88.0)
Age at the diagnosis of distant metastasis, years*	51 ± 11
Child-Pugh grade at the diagnosis of distant metastasis (*n* = 302)	
A	254 (84.1)
B/C	48 (15.9)
Interval to first recurrence, months*	
Early recurrence (within 1 year after surgery)	230 (67.3)
Late recurrence (beyond 1 year after surgery)	112 (32.7)
Lesion numbers of distant metastasis	
Single metastatic lesion	87 (25.4)
Multiple metastatic lesions	255 (74.6)
Metastatic site of distant metastasis	
Extrahepatic gross vascular metastasis	124 (36.2)
Lung metastasis	89 (26.0)
Lymph node metastasis	71 (20.8)
Peritoneal seeding metastasis	48 (14.0)
Adrenal metastasis	26 (7.6)
Bone metastasis	10 (2.9)
Brain metastasis	4 (1.2)

*Values are mean ± standard deviation.

As shown in Table 3 and Figure 1, patients with distant metastasis had significantly poorer survival outcomes. The 5-year OS rates were 9.1%, 41.1%, and 90.8% for patients with distant metastasis, only intrahepatic recurrence, and no recurrence, respectively ($p < 0.001$). The median PRS for patients with distant metastasis was significantly shorter than for those with only intrahepatic recurrence (7.0 *vs.* 24.6 months, $p < 0.001$). With regard to the treatment modalities, the proportion of patients undergoing potentially curative treatment for recurrent lesions among patients with distant metastasis was significantly lower than among patients with only intrahepatic recurrence (14.0% *vs.* 47.1%, $p < 0.001$).

3.4. Risk factors of distant metastasis

Supplemental Table S1 (*https://www.biosciencetrends.com/action/getSupplementalData.php?ID=246*) and Table 4 summarize the independent risk factors associated with overall recurrence and distant metastasis at first recurrence after surgery, as identified through univariate and multivariate Cox-regression analyses. Several variables, including preoperative AFP level > 400 ng/mL, preoperative tumor rupture, largest tumor size > 5.0 cm, multiple tumors, microvascular and macrovascular invasion, satellite nodules, narrow surgical margin, and intraoperative blood transfusion, were identified as independent risk factors of distant metastasis at the first recurrence after HCC resection.

Further analysis of predictors for worse PRS was conducted among patients who experienced distant metastasis at first recurrence. As noted in Table 5, independent risk factors associated with PRS included short time interval to recurrence (within 1 year after hepatectomy), concurrent intrahepatic recurrence, and

Table 3. Comparison of recurrent patterns, treatment modalities and post-recurrence survival between patients with intrahepatic recurrence only and patients with distant metastasis at first recurrenc

n (%)	Overall Recurrence (*n* = 1,507)	Only Intrahepatic Recurrence (*n* = 1,165)	Distant Metastasis (*n* = 342)	*p*
Male sex	1,331 (88.3)	1,030 (88.4)	301 (88.0)	0.867
Age at first recurrence, years*	54 ± 12	55 ± 12	51 ± 12	< 0.001
Child-Pugh grade B/C at the diagnosis of recurrence	181/1,334 (13.6)	133/1,032 (12.9)	48/302 (15.9)	< 0.001
Interval to recurrence				
Early recurrence (within 1 year after surgery)	791 (52.5)	561 (48.2)	230 (67.3)	< 0.001
Late recurrence (beyond 1 year after surgery)	716 (47.5)	604 (51.8)	112 (32.7)	
Treatment modality for the recurrent tumor				
Potentially curative treatments	597 (39.6)	549 (47.1)	48 (14.0)	< 0.001
Non-curative treatments	910 (60.4)	616 (62.9)	294 (86.0)	
Deaths during the follow-up	1,118 (74.2)	792 (68.0)	326 (95.3)	< 0.001
Median OS (95% CI), months	37.8 (34.5, 41.0)	49.6 (45.9, 53.2)	17.6 (15.4, 19.7)	< 0.001
1-year OS rate, %	81.2	86.6	62.9	
3-year OS rate, %	51.1	59.8	21.6	
5-year OS rate, %	33.8	41.1	9.1	
Median PRS (95% CI), months	18.2 (16.6, 19.9)	24.6 (22.3, 26.9)	7.0 (6.1, 7.9)	< 0.001
1-year PRS rate, %	62.0	70.9	32.5	
3-year PRS rate, %	32.4	39.7	8.4	
5-year PRS rate, %	19.7	25.1	2.7	

*Values are mean ± standard deviation; CI, confidence interval; OS, overall survival; PRS, post-recurrence survival.

receiving non-curative treatment modalities for recurrent tumors. Furthermore, survival curves also demonstrated that patients who experienced early recurrence or had concurrent intrahepatic recurrence had worse PRS rates in comparison to those who experienced late recurrence (beyond 1 year after surgery) or did not have concurrent intrahepatic recurrence (Supplemental Figure S2-S3, *https://www.biosciencetrends.com/action/getSupplementalData.php?ID=246*).

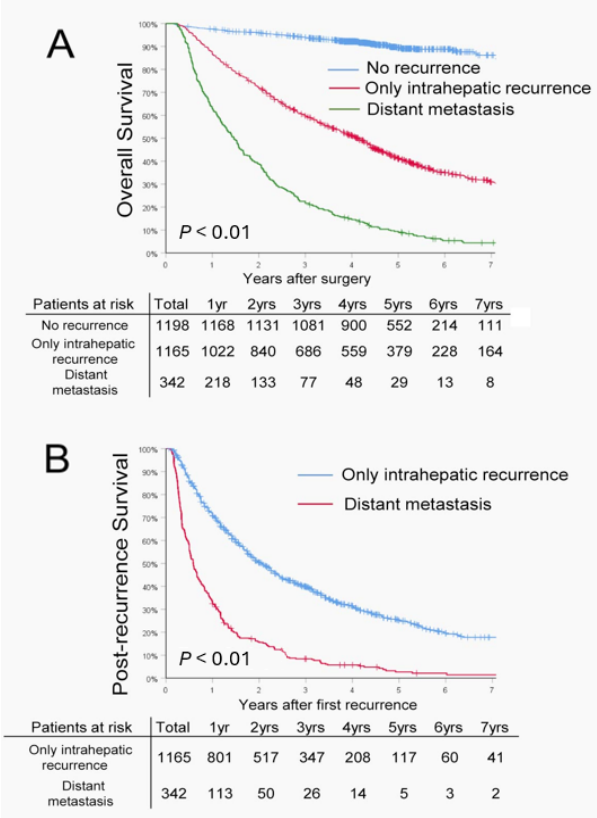


Figure 1. Kaplan-Meier survival curves comparing (A) overall survival and (B) post-recurrence survival among patients with no recurrence, only intrahepatic recurrence, and distant metastasis at first recurrence after hepatectomy for hepatocellular carcinoma.

3.5. Prediction model for distant metastasis

Based on the independent risk factors of distant metastasis, a novel nomogram for predicting 1-year and 3-year distant metastasis at first recurrence following HCC resection has been constructed (Figure 2A). Each predictive variable in the nomogram was assigned a weighted score, which was determined by its regression coefficient in the multivariable analysis (Supplemental Table S2, *https://www.biosciencetrends.com/action/getSupplementalData.php?ID=246*). These scores were then summed for each patient, representing their total scores and corresponding to the predicted probabilities of developing distant metastasis. The nomograms exhibited excellent discriminatory and calibration abilities across the entire cohort, with C-indices of 0.875, 0.865, and 0.871 for predicting distant metastasis at 1-year, 2-year, and 3-year intervals, respectively (Figure 2B). The calibration curves further demonstrated a robust alignment between the predicted probabilities and the observed occurrences of distant metastasis (Figure 2, C-D).

3.6. Risk stratification of distant metastasis

Based on the nomogram scores, patients were stratified

Table 4. Univariate and multivariate Cox-regression analyses predicting distant metastasis at first recurrence after hepatectomy for hepatocellular carcinoma

Variables	HR comparison	UV HR (95% CI)	UV *p*	MV HR (95% CI)	MV *p*
Age	> 65 *vs.* ≤ 65 years	0.622 (0.441-0.879)	0.007	NS	0.383
Sex	Male *vs.* female	0.936 (0.676-1.298)	0.692		
ASA score	> 2 *vs.* ≤ 2	0.871 (0.688-1.096)	0.164		
HBV (+)	Yes *vs.* no	1.093 (0.808-1.480)	0.564		
HCV (+)	Yes *vs.* no	0.881 (0.437-1.776)	0.723		
Cirrhosis	Yes *vs.* no	0.900 (0.717-1.130)	0.364		
Portal hypertension	Yes *vs.* no	1.017 (0.790-1.309)	0.896		
Preoperative Child-Pugh grade	B *vs.* A	1.293 (0.894-1.872)	0.172		
Preoperative AFP level > 400 ng/mL	Yes *vs.* no	2.766 (2.213-3.456)	< 0.001	1.687 (1.344-2.117)	< 0.001
Preoperative tumor rupture	Yes *vs.* no	5.133 (3.687-7.145)	< 0.001	2.558 (1.808-3.617)	< 0.001
Largest tumor size	> 5.0 *vs.* ≤ 5.0 cm	4.946 (3.830-6.389)	< 0.001	2.430 (1.845-3.202)	< 0.001
Multiple tumors	Yes *vs.* no	3.255 (2.603-4.071)	< 0.001	1.388 (1.013-1.902)	0.042
Macrovascular invasion	Yes *vs.* no	9.703 (7.640-12.323)	< 0.001	3.442 (2.639-4.489)	< 0.001
Microvascular invasion	Yes *vs.* no	4.420 (3.488-5.601)	< 0.001	2.079 (1.597-2.706)	< 0.001
Satellite nodules	Yes *vs.* no	4.615 (3.707-5.746)	< 0.001	1.716 (1.246-2.364)	0.001
Poor tumor differentiation	Yes *vs.* no	2.114 (1.469-3.044)	< 0.001	NS	0.792
Surgical approach	Open *vs.* laparoscopic	1.170 (0.882-1.486)	0.451		
Narrow resection margin	Yes *vs.* no	1.761 (1.414-2.193)	< 0.001	1.653 (1.325-2.063)	< 0.001
Major hepatectomy	Yes *vs.* no	3.816 (3.078-4.730)	< 0.001	NS	0.563
Intraoperative blood loss > 600 mL	Yes *vs.* no	3.159 (2.528-3.947)	< 0.001	NS	0.534
Intraoperative blood transfusion	Yes *vs.* no	3.670 (2.925-4.604)	< 0.001	1.486 (1.157-1.910)	0.002

AFP, alpha-fetoprotein; ASA, American Society of Anesthesiologists; CI, confidence interval; HBV, hepatitis B virus; HCV, hepatitis C virus; HR, hazard ratio; MV, multivariate; NS, not significant; UV, univariate.

Table 5. Univariate and multivariate Cox-regression analyses predicting post-recurrence survival among patients who developed distant metastasis at first recurrence after hepatectomy for hepatocellular carcinoma

Variables	HR comparison	UV HR (95% CI)	UV *p*	MV HR (95% CI)	MV *p*
Sex	Male *vs.* female	1.693 (0.839-3.414)	0.141		
Age at first recurrence	> 65 *vs.* ≤ 65 years	0.406 (0.673-2.940)	0.365		
HBV (+)	Yes *vs.* no	0.972 (0.561-1.685)	0.920		
Cirrhosis at first recurrence	Yes *vs.* no	1.004 (0.631-1.596)	0.987		
Portal hypertension at first recurrence	Yes *vs.* no	1.902 (0.753-4.805)	0.174		
Child-Pugh grade at first recurrence	B/C *vs.* A	1.677 (1.015-2.771)	0.044	NS	0.213
Interval to recurrence	<1 *vs.* ≥ 1 year	1.433 (0.935-2.196)	0.099	2.340 (1.477-3.706)	< 0.001
Largest recurrent tumor size	> 5.0 *vs.* ≤ 5.0 cm	1.671 (0.947-2.948)	0.076	NS	0.356
Multiple metastatic lesions	Yes *vs.* no	2.322 (0.918-5.872)	0.075	NS	0.546
Metastatic site	Others *vs.* lung	1.285 (0.831-1.986)	0.259		
Concurrent intrahepatic recurrence	Yes *vs.* no	1.575 (1.175-2.111)	0.002	3.169 (1.303-7.706)	0.011
Treatment modality for recurrent tumor	Curative *vs.* non-curative	0.556 (0.279-0.782)	0.001	0.423 (0.268-0.669)	< 0.001

AFP, alpha-fetoprotein; ASA, American Society of Anesthesiologists; CI, confidence interval; HBV, hepatitis B virus; HCV, hepatitis C virus; HR, hazard ratio; MV, multivariate; NS, not significant; UV, univariate.

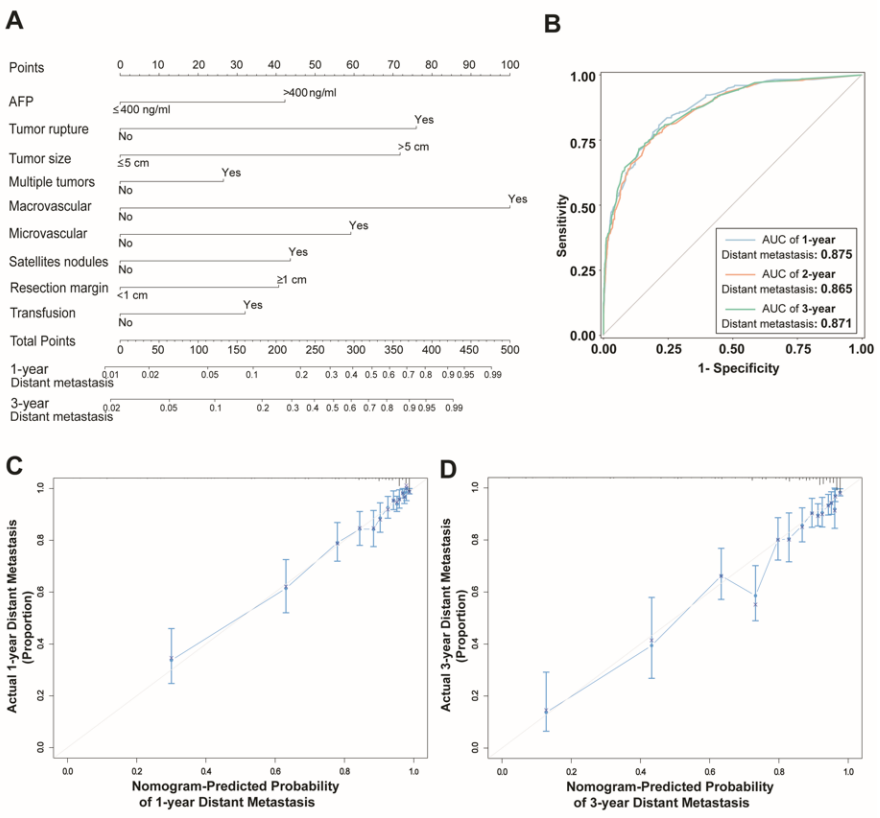


Figure 2. Nomogram for predicting distant metastasis at first recurrence after hepatectomy for hepatocellular carcinoma. **(A)** The nomogram incorporating independent risk factors. **(B)** Receiver operating characteristic curves for predicting 1-, 2-, and 3-year distant metastasis. **(C)** Calibration curve for 1-year distant metastasis prediction. **(D)** Calibration curve for 3-year distant metastasis prediction. The nomogram-predicted probability of distant metastasis is plotted on the X axis, and the actual distant metastasis is plotted on the Y axis. AUC, Area under the curve.

into low, intermediate, and high-risk groups, with cut-off values setting at the 50th and 85th percentiles. The high-risk group had a 2.98-fold higher probability of developing distant metastasis compared to the low-risk group (HR: 2.981, 95% CI: 2.639-3.268, $p < 0.001$), while the intermediate-risk group had a 1.54-fold higher probability (HR: 1.544, 95% CI: 1.253-1.845, $p < 0.001$) (Supplemental Table S3, *https://www.biosciencetrends.com/action/getSupplementalData.php?ID=246*). Kaplan-Meier survival curves for the low, intermediate, and high-risk groups stratified by the nomograms for distant metastasis are depicted in Figure 3. The cumulative rate of distant metastasis at first recurrence after hepatectomy for HCC is significantly higher in the high-risk group compared to the low-risk and moderate-risk groups ($p < 0.001$).

4. Discussion

This large-scale, multi-institutional study comprehensively

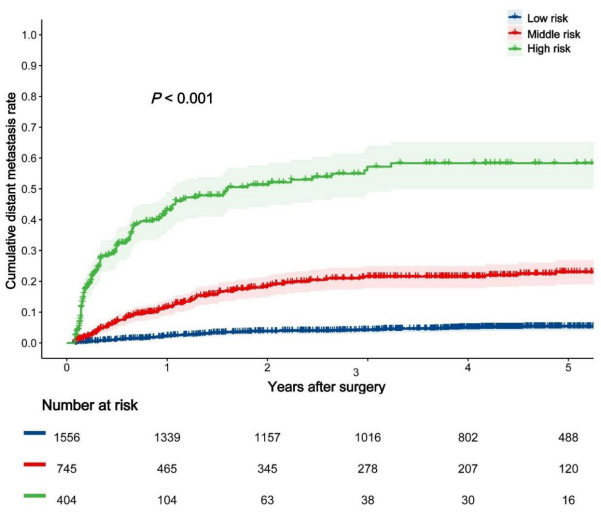


Figure 3. Kaplan-Meier curves showing cumulative rate of distant metastasis for low-, intermediate-, and high-risk groups stratified by the nomogram.

analyzed the patterns, risk factors, and outcomes of distant metastasis at first recurrence following curative hepatectomy for HCC. The findings of the present study reveal that distant metastasis occurs in a substantial proportion of patients with postoperative recurrence (22.7%) and is associated with dismal long-term outcomes. We have developed and internally validated a novel nomogram that accurately predicts the risk of distant metastasis, potentially enabling more personalized postoperative management strategies. Meanwhile, the observed patterns of distant metastasis in our cohort provide important insights into the biological behavior of recurrent HCC. The predominance of extrahepatic vascular and pulmonary metastases underscores the hematogenous spread as a key mechanism of distant dissemination. This finding is consistent with previous studies highlighting the role of circulating tumor cells in HCC metastasis (*32-35*) and suggests potential targets for future interventions aimed at preventing distant spread.

Our study identified several independent risk factors for distant metastasis, many of which reflect aggressive tumor biology. Among these factors, preoperative tumor rupture warrants particular attention. Through stringent inclusion criteria, we selected only patients whose preoperative rupture was promptly controlled and showed no evidence of peritoneal seeding upon intraoperative exploration. Since tumor rupture typically occurs on the diaphragmatic or visceral surface rather than affecting intrahepatic tumor boundaries, surgical resection with adequate margins was technically feasible following careful evaluation of tumor size and location, as confirmed by negative surgical margins on postoperative pathological examination. Notably, among the 100 patients with tumor rupture, 40% developed distant metastasis, a significantly higher rate compared to the non-rupture group, indicating that tumor rupture remains a crucial risk factor for distant dissemination even when R0 resection is achieved. Other risk factors, including vascular invasion, tumor size, and multiplicity, further emphasize the importance of early detection and timely intervention.

Our study on predicting post-hepatectomy distant metastasis demonstrates several distinctive features. First, unlike most existing models in previous studies that predict overall recurrence or survival (*23-26*), this work represents the first large-scale investigation specifically addressing distant metastasis, the recurrence pattern associated with the poorest prognosis. Then, we observed that 36.2% (124 cases) of distant metastases occurred in extrahepatic vessels, a critical pattern not well documented in current literature. More importantly, 67.3% of distant metastases developed within the first postoperative year, identifying a crucial temporal window for clinical intervention. Lastly, our prediction model achieved superior discriminative ability, with C-indices exceeding 0.85, surpassing most existing models, thus enabling more precise risk stratification and individualized surveillance protocols.

Our findings have important clinical implications, particularly in the context of regional differences in HCC management. While there are notable variations in disease etiology and treatment paradigms between Eastern and Western countries, especially regarding surgical intervention for intermediate/advanced HCC, carefully selected patients with high-risk features can achieve survival benefits through hepatectomy. The proposed nomogram effectively stratifies these patients into distinct risk groups, enabling adaptation of post-resection management strategies. Unlike intrahepatic recurrence where curative local treatments remain viable options, patients with distant metastasis primarily rely on systemic therapies. For those identified as high-risk, our prediction model supports administration of intensified surveillance protocols, including earlier and more frequent imaging examinations to detect metastatic lesions at a treatable stage. This risk-stratified approach proves particularly valuable as it guides both personalized monitoring schedules and the timing of systemic therapy initiation, with high-risk patients being potential candidates for adjuvant treatment or enrollment in clinical trials evaluating novel therapeutic strategies (*29,30*).

Our study has several limitations. First, despite the multi-institutional nature of our cohort, all participating centers were in China, potentially limiting the generalizability of our findings to other populations with different HCC etiologies. Second, our model is based solely on clinicopathological factors and does not incorporate molecular markers, which could potentially enhance its predictive accuracy. Future studies integrating genomic and proteomic data may further refine risk stratification for distant metastasis in HCC. Third, as a real-world retrospective study, standardization

of adjuvant therapy was challenging due to multiple factors, including evolving evidence for adjuvant treatment efficacy and varying institutional protocols. Current clinical trials have yielded conflicting results regarding the effectiveness of postoperative adjuvant therapy in reducing HCC recurrence, underscoring the need for large-scale prospective studies to establish optimal adjuvant treatment strategies, particularly for patients identified as high-risk for distant metastasis.

In conclusion, this study provides a comprehensive analysis of distant metastasis patterns following HCC resection and presents a novel, internally validated nomogram for predicting this adverse outcome. The ability to accurately stratify patients according to their risk of distant metastasis may inform personalized postoperative surveillance strategies and guide early intervention in high-risk individuals. Future prospective studies are warranted to evaluate the clinical impact of risk-adapted management based on our prediction model and to explore targeted approaches for preventing distant metastasis in HCC.

Acknowledgements

We would like to express our gratitude to Prof. Chen Zhong and his team from the Department of Hepatobiliary Surgery, Affiliated Hospital of Nantong University, for their participation in this multicenter retrospective study and their valuable contribution of HCC patient data.

Funding: This study was supported by the National Natural Science Foundation of China (No. 82273074, 82425049 for Yang T; 82372813 for Wang MD), Shanghai Health and Hygiene Discipline Leader Project (No. 2022XD001 for Yang T), Shanghai Outstanding Academic Leader Program (No. 23XD1424900 for Yang T), Natural Science Foundation of Shanghai (No. 22ZR1477900 for Wang MD), and Shanghai Science and Technology Committee Rising-Star Program (No. 22QA1411600 for Wang MD; 24YF2758600 for Diao YK).

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Vogel A, Meyer T, Sapisochin G, Salem R, Saborowski A. Hepatocellular carcinoma. Lancet. 2022; 400:1345-1362.
2. Wang MD, Diao YK, Yao LQ, Fan ZQ, Wang KC, Wu H, Gu LH, Xu JH, Li C, Lv GY, Yang T. Emerging role of molecular diagnosis and personalized therapy for hepatocellular carcinoma. *iLIVER.* 2024; 3:100083.
3. Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR. A global view of hepatocellular carcinoma: Trends, risk, prevention and management. Nat Rev Gastroenterol Hepatol. 2019; 16:589-604.
4. Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM, Roberts LR, Heimbach JK. Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the Study of Liver Diseases. Hepatology. 2018; 68:723-750.
5. Yao LQ, Chen ZL, Feng ZH, *et al.* Clinical features of recurrence after hepatic resection for early-stage hepatocellular carcinoma and long-term survival outcomes of patients with recurrence: A multi-institutional analysis. Ann Surg Oncol. 2022; 29:4291-4303.
6. Yan WT, Li C, Yao LQ, *et al.* Predictors and long-term prognosis of early and late recurrence for patients undergoing hepatic resection of hepatocellular carcinoma: A large-scale multicenter study. Hepatobiliary Surg Nutr. 2023; 12:155-168.
7. Wang MD, Li C, Liang L, Xing H, Sun LY, Quan B, Wu H, Xu XF, Wu MC, Pawlik TM, Lau WY, Shen F, Yang T. Early and late recurrence of hepatitis B virus-associated hepatocellular carcinoma. Oncologist. 2020; 25:e1541-e1551.
8. Chen WY, Li C, Liu ZP, *et al.* Novel online calculator to predict reduced risk of early recurrence from adjuvant transarterial chemoembolisation for patients with hepatocellular carcinoma. eGastroenterology. 2023; 1:e100008.
9. Xu X, Wang MD, Xu JH, *et al.* Adjuvant immunotherapy improves recurrence-free and overall survival following surgical resection for intermediate/advanced hepatocellular carcinoma a multicenter propensity matching analysis. Front Immunol. 2023; 14:1322233.
10. Gou XX, Shi HY, Li C, Chen ZL, Ouyang W, Sun LY, Diao YK, Wang MD, Yao LQ, Gu LH, Pawlik TM, Lau WY, Shen F, Xue J, Yang T. Association of adjuvant radiation therapy with long-term overall and recurrence-free survival after hepatectomy for hepatocellular carcinoma: A multicenter propensity-matched study. Int J Radiat Oncol Biol Phys. 2022; 114:238-249.
11. Tabrizian P, Jibara G, Shrager B, Schwartz M, Roayaie S. Recurrence of hepatocellular cancer after resection: patterns, treatments, and prognosis. Ann Surg. 2015; 261:947-955.
12. Tsilimigras DI, Bagante F, Moris D, *et al.* Recurrence patterns and outcomes after resection of hepatocellular carcinoma within and beyond the Barcelona Clinic Liver Cancer Criteria. Ann Surg Oncol. 2020; 27:2321-2331.
13. Ivanics T, Murillo Perez CF, Claasen M, Claasen MPAW, Patel MS, Morgenshtern G, Erdman L, Shwaartz C, Rajendran L, O'Kane GM, Hansen BE, Cleary SP, Sapisochin G. Dynamic risk profiling of HCC recurrence after curative intent liver resection. Hepatology. 2022; 76:1291-1301.
14. Papaconstantinou D, Tsilimigras DI, Pawlik TM. Recurrent hepatocellular carcinoma: Patterns, detection, staging and treatment. J Hepatocell Carcinoma. 2022; 9:947-957.
15. Yoon JH, Choi SK, Cho SB, Kim HJ, Ko YS, Jun CH. Early extrahepatic recurrence as a pivotal factor for survival after hepatocellular carcinoma resection: A 15-year observational study. World J Gastroenterol. 2022; 28:5351-5363.
16. Hong SK, Jin XL, Suh S, Hong SY, Hong K, Han ES, Lee JM, Choi Y, Yi NJ, Lee KW, Suh KS. Different risk factors for early and late recurrence after curative resection of hepatocellular carcinoma. World J Surg. 2022; 46:197-206.
17. Xu XF, Xing H, Han J, Li ZL, Lau WY, Zhou YH, Gu WM, Wang H, Chen TH, Zeng YY, Li C, Wu MC,

Shen F, Yang T. Risk factors, patterns, and outcomes of late recurrence after liver resection for hepatocellular carcinoma: A multicenter study from China. JAMA Surg. 2019; 154:209-217.
18. Xing H, Zhang WG, Cescon M, *et al.* Defining and predicting early recurrence after liver resection of hepatocellular carcinoma: A multi-institutional study. HPB (Oxford). 2020; 22:677-689.
19. Tang SC, Diao YK, Lin KY, *et al.* Association of Pringle maneuver with postoperative recurrence and survival following hepatectomy for hepatocellular carcinoma: A multicenter propensity score and competing-risks regression analysis. Hepatobiliary Surg Nutr. 2024; 13:412-424.
20. Feng ZH, Wang MD, Chen Z, *et al.* Risk factors and long-term prognosis of beyond-Milan recurrence after hepatectomy for BCLC stage 0/A hepatocellular carcinoma: A large-scale multicenter study. Surgery. 2022; 172:1147-1155.
21. Zhang H, Han J, Xing H, Li ZL, Schwartz ME, Zhou YH, Chen TH, Wang H, Gu WM, Lau WY, Wu H, Wu MC, Shen F, Yang T. Sex difference in recurrence and survival after liver resection for hepatocellular carcinoma: A multicenter study. Surgery. 2019; 165:516-524.
22. Wang MD, Sun LY, Qian GJ, Li C, Gu LH, Yao LQ, Diao YK, Pawlik TM, Lau WY, Huang DS, Shen F, Yang T. Prothrombin induced by vitamin K Absence-II versus alpha-fetoprotein in detection of both resectable hepatocellular carcinoma and early recurrence after curative liver resection: A retrospective cohort study. Int J Surg. 2022; 105:106843.
23. Liang L, Quan B, Wu H, *et al.* Development and validation of an individualized prediction calculator of postoperative mortality within 6 months after surgical resection for hepatocellular carcinoma: an international multicenter study. Hepatol Int. 2021; 15:459-471.
24. Ni X, Li D, Dai S, Pan H, Sun H, Ao J, Chen L, Kong H. Development and evaluation of nomograms to predict the cancer-specific mortality and overall mortality of patients with hepatocellular carcinoma. Biomed Res Int. 2021; 2021:1658403.
25. Lu Y, Ren S, Jiang J. Development and validation of a nomogram for survival prediction in hepatocellular carcinoma after partial hepatectomy. BMC Surg. 2023; 23:27.
26. Diao YK, Sun L, Wang MD, *et al.* Development and validation of nomograms to predict survival and recurrence after hepatectomy for intermediate/advanced (BCLC stage B/C) hepatocellular carcinoma. Surgery. 2024; 176:137-147.
27. Yoh T, Seo S, Taura K, Iguchi K, Ogiso S, Fukumitsu K, Ishii T, Kaido T, Uemoto S. Surgery for recurrent hepatocellular carcinoma: Achieving long-term survival. *Ann Surg.* 2021; 273:792-799.
28. Midorikawa Y, Takayama T, Nakayama H, Moriguchi M, Aramaki O, Yamazaki S, Teramoto K, Yoshida N, Kobayashi N, Tsuji S, Higaki T. Favorable outcomes of surgical resection for extrahepatic recurrent hepatocellular carcinoma. Hepatol Res. 2020; 50:978-984.
29. Mo A, Lin B, Chen D. Efficacy of sequential TACE on primary hepatocellular carcinoma with microvascular invasion after radical resection: a systematic review and meta-analysis. World J Surg Oncol. 2023; 21:277.
30. Feng X, Feng GY, Tao J, Ao YP, Wu XH, Qi SG, Shi ZR. Comparison of different adjuvant therapy regimen efficacies in patients with high risk of recurrence after radical resection of hepatocellular carcinoma. J Cancer Res Clin Oncol. 2023; 149:10505-10518.
31. Strasberg SM, Phillips C. Use and dissemination of the brisbane 2000 nomenclature of liver anatomy and resections. Ann Surg. 2013; 257:377-382.
32. Lu M, Zhu WW, Wang X, *et al.* ACOT12-dependent alteration of acetyl-CoA drives hepatocellular carcinoma metastasis by epigenetic induction of epithelial-mesenchymal transition. Cell Metab. 2019; 29:886-900. e5.
33. Sun YF, Xu Y, Yang XR, Guo W, Zhang X, Qiu SJ, Shi RY, Hu B, Zhou J, Fan J. Circulating stem cell-like epithelial cell adhesion molecule-positive tumor cells indicate poor prognosis of hepatocellular carcinoma after curative resection. Hepatology. 2013; 57:1458-1468.
34. Hu ZQ, Zhou SL, Li J, Zhou ZJ, Wang PC, Xin HY, Mao L, Luo CB, Yu SY, Huang XW, Cao Y, Fan J, Zhou J. Circular RNA sequencing identifies CircASAP1 as a key regulator in hepatocellular carcinoma metastasis. Hepatology. 2020; 72:906-922.
35. Guo W, Sun YF, Shen MN, *et al.* Circulating tumor cells with stem-like phenotypes for diagnosis, prognosis, and therapeutic response evaluation in hepatocellular carcinoma. Clin Cancer Res. 2018; 24:2203-2213.

Received December 6, 2024; Revised January 26, 2025; Accepted February 21, 2025.

[§]These authors contributed equally to this work.
**Address correspondence to:*
Ming-Da Wang and Tian Yang, Department of Hepatobiliary Surgery, Eastern Hepatobiliary Surgery Hospital, Second Military Medical University (Naval Medical University), No. 225, Changhai Road, Shanghai 200438, China.
E-mail: wangmingda1987@163.com (MW); or yangtianehbh@smmu.edu.cn (TY)

Released online in J-STAGE as advance publication March 4, 2025.

Original Article

DOI: 10.5582/bst.2024.01424

The profile and clinical predicting effect of non-rash dermatologic toxicity related to targeted therapy in stage-IV non-small cell lung cancer patients

Yanmei Peng[1,2], Collin M. Costello[3], Zhaoheng Liu[2,4], Ashok V. Kumar[2], Zhong Gu[2], Nikhila Kosuru[2], Jason A. Wampfler[5], Pedro A. Reck dos Santos[6], Jonathan D'Cunha[6], Vinicius Ernani[7], Ping Yang[2,*]

[1] Department of Oncology, Fangshan Hospital Beijing University of Chinese Medicine, Beijing, China;
[2] Division of Epidemiology, Department of Quantitative Health Sciences, Mayo Clinic, AZ, USA;
[3] Department of Dermatology, Mayo Clinic, AZ, USA;
[4] School of Life Sciences, Beijing University of Chinese Medicine, Beijing, China;
[5] Division of Clinical Trials and Biostatistics, Department of Quantitative Health Sciences, Mayo Clinic, MN, USA;
[6] Division of Thoracic Surgery, Mayo Clinic, AZ, United States of America;
[7] Division of Hematology and Medical Oncology, Department of Medicine, Mayo Clinic, AZ, USA.

SUMMARY: Dermatologic toxicities associated with targeted therapies may impact drug intolerance and predict drug response, among which rash is most frequently reported and well delineated. However, the profile and effect of non-rash dermatologic toxicity are not fully understood. We identified stage-IV non-small cell lung cancer patients diagnosed at Mayo Clinic in 2006-2019 and systematically analyzed demographics, targeted agents, toxicity, response, and survival outcomes of patients who received targeted therapy. Five toxicity subgroups-none, only non-rash dermatologic, concurrent non-rash and rash (concurrent) dermatologic, only rash, and others-were compared; multivariable survival analyses employed Cox Proportional Hazard models. This study included 533 patients who had taken targeted therapies: 36 (6.8%) had no toxicity, 26 (4.9%) only non-rash dermatologic, 193 (36.2%) only rash, 134 (25.1%) concurrent dermatologic, 144 (27.0%) other toxicities. Non-rash dermatologic toxicities predominately included xerosis (12.8%), pruritus (8.5%), paronychia (7.0%). Rash was the most frequent (59.4%) and the earliest occurring (21 median onset days [MOD]) dermatologic toxicity; paronychia was the latest (69 MOD) occurring. In 329 epidermal growth factor receptor inhibitors-treated patients with dermatologic toxicity, mild toxicity occurred the most frequently in patients with only non-rash (81.8%), then those with only rash (64.8%), and the least in the concurrent (50.4%, P=0.013). Patients with concurrent dermatologic toxicities had a significantly higher response rate (67.9%) than those with only non-rash (53.8%) or only rash (41.1%, $p < 0.001$). Multivariable analysis demonstrated concurrent dermatologic toxicity independently predicted a lower risk of death (harzard ratio [HR] 0.48 [0.30-0.77], $p < 0.001$). Compared to rash, non-rash dermatologic toxicity might be a stronger predictor of better treatment response and longer survival in patients who received targeted therapy.

Keywords: lung cancer, target therapy, dermatologic toxicity, non-rash, survival

1. Introduction

Lung cancer accounts for approximately 21% of cancer deaths in the United States; however, the mortality has been declining with advances in targeted therapy, partly driven by improved overall survival time of stage-IV non-small cell lung cancer (NSCLC) (*1*). Meanwhile, certain adverse effects associated with these novel anti-cancer agents, particularly dermatologic toxicity, have been significant and draw attention of care providers (*2-4*).

NSCLC accounts for 85% of lung cancer (*5*), predominantly consisting of adenocarcinoma and squamous cell carcinoma. Approximately 78% of Asian populations and 60% of Western populations of lung adenocarcinoma patients have driver gene mutations, including *epidermal growth factor receptor* (*EGFR*), *anaplastic lymphoma kinase* (*ALK*), *c-ros oncogene 1* (*ROS1*), *Kirsten rat sarcoma virus* (*KRAS*), *V-raf murine sarcoma oncogene homolog B1* (*BRAF*), *MET*,

and *human epidermal growth factor receptor* (*HER2*), rearranged during transfection *(RET)* and other genetic alterations (*6*). Other targeted agents included anti-*vascular endothelial growth factor* (*VEGF*) therapy, widely used for targeting tumor angiogenesis (*7*), and *mammalian target of rapamycin* (mTOR) inhibitors, which targets a cellular pathway driving oncogenesis and tumor progression (*8*) independent of specific gene mutations.

Dermatologic toxicity was mostly reported in *EGFR* inhibitors compared to other targeted drugs, typically presenting as papulopustular (acneiform) rash, xerosis, pruritus, paronychia, hair changes, and mucositis, and their incidences ranged from 47% to 100%, 10% to 49%, 8% to 57%, 3% to 25%, 0 to 13%, and 0 to 44% (*9*), respectively. Even though most dermatologic toxicities are not life-threatening, their symptoms are unfavorably correlated with quality of life (*10*). One of the earlier clinical studies to explore the relationship between rash and clinical outcomes showed patients who developed cutaneous rash were associated with better response and prolonged survival in 57 NSCLC patients treated with erlotinib, a classic *EGFR* inhibitor (*11*). A similar result between rash and survival has been observed in a real-world cohort of 79 patients with erlotinib (*12*). Higher severity of rash was also found to be a potential marker for the long-term efficacy of afatinib in 32 NSCLC patients (*13*). We also validated that dermatologic toxicity was a protective predictor for treatment response and survival (*14*). However, the specific relationship between non-rash dermatological and drug response, as well as survival length, is not documented, especially from real-world settings.

Our current study was designed to provide additional perspectives to fulfill the knowledge gap on the profile and predictive value of non-rash dermatologic toxicities in stage-IV lung cancer patients based on a 14-year prospectively enrolled and followed clinical cohort.

2. Patients and Methods

2.1. Study population and grouping

A total of 3,767 patients with newly diagnosed stage-IV NSCLC were identified from January 1, 2006 to December 31, 2016 (*15-17*) in Mayo Clinic Lung Cancer Cohort and consecutive case series from January 1, 2017 to December 31, 2019 (*18*). Patients were staged at the time of original diagnosis according to the 5th (*19*) or 7th (*20*) edition of TNM staging system. Inclusion criteria were *i*) patients were newly diagnosed stage-IV NSCLC from January 1, 2006 to December 31, 2019, *ii*) patients were treated with targeted therapy at Mayo Clinic, and (iii) patients signed content form. Exclusion criteria were *i*) patients had no documented toxicity information relevant to targeted therapy, *ii*) patients were lost to follow-up or terminated targeted therapy within one month from treatment initiation, and *iii*) patients were treated with concurrent chemoradiation and targeted therapy. Targeted agents targeted specific driver genes (*e.g.*, *EGFR*, *ALK*/*ROS1*) and other antagonists targeting mTOR and *VEGF/VEGFR*.

The included patients were divided into five toxicity subgroups-none, only non-rash dermatologic, concurrent non-rash and rash (concurrent) dermatologic, only rash, and others based on the targeted therapy status. Patients without any targeted therapy-induced toxicity were grouped into none group; those who had dermatologic toxicity but not rash were put into only non-rash dermatologic group, and they could also have non-dermatologic toxicity or not; those who had dermatologic toxicity with concurrent rash and other dermatologic toxicities that included but not limited to xerosis, pruritus, paronychia, erythema, mucositis, and nail changes were put into concurrent dermatologic group; those with only rash rather than other dermatologic toxicities were put into only rash group, and they could also have non-dermatologic toxicity or not; and those with only non-dermatologic toxicity were put into others group.

2.2. Data collection

The electronic medical records were reviewed, including detailed information on demographics, smoking history, lung cancer diagnosis, treatment, targeted therapy-associated toxicity, treatment response, and vital status under the approval of the Mayo Foundation Institutional Review Board approval (IRB# 225-99).

Dermatologic toxicities were identified and categorized into rash and non-rash toxicities. Rash referred to acneiform, maculopapular, erythematous papular/pustular; non-rash dermatologic toxicity included xerosis, pruritus, paronychia, erythema, mucositis, nail changes, and other dermatological reactions that have been reported previously (*21*). The severity of dermatologic toxicities was evaluated by Common Terminology Criteria for Adverse Events (CTCAE) v5.0 (*22*), were graded into mild (grade 1), moderate (grade 2), severe (grade 3), life-threatening (grade 4) and death (grade 5). If the grade of toxicity based on CTCAE was not documented, descriptions related to severity were employed. Definitions of severity in medical records were identified by oncologists as follows: "tolerable or tolerated, sporadic, some, notable, occasional, manageable, faint (skin disorders)" were graded as mild; "intermittent, some continued, some modest, worsen" were categorized as moderate; "extremely, profound, generalized, faint (anemia, weak, fatigue), generalized, outstanding, persistent, quite a bit, really bad, significant, prominent, considerable, substantial, very" were considered as severe. The toxicity onset time was defined as the time from targeted drug initiation to toxicity occurrence.

Treatment response was determined by the best response to targeted drugs, evaluated by Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 (*23*), categorized by complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD). Objective response rate was the percentage of CR and PR (abbreviated as "response"). If original imaging tests and biopsy of the new suspected disease were not available, descriptions of response noted in medical records were utilized to define the best response. Definition of the descriptions for response were identified by oncologists based on the following criteria: "Totally resolved, complete remission, complete response, free of disease, negative bronchial margins, negative for tumor, no evidence of disease" were classified as CR; "Interval response, near complete response, good response, nice response, a remarkable response, nice regression, responded well, reduction, improvement of disease, and dramatic shrinkage" were categorized as PR; "Stable, stable disease, and good control" were considered as SD; "Disease progression, recurrent, progressive, recurrence, and new metastatic" were identified as PD.

2.3. Statistical methods

Age was analyzed by Kruskal-Wallis test; sex, race, smoking status, cell type, treatment modality, treatment line, treatment response, toxicity severity and gene status were evaluated by Chi-Square test to identify differences in five toxicity groups: none, only non-rash dermatologic, concurrent non-rash and rash (concurrent) dermatologic, only rash, and other (non-dermatologic toxicities), as well as the subgroup analysis in three dermatologic toxicity groups (only non-rash, concurrent, only rash). A Logrank test assessed overall survival (OS), defined as the date of targeted drug initiation to the date of last follow-up or patient death with the endpoint on April 30, 2022. A Cox proportional hazards model was developed for multivariable analysis to evaluate the toxicity status and known prognostic factors, including age, sex, race, smoking status, cell type, treatment modality, treatment line and treatment response. Hazard ratios (HR) with 95% confidence intervals (CI) were calculated. Statistical analyses were performed using SAS, v.9.4 (SAS Institute Inc., Cary, NC, USA). Means (standard deviation, SD) and medians were reported for continuous data, and counts (*n*) and frequency (%) were used for categorical data. A two-sided $p < 0.05$ was statistically significant.

3. Results

In 3,767 stage-IV NSCLC patients, 1,856 (49.3%) received systemic therapy. Among the remaining 1,911 (50.7%) patients who did not receive systemic therapy, 185 had surgery and 625 had radiation. Excluding 1,224 patients that were treated only with chemotherapy, immunotherapy, or concurrent chemoradiation. Six hundred and thirty-two (34.1%) of the 1856 patients were treated with targeted therapy. After excluding 99 patients without information on toxicity, 533 were included in the analyses (Figure 1): dermatologic toxicity occurred in 353/533 (66.2%), including 26/533 (4.9%) only non-rash, 193 (36.2%) only rash, 134 (25.1%) concurrent, other toxicity in 144 (27.0%), and none in 36

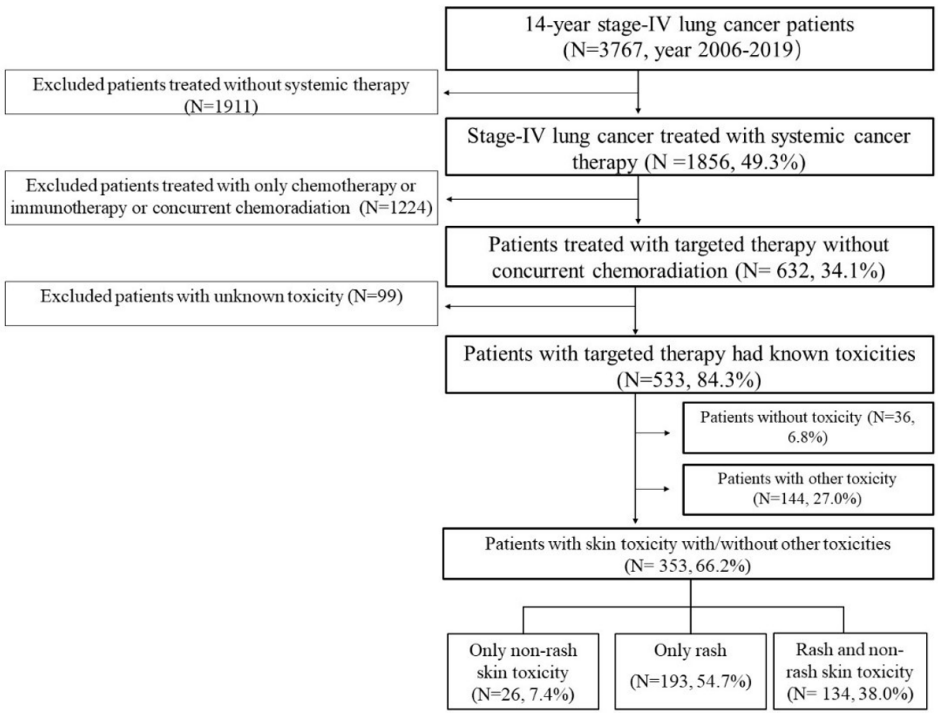


Figure 1. The flow chart for the study population with patient inclusion and exclusion criteria.

(6.8%). Of the 533 patients, the mean age (± SD) at lung cancer diagnosis was 62.4 (± 12.74) years, with 58.9% being female, 87.4% Whites, and 90.4% adenocarcinoma (Table 1). We observed that dermatologic toxicity was more frequent in *EGFR*+ than *EGFR*- patients (73.4% *vs.* 45.0%; $p < 0.001$). Among 353 patients with dermatologic reactions, 26/353 (7.4%) had only non-rash, 193 (54.7%) only rash, 134 (38.0%) concurrent (Supplemental Table S1, *https://www.biosciencetrends.com/action/getSupplementalData.php?ID=245*); concurrent dermatologic toxicity was more frequent in patients with EGFR+ than EGFR- tumors (44.0% *vs.* 22.0%, $p = 0.014$), and in those who received targeted therapy as the first-line treatment than other lines (45% *vs.* 30.8%, $p = 0.005$). A total of 727 dermatological events were observed among 353 patients; the frequencies (*n*, %) from the highest to the lowest were rash (432, 59.4%), xerosis (93, 12.8%), pruritus (62, 8.5%), paronychia (51, 7.0%), erythema (36, 5.0%), mucositis (12, 1.7%), nail changes (12, 1.7%) and others (30/727, 4.1%), predominately included dermatitis, skin pigmentation, and eyelash changes (Table 2A). When comparing the distribution of non-rash dermatologic toxicities, more patients had xerosis in the only non-rash dermatologic group (14/26, 53.8%) than those in the concurrent dermatologic group (68/134, 50.7%); conversely, more patients had pruritus (58/134, 43.3%), paronychia (31, 23.1%) and mucositis (7, 5.2%) in the concurrent dermatologic group compared with those in only the non-rash dermatologic group (both pruritus and paronychia 1/26, 3.8%; mucositis 0; $p = 0.003$).

In 353 patients with dermatologic toxicity, 348 (98.6%) had known severity of toxicity: 206/353 (59.2%) patients experienced grade 1 toxicity, 49 (14.1%) grade 2, 93 (26.7%) grade 3-4. Grade 1 toxicity was found in the most patients with only non-rash (76.0%), then those with only rash (62.4%), and the least in those with concurrent (51.5%) dermatologic toxicity, though significance did not reach the p-value threshold ($p = 0.069$). The incidence, severity, and onset days for the common dermatologic toxicities varied by drugs were identified (Table 3). Rash was found the earliest occurring (21 median onset days [MOD]) while paronychia the latest (69 MOD). A similar distribution of incidence, severity and onset time were observed in patients with *EGFR* inhibitors. The responsible drugs associated with the common dermatologic toxicities were scrutinized (Figure 2). Erlotinib (64.5%), then osimertinib (11.3%) and afatinib (11.1%) were preponderantly drugs associated with these dermatologic events but occurrence rates varied, Figure 2A. Rash occurred more frequently in erlotinib (66.5%) than in afatinib (58.9%) and osimertinib (44.8%); and paronychia was more associated with osimertinib (14.9%) and afatinib (10.7%) than erlotinib (4.0%), Figure 2B. Other non-skin toxicities mainly reported fatigue (14.0%), diarrhea (14.9%), nausea (10.8%), anorexia (5.7%), vomiting (4.1%) and anemia (3.4%) (Table 2B).

When looking into the 442 *EGFR* inhibitors-treated patients, 24 (5.4%) patients had none, 23 (5.2%) only non-rash dermatologic, 127 (28.7%) concurrent dermatologic, 179 (40.5%) only rash, and 89 (20.1%) other toxicities. Among them, more patients (315/442, 71.3%) had *EGFR*+ tumors. Dermatologic toxicity subgroup analysis showed patients with only non-rash (81.8%) were the most frequently observed grade 1 toxicity, then those with rash (64.8%), and the least in those with concurrent (50.4%, $p = 0.013$). In 63 patients treated with *ALK/ROS1* inhibitors, 6 (9.5%) patients had no toxicity, 2 (3.2%) only non-rash dermatologic toxicity, 3 (4.8%) concurrent dermatologic toxicity, 5 (7.9%) only rash and 47 (74.6%) other toxicities. *ALK* and *ROS1* mutations were identified in 41/63 (65.1%) and 4/63 (6.3%), respectively. Comparison of variables was limited by sample size.

Patients with dermatologic toxicity (52.3%) had similar ORR compared with those without (43.8%, $p = 0.127$). However, when focusing on dermatologic toxicity subgroups, we found patients with concurrent dermatologic (67.9%) had a significantly higher ORR than those with only non-rash dermatologic (53.8%) and only rash (41.1%, $p < 0.001$) toxicities. Similar differences were also identified in those with *EGFR* inhibitors ($p < 0.001$), indicating non-rash dermatologic toxicity was more likely to enhance the drug-efficacy predictor (Supplemental Table S2, *https://www.biosciencetrends.com/action/getSupplementalData.php?ID=245*). For all patients with targeted therapy, multivariable analysis showed patients in concurrent dermatologic toxicity group had longer median survival years (2.6 years) than those in other groups (1.5-1.9 years) and lower risk of death (HR 0.71, 95% CI [0.46-1.10], $p = 0.009$) adjusting for smoking status, cell type, treatment modality, treatment response and age (Supplemental Table S3, *https://www.biosciencetrends.com/action/getSupplementalData.php?ID=245*). Furthermore, when focusing on dermatologic toxicity subgroups, the concurrent group was an independent predictor of a lower risk of death (HR 0.48, 95% CI [0.30-0.77]) ($p < 0.001$, Figure 3A) adjusting for cell type, treatment modality, treatment response, and age (Table 4). Severity of drug-related dermatologic toxicities showed no correlation with drug efficacy. A similar association with dermatologic toxicity and survival benefits was found in anti-*EGFR* agents (Figure 3B). However, the limited amount of patients with anti-*ALK*/*ROS1* agents could not be used for Cox model analysis.

These results highlighted the importance of non-rash dermatologic toxicities in the drug-efficacy predictive value of dermatologic toxicities in targeted therapy-treated and *EGFR* inhibitors-treated patients. Concurrent dermatologic toxicity predicted a strengthened efficacy and longer survival.

Table 1. The characteristics of 533 patients with targeted therapy in five toxicity status groups

	Toxicity					Total (n = 533)	p value
	None (n = 36)	Only Non-rash (n = 26)	Non-rash and rash (n = 134)	Only rash (n = 193)	Other (n = 144)		
Age at diagnosis							0.955
Mean (SD)	62.8 (13.53)	63.0 (14.55)	62.9 (12.71)	62.3 (13.04)	61.9 (11.97)	62.4 (12.74)	
Median	65	65	64	63	64	64	
Range	29.0, 94.0	36.0, 84.0	31.0, 87.0	26.0, 86.0	34.0, 94.0	26.0, 94.0	
Sex, n (%)							0.709
0 = Female	18 (50.0)	14 (53.8)	81 (60.4)	112 (58.0)	89 (61.8)	314 (58.9)	
1 = Male	18 (50.0)	12 (46.2)	53 (39.6)	81 (42.0)	55 (38.2)	219 (41.1)	
Race, n (%)							0.923
White	31 (88.6)	24 (92.3)	116 (86.6)	166 (86.5)	126 (88.1)	463 (87.4)	
Other	4 (11.4)	2 (7.7)	18 (13.4)	26 (13.5)	17 (11.9)	67 (12.6)	
NA	1	0	0	1	1	3	
Smoking history, n (%)							0.451
Never	19 (52.8)	18 (69.2)	85 (63.4)	121 (62.7)	81 (56.3)	324 (60.8)	
Ever	17 (47.2)	8 (30.8)	49 (36.6)	72 (37.3)	63 (43.8)	209 (39.2)	
Cell type, n (%)							0.876
Adenocarcinoma	34 (94.4)	24 (92.3)	121 (90.3)	172 (89.1)	131 (91.0)	482 (90.4)	
Other	2 (5.6)	2 (7.7)	13 (9.7)	21 (10.9)	13 (9.0)	51 (9.6)	
Treatment modality, n (%)							0.846
Drug therapy	28 (77.8)	20 (76.9)	100 (74.6)	132 (68.4)	113 (78.5)	393 (73.7)	
Surgery & drug	1 (2.8)	1 (3.8)	8 (6.0)	12 (6.2)	6 (4.2)	28 (5.3)	
Surgery, rad, drug	1 (2.8)	0	4 (3.0)	4 (2.1)	3 (2.1)	12 (2.3)	
Radiation & drug	6 (16.7)	5 (19.2)	22 (16.4)	45 (23.3)	22 (15.3)	100 (18.8)	
Treatment line of targeted therapy, n (%)							0.004
First line	19 (52.8)	16 (66.7)	86 (64.2)	89 (47.1)	62 (44.0)	272 (51.9)	
Other	17 (47.2)	8 (33.3)	48 (35.8)	100 (52.9)	79 (56.0)	252 (48.1)	
NA	0	2	0	3	4	9	
EGFR mutation, n (%)							< 0.0001
Negative	12 (42.9)	5 (26.3)	11 (9.6)	34 (22.5)	49 (41.5)	111 (25.8)	
Positive	16 (57.1)	14 (73.7)	103 (90.4)	117 (77.5)	69 (58.5)	319 (74.2)	
ALK mutation, n (%)							
Negative	4 (40.0)	2 (100.0)	2 (33.3)	8 (72.7)	11 (28.2)	27 (39.7)	0.036
Positive	6 (60.0)	0	4 (66.7)	3 (27.3)	28 (71.8)	41 (60.3)	
ROS1 mutation, n (%)							
Negative	1 (100.0)	1 (100.0)	2 (100.0)	3 (60.0)	4 (57.1)	11 (68.8)	-
Positive	0	0	0	2 (40.0)	3 (42.9)	5 (31.3)	
Other genetic mutation, n (%)							
Negative	0	0	1 (20.0)	4 (57.1)	6 (50.0)	11 (42.3)	-
Positive	1 (100.0)	1 (100.0)	4 (80.0)	3 (42.9)	6 (50.0)	15 (57.7)	

Abbreviations: SD, standard deviation; *EGFR,* epidermal growth factor receptor; *ALK,* anaplastic lymphoma kinase; *ROS1,* c-ros oncogene 1.

Table 2 (A). The distribution of common dermatologic toxicities in patients with different type of drugs

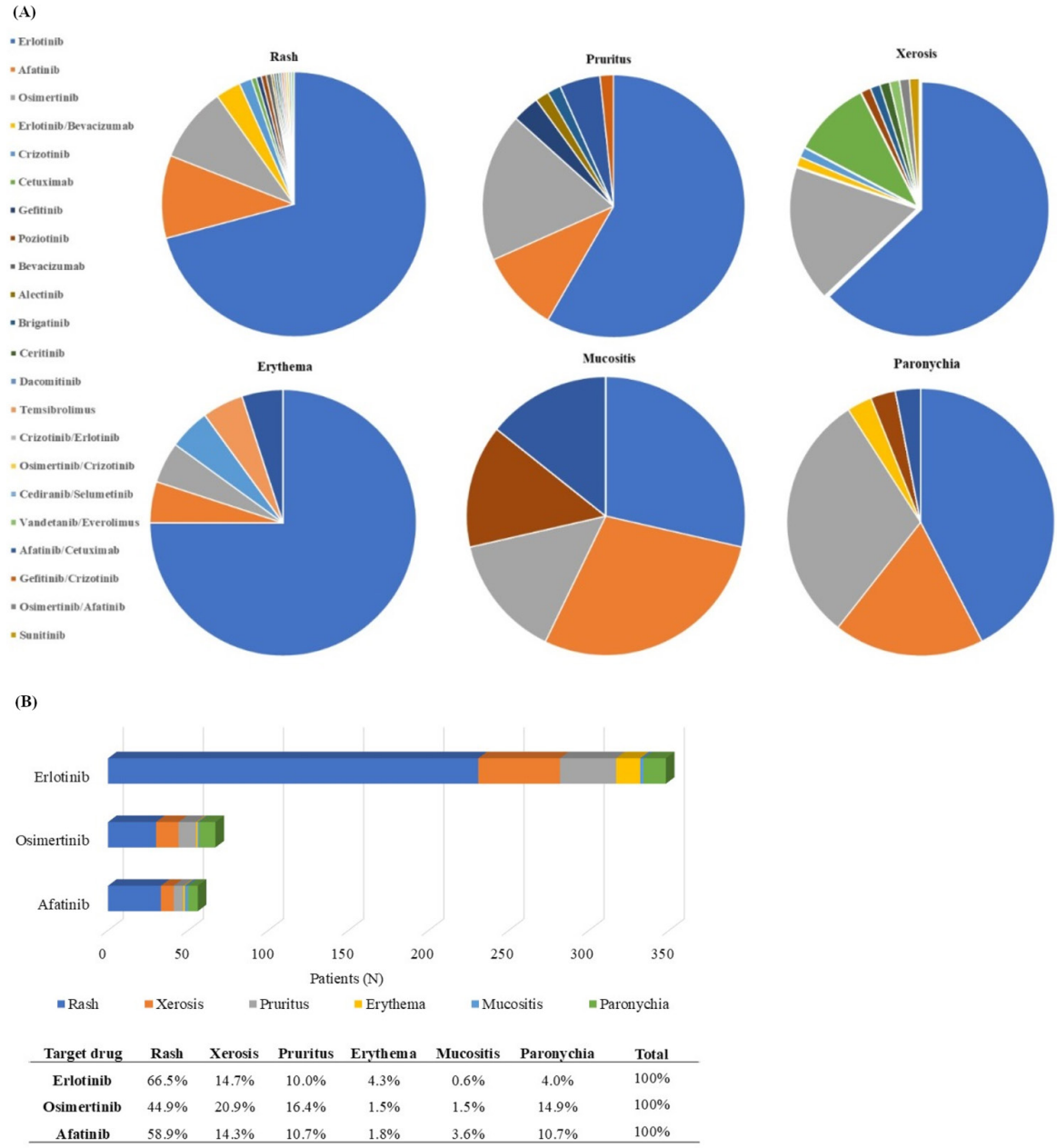
Most common toxicities (N*, %)	Target therapy N = 353 patients with 727 toxicities*	*EGFRIs* N = 329 patients with 681 toxicities*	*ALK/ROS1* inhibitors N = 10 patients with 40 toxicities*
Rash	432 (59.4)	403 (59.2)	27 (67.5)
Xerosis	93 (12.8)	87 (12.8)	7 (17.5)
Pruritus	62 (8.5)	60 (8.8)	4 (10.0)
Paronychia	51 (7.0)	51 (7.5)	0
Erythema	36 (5.0)	30 (4.4)	2 (5.0)
Mucositis	12 (1.7)	11 (1.6)	0
Nail changes	12 (1.7)	12 (1.8)	0

**N means the observed toxicities rather than the patient number. *Abbreviations*: EGFRIs, epidermal growth factor receptor inhibitors; ALK/ROS1, anaplastic lymphoma kinase/ c-ros oncogene 1.

Table 3. The incidence, severity, and onset days for skin toxicity in patients with different type of drugs

	Targeted therapy (*n* = 353)	*EGFR* inhibitors (*n* = 339)	ALK/ROS1 inhibitors (*n* = 10)
Rash, *n* (incidence, %)	327 (92.6)	306 (93.0)	8 (80.0)
Severity, *n* (%)			
Grade 1	170 (60.7)	161 (61.7)	4 (66.7)
Grade 2	44 (15.7)	43 (16.5)	1 (16.7)
Grade 3-4	65 (23.2)	56 (21.5)	1 (16.7)
NA	48	46	2
Onset days (median)	21	21	44
Xerosis, *n* (incidence, %)	82 (23.2)	76 (23.1)	3 (30.0)
Severity, *n* (%)			
Grade 1	45 (73.8)	41 (71.9)	3 (100.0)
Grade 2	4 (6.6)	4 (7.0)	-
Grade 3	12 (19.7)	12 (21.1)	-
NA	21	19	-
Onset days (median)	47	47	177
Erythema, *n* (incidence, %)	20 (5.7)	18 (5.5)	1 (10.0)
Severity, *n* (%)			
Grade 1	12 (75.0)	11 (73.3)	-
Grade 2	1 (6.3)	1 (6.7)	-
Grade 3	3 (18.8)	3 (20.0)	-
NA	4	3	-
Onset days (median)	30.5	28.5	109
Mucositis, *n* (incidence, %)	7 (2.0)	7 (2.1)	-
Severity, *n* (%)			
Grade 1	1 (20.0)	1 (20.0)	-
Grade 2	1 (20.0)	1 (20.0)	-
Grade 3	3 (60.0)	3 (60.0)	-
NA	2	2	-
Onset days (median)	34	34	-
Pruritus, *n* (incidence, %)	60 (17.0)	57 (17.3)	2 (20.0)
Severity, *n* (%)			
Grade 1	35 (74.5)	34 (77.3)	1 (50.0)
Grade 2	3 (6.4)	3 (6.8)	-
Grade 3	9 (19.1)	7 (15.9)	1 (50.0)
NA	13	13	-
Onset days (median)	36	34	163.5
Paronychia, *n* (incidence, %)	33 (9.3)	33 (10.0)	-
Severity, *n* (%)			
Grade 1	21 (84.0)	21 (84.0)	-
Grade 2	1 (4.0)	1 (4.0)	-
Grade 3	3 (12.0)	3 (12.0)	-
NA	8	8	-
Onset days (median)	69	69	-

Abbreviations: EGFR, epidermal growth factor receptor inhibitors; ALK/ROS1, anaplastic lymphoma kinase/ c-ros oncogene 1.



Target drug	Rash	Xerosis	Pruritus	Erythema	Mucositis	Paronychia	Total
Erlotinib	66.5%	14.7%	10.0%	4.3%	0.6%	4.0%	100%
Osimertinib	44.9%	20.9%	16.4%	1.5%	1.5%	14.9%	100%
Afatinib	58.9%	14.3%	10.7%	1.8%	3.6%	10.7%	100%

Figure 2. (A) The proportion of responsible drugs for skin toxicities. A total of 22 single agent or combined therapies were related to 6 primary skin toxicities. Erlotinib, afatinib and osimertinib were the most frequent associated drugs for rash, pruritus, erythema, mucositis, and paronychia. However, erlotinib, osimertinib and cetuximab were more common to cause xerosis. **(B)** The distribution and proportion of dermatologic toxicity in erlotinib, osimertinib, and afatinib. The skin toxicities had different distributions among the three predominate responsible drugs: rash was the most common skin toxicity and occurred more frequently in erlotinib (66.5%) than osimertinib (44.9%) and afatinib (58.9%), however, paronychia had a lower rate with erlotinib (4.0%) than osimertinib (14.9%) and afatinib (10.7%).

Table 2 (B). The distribution of common non-dermatologic toxicities in patients with different type of drugs

Most common toxicities (N*, %)	Target therapy N = 144 patients with 2,383 toxicities**	EGFRIs N = 89 patients with 1,857 toxicities**	ALK/ROS1 inhibitors N = 47 patients with 377 toxicities**
Fatigue	334 (14.0)	273 (14.7)	48 (12.7)
Diarrhea	355 (14.9)	321 (17.3)	36 (9.5)
Nausea	257 (10.8)	202 (10.9)	39 (10.3)
Anorexia	137 (5.7)	114 (6.1)	16 (4.2)
Vomiting	97 (4.1)	74 (4.0)	16 (4.2)
Anemia	81 (3.4)	60 (3.2)	17 (4.5)

**N means the observed toxicities rather than the patient number. *Abbreviations*: EGFRIs, epidermal growth factor receptor inhibitors; ALK/ROS1, anaplastic lymphoma kinase/ c-ros oncogene 1.

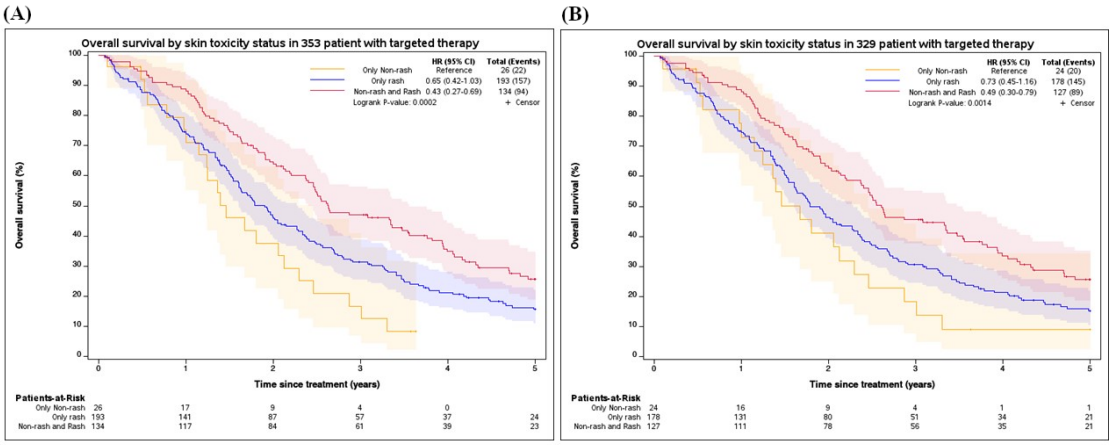


Figure 3. Kaplan-Meier curves of overall survival in targeted therapy-treated and *EGFR* inhibitors-treated patients with dermatologic toxicity respectively truncated at 5 years. (A) In all 353 targeted therapy-treated patients, patient with concurrent non-rash and rash had better survival than those with only non-rash or only rash dermatologic toxicity. **(B)** In 329 *EGFR* inhibitors-treated, patient with concurrent non-rash and rash had better survival than those with only non-rash or only rash dermatologic toxicity.

4. Discussion

Dermatologic toxicity is a commonly observed adverse effect of targeted therapy, reporting a frequency of 60.6% among NSCLC patients in our study. As is known, we observed that rash (59.3%) is the most frequent dermatologic toxicity associated with targeted therapy, consistent with that in the literature (*24*). Furthermore, we delineated the profile of non-rash dermatologic toxicities and discovered the strengthened prognostic predicting value of concurrent dermatologic toxicities in targeted therapy-treated patients.

Administration of targeted drugs is standard treatment for driver gene-mutated patients (*25*). Meanwhile, gene tests have become a routine recommendation by conventional methods, even novel next-generation sequencing for screening oncogenic targets (*26*). *EGFR* mutation is the most common targetable genetic driver alteration in lung adenocarcinoma, accounting for approximately 40% and 20% of NSCLC patients in Asian and non-Asia populations, respectively (*27*). Frequently administered *EGFR* inhibitors are divided into intracellular tyrosine kinase inhibitors (TKIs) and monoclonal antibodies inhibitors (mAbs) against the extracellular domain of *EGFR* (*28*). The main mechanism of *EGFR* inhibitors-related dermatologic toxicities is due to the prominent role of *EGFR* in maintaining dermatological homeostasis; *EGFR* inhibitors instigate pathological changes of growth and migration arrest and apoptosis, chemokine expression, and abnormal maturation and differentiation in skin cells, eventually, causing skin disorders (*29*).

In our study, dermatologic toxicity was mostly prevalent in erlotinib, afatinib, and osimertinib, which were typical three generations of EGFR tyrosine kinase inhibitors (EGFR-TKIs). To date, first-generation (gefitinib, erlotinib), second-generation (afatinib, dacomitinib), and third-generation (osimertinib) EGFR-TKIs are approved as standard management for sensitive *EGFR* mutations (*30*). The frequency of various dermatologic toxicities differed in first-, second-, third- generation of EGFR-TKIs in phase III trials were reported rash at 51.3%, 75.2%, 45.7%, stomatitis or mucositis at 11.2%, 27.5%, 21.7%, paronychia at 9.2%, 30.7%, 28.3%, respectively; additionally, xerosis occurred at 23%-36% in osimertinib and pruritus at 7% in gefitinib (*31*). Our results showed a higher frequency of rash in erlotinib than afatinib, and similar incidences of xerosis and pruritis with those in clinical trials. For non-dermatologic toxicities, most notably, diarrhea occurred at any grade (grade $\geq$ 3) was 45.3% (2.6%), 79.2% (6.8%), 49.1% (1.6%) for first-, second-, third-generation of EGFR-TKIs respectively, as validated by our study (*31*).

We validated that rash was the earliest occurring and most frequent dermatologic toxicity caused by *EGFR* inhibitors, and median onset time was in the range of 2-4 weeks (*32*). Further subgroup analysis on the patients showed that less patients had non-rash (7.4%) than only rash (54.7%) or concurrent (38.0%) dermatologic toxicities. More specifically, we found that pruritus, paronychia, and mucositis tended to occur with rash. Pruritus concurrent with rash may be related to the inflammatory response and probably increased keratinocyte expression in growth factors significant to mast cells (*33*). Paronychia is a disorder characterized by an inflammatory process involving the soft tissues around the nail (*34*), which emerged latest at a median onset 69 days in our study.

Although dermatologic symptoms induced by *EGFR* inhibitors appeared to be significantly correlated with poor quality of life and compliance (*35*), rash in *EGFR* inhibitors has been varied to be a surrogate biomarker of therapeutic efficacy and improved survival for *EGFR*-mutated patients (*36,37*) and validated our previous study (*14*). This study highlighted that non-

Table 4. The characteristics of 353 patients with targeted therapy in three dermatologic toxicity groups

Variable	*n*	Events (%)	Median Years	5-year survival % (95% CI)	Cox univariate hazard ratio (95% CI)	Cox univariate score *p* value	Cox multivariate hazard ratio (95% CI)	Cox multivariate likelihood ratio *p* value (*n* = 352)
Toxicity						< 0.001		< 0.001
Only Non-rash	26	22 (85)	1.5	8.4 (0, 19.6)	-		--	
Non-rash and Rash	134	94 (70)	2.6	27.7 (19.9, 35.6)	0.45 (0.28, 0.72)		0.48 (0.30, 0.77)	
Only rash	193	157 (81)	1.9	16.3 (11.0, 21.7)	0.70 (0.45, 1.09)		0.75 (0.48, 1.18)	
Sex						0.444		
0 = Female	207	174 (84)	2.3	21.9 (16.0, 27.7)	-			
1 = Male	146	126 (86)	2.0	17.8 (11.5, 24.1)	1.09 (0.87, 1.38)			
Race						0.670		
White	306	263 (86)	2.1	20.0 (15.4, 24.6)	-			
Other	46	36 (78)	2.3	21.7 (9.1, 34.4)	0.93 (0.65, 1.31)			
Ever smoke cigarettes						0.019		0.074
0 = Never	224	165 (74)	2.3	24.1 (18.3, 29.9)	-		-	
1 = Ever	129	108 (84)	2.0	13.3 (7.3, 19.4)	1.34 (1.05, 1.70)		1.26 (0.98, 1.62)	
Cell type						< 0.001		< 0.001
Adenocarcinoma	317	240 (76)	2.3	21.5 (16.8, 26.2)	-		-	
Other	36	33 (92)	1.3	8.3 (0, 17.4)	1.97 (1.37, 2.84)		2.04 (1.40, 2.96)	
Drug treatment/combination						< 0.001		< 0.001
Drug therapy	252	209 (83)	1.9	14.1 (9.7, 18.6)	-		-	
Drug therapy & other	101	64 (63)	3.3	34.8 (25.3, 44.3)	0.54 (0.41, 0.71)		0.57 (0.42, 0.75)	
Treatment line of targeted therapy, *n* (%)						0.011		0.087
First line	191	141 (74)	2.5	23.7 (17.5, 29.9)	-		-	
Other	156	129 (83)	1.8	15.0 (9.2, 20.8)	1.36 (1.07, 1.73)		1.24 (0.97, 1.60)	
Treatment response						< 0.001		< 0.001
Response	184	128 (70)	2.6	27.2 (20.6, 33.9)	-		-	
No response	168	145 (86)	1.6	12.4 (7.3, 17.5)	1.79 (1.41, 2.28)		1.53 (1.20, 1.96)	
Toxicity severity						0.937		
Mild	206	176 (85)	2.1	20.3 (14.6, 25.9)	-			
Moderate & Severe & life threatening	142	120 (85)	2.4	19.4 (12.6, 26.2)	1.01 (0.80, 1.27)			
Age at diagnosis	353	273 (77)	2.1	20.1 (15.8, 24.5)	1.017 (1.007, 1.027)	0.001	1.013 (1.003, 1.023)	0.010

rash dermatologic toxicity played a critical role in predicting better response to treatment when looking into the higher ORR in patients with concurrent (67.9%) or only non-rash (53.8%) dermatologic toxicities than those with only rash (41.1%). Furthermore, the relationship between dermatologic toxicity and survival analysis showed longer survival among patients with concurrent toxicity, providing more detailed evidence of targeted therapy-induced dermatologic toxicity predicting positive treatment response and OS benefit (*38*). Considering prevalence of non-rash dermatologic toxicity in the concurrent dermatologic toxicity group, rash with pruritus, paronychia or mucositis appeared to be associated with improved outcomes. Therefore, dermatologic toxicity as a drug-efficacy marker for patients with treated therapy called for further investigations to differentiate various toxicities, especially non-rash dermatologic toxicities.

ALK and *ROS1* define unique subsets of NSCLC patients highly sensitive to *ALK*/*ROS1* targeted drugs. However, *ALK*+ and *ROS1*+ have a low frequency of 1.7% and 2.9% among NSCLC patients, respectively (*39*). Therefore, only 63 patients who received *ALK*/*ROS1* inhibitors were included in our study. Dermatologic toxicity was uncommon in the toxicity profile of *ALK/ROS1* inhibitors. Rash was the primary complaint of dermatologic side effects, reporting rates of any grade (grade ≥ 3) at 8.4% in crizotinib, 14.7% (0.9%) in alectinib, 15.4% (0.7%) in brigatinib, 12.4% in ceritinib, 62.9% (8.0%) in ensartinib, 6.6% (0.2%) in lorlatinib (*40*). We found that *ALK/ROS1* inhibitors had a low incidence of dermatologic toxicity (15.9%), as reported in previous studies.

Due to the nature of back-reviewed information, clinical data unavoidably produced some bias, such as the inaccurately reported and recorded toxicity information, even though we have carefully defined each variable. Additionally, patients with unavailable or unjudgeable toxicity were not included because of outside medical records, less than one-month treatment duration and loss of follow-up, which might lead to underestimation of the occurrence and effects of dermatologic toxicity.

In conclusion, non-rash dermatologic toxicity appeared to be milder than rash compared to rash toxicity, but might be a stronger protective indicator for treatment response and survival length in patients who received targeted therapy or *EGFR* inhibitors. Severity of dermatologic toxicity showed no correlation with survival length. Oncologists and dermatologists need to collaborate effectively on the awareness, prevention, and treatment of dermatologic toxicity associated with targeted drugs.

Funding: This work was supported by the National Institutes of Health [Grant Numbers R03 CA77118, R01 CA80127 and R01 CA84354]; the National Institute on Aging [Grant Numbers R01 AG034676 and R01 AG052425]; the Mayo Clinic Foundation. Yanmei Peng was funded by National Natural Science Foundation of China (8187151262) and a grant (Z221100007422103) from the Beijing Municipal Science and Technology Commission.

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin. 2023; 73:17-48.
2. Hines JB, Bowar B, Levine E, Esposito A, Garassino MC, Bestvina CM. Targeted toxicities: Protocols for monitoring the adverse events of targeted therapies used in the treatment of non-small cell lung cancer. Int J Mol Sci. 2023; 24:9429.
3. O'Kane GM, Labbé C, Doherty MK, Young K, Albaba H, Leighl NB. Monitoring and management of immune-related adverse events associated with programmed cell death protein-1 axis inhibitors in lung cancer. Oncologist. 2017; 22:70-80.
4. Kroschinsky F, Stölzel F, von Bonin S, Beutel G, Kochanek M, Kiehl M, Schellongowski P; Intensive Care in Hematological and Oncological Patients (iCHOP) Collaborative Group. New drugs, new toxicities: Severe side effects of modern targeted and immunotherapy of cancer and their management. Crit Care. 2017; 21:89.
5. Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. Lancet. 2021; 398:535-554.
6. Tan AC, Tan DSW. Targeted therapies for lung cancer patients with oncogenic driver molecular alterations. J Clin Oncol. 2022; 40:611-625.
7. Crinò L, Foglietta J, Hamzaj A. Lung cancer. J Thorac Oncol. 2007; 2:S24-S26.
8. Dempke WC, Suto T, Reck M. Targeted therapies for non-small cell lung cancer. Lung Cancer. 2010; 67:257-274.
9. Tischer B, Huber R, Kraemer M, Lacouture ME. Dermatologic events from EGFR inhibitors: The issue of the missing patient voice. Support Care Cancer. 2017; 25:651-660.
10. Chan JC, Lee YH, Liu CY, Shih HH, Tsay PK, Tang WR. A correlational study of skin toxicity and quality of life in patients with advanced lung cancer receiving targeted therapy. J Nurs Res. 2019; 27:e51.
11. Pérez-Soler R, Chachoua A, Hammond LA, Rowinsky EK, Huberman M, Karp D, Rigas J, Clark GM, Santabárbara P, Bonomi P. Determinants of tumor response and survival with erlotinib in patients with non--small-cell lung cancer. J Clin Oncol. 2004; 22:3238-3247.
12. Kainis I, Syrigos N, Kopitopoulou A, Gkiozos I, Filiou E, Nikolaou V, Papadavid E. Erlotinib-associated rash in advanced non-small cell lung cancer: Relation to clinicopathological characteristics, treatment response, and survival. Oncol Res. 2018; 26:59-69.
13. Nasu S, Suzuki H, Shiroyama T, Tanaka A, Iwata K, Ryota N, Ueda Y, Takata SO, Masuhiro K, Morita S, Morishita N, Okamoto N, Hirashima T. Skin rash can be a useful marker for afatinib efficacy. Anticancer Res. 2018; 38:1783-1788.
14. Peng Y, Kumar AV, Gu Z, Kosuru N, Wampfler JA, Reck Dos Santos PA, Yu NY, D'Cunha J, Ernani V, Costello

CM, Yang P. Dermatologic toxicity associated with targeted and immunotherapies in stage-IV non-small cell lung cancer patients: A 14-year cohort. J Am Acad Dermatol. 2024; 91:172-175.
15. Jatoi A, Williams B, Nichols F, Marks R, Aubry MC, Wampfler J, Finke EE, Yang P. Is voluntary vitamin and mineral supplementation associated with better outcome in non-small cell lung cancer patients? Results from the Mayo Clinic lung cancer cohort. Lung Cancer. 2005; 49:77-84.
16. Dai J, Liu M, Swensen SJ, Stoddard SM, Wampfler JA, Limper AH, Jiang G, Yang P. Regional emphysema score predicting overall survival, quality of life, and pulmonary function recovery in early-stage lung cancer patients. J Thorac Oncol. 2017; 12:824-832.
17. Luo YH, Luo L, Wampfler JA, Wang Y, Liu D, Chen YM, Adjei AA, Midthun DE, Yang P. 5-year overall survival in patients with lung cancer eligible or ineligible for screening according to US Preventive Services Task Force criteria: A prospective, observational cohort study. Lancet Oncol. 2019; 20:1098-1108.
18. Luo YH, Liu H, Wampfler JA, Tazelaar HD, Li Y, Peikert T, Liu D, Leventakos K, Chen YM, Yang Y, Chiou SH, Yang P. Real-world efficacy of osimertinib in previously EGFR-TKI treated NSCLC patients without identification of T790M mutation. J Cancer Res Clin Oncol. 2022; 148:2099-2114.
19. Mountain CF. Revisions in the international system for staging lung cancer. Chest. 1997; 111:1710-1717.
20. Shepherd FA, Crowley J, Van Houtte P, Postmus PE, Carney D, Chansky K, Shaikh Z, Goldstraw P. The International Association for the Study of Lung Cancer lung cancer staging project: Proposals regarding the clinical staging of small cell lung cancer in the forthcoming (seventh) edition of the tumor, node, metastasis classification for lung cancer. J Thorac Oncol. 2007; 2:1067-1077.
21. Huynh Dagher S, Blom A, Chabanol H, Funck-Brentano E. Cutaneous toxicities from targeted therapies used in oncology: Literature review of clinical presentation and management. Int J Womens Dermatol. 2021; 7:615-624.
22. National Institutes of Health NCI. Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0. *https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/CTCAE_v5_Quick_Reference_8.5x11.pdf* (accessd August 4, 2023).
23. Eisenhauer EA, Therasse P, Bogaerts J, *et al.* New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer. 2009; 45:228-247.
24. Pugliese SB, Neal JW, Kwong BY. Management of dermatologic complications of lung cancer therapies. Curr Treat Options Oncol. 2015; 16:50.
25. Fois SS, Paliogiannis P, Zinellu A, Fois AG, Cossu A, Palmieri G. Molecular epidemiology of the main druggable genetic alterations in non-small cell lung cancer. Int J Mol Sci. 2021; 22:612.
26. Planchard D, Popat S, Kerr K, Novello S, Smit EF, Faivre-Finn C, Mok TS, Reck M, Van Schil PE, Hellmann MD, Peters S. Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2018; 29:iv192-iv237.
27. Zhang YL, Yuan JQ, Wang KF, Fu XH, Han XR, Threapleton D, Yang ZY, Mao C, Tang JL. The prevalence of EGFR mutation in patients with non-small cell lung cancer: A systematic review and meta-analysis. Oncotarget. 2016; 7:78985-78993.
28. Ray M, Salgia R, Vokes EE. The role of EGFR inhibition in the treatment of non-small cell lung cancer. Oncologist. 2009; 14:1116-1130.
29. Braden RL, Anadkat MJ. EGFR inhibitor-induced skin reactions: Differentiating acneiform rash from superimposed bacterial infections. Support Care Cancer. 2016; 24:3943-3950.
30. Gelatti ACZ, Drilon A, Santini FC. Optimizing the sequencing of tyrosine kinase inhibitors (TKIs) in epidermal growth factor receptor (EGFR) mutation-positive non-small cell lung cancer (NSCLC). Lung Cancer. 2019; 137:113-122.
31. Zhou JY, Liu SY, Wu YL. Safety of EGFR-TKIs for EGFR mutation-positive non-small cell lung cancer. Expert Opin Drug Saf. 2020; 19:589-599.
32. Shah DR, Shah RR, Morganroth J. Tyrosine kinase inhibitors: Their on-target toxicities as potential indicators of efficacy. Drug Saf. 2013; 36:413-426.
33. Bryce J, Boers-Doets CB. Non-rash dermatologic adverse events related to targeted therapies. Semin Oncol Nurs. 2014; 30:155-168.
34. Califano R, Tariq N, Compton S, Fitzgerald DA, Harwood CA, Lal R, Lester J, McPhelim J, Mulatero C, Subramanian S, Thomas A, Thatcher N, Nicolson M. Expert consensus on the management of adverse events from EGFR tyrosine kinase inhibitors in the UK. Drugs. 2015; 75:1335-1348.
35. Tseng LC, Chen KH, Wang CL, Weng LC. Effects of tyrosine kinase inhibitor therapy on skin toxicity and skin-related quality of life in patients with lung cancer: An observational study. Medicine (Baltimore). 2020; 99:e20510.
36. Petrelli F, Borgonovo K, Cabiddu M, Lonati V, Barni S. Relationship between skin rash and outcome in non-small-cell lung cancer patients treated with anti-EGFR tyrosine kinase inhibitors: A literature-based meta-analysis of 24 trials. Lung Cancer. 2012; 78:8-15.
37. Perez-Soler R. Rash as a surrogate marker for efficacy of epidermal growth factor receptor inhibitors in lung cancer. Clin Lung Cancer. 2006; 8 Suppl 1:S7-S14.
38. Faehling M, Eckert R, Kuom S, Kamp T, Stoiber KM, Schumann C. Benefit of erlotinib in patients with non-small-cell lung cancer is related to smoking status, gender, skin rash and radiological response but not to histology and treatment line. Oncology. 2010; 78:249-258.
39. Bergethon K, Shaw AT, Ou SH, *et al.* ROS1 rearrangements define a unique molecular class of lung cancers. J Clin Oncol. 2012; 30:863-870.
40. Tao Y, Zhou Y, Tang L, Chen H, Feng Y, Shi Y. Toxicity profile of anaplastic lymphoma kinase tyrosine kinase inhibitors for patients with non-small cell lung cancer: A systematic review and meta-analysis. Invest New Drugs. 2022; 40:831-840.

Received December 26, 2024; Revised February 15, 2025; Accepted February 27, 2025.

**Address correspondence to:*
Ping Yang, Division of Epidemiology, Department of Quantitative Health Sciences, Mayo Clinic, 13400 E. Shea Blvd. Scottsdale, AZ 85259, United States of America.
E-mail: Yang.Ping@mayo.edu

Released online in J-STAGE as advance publication March 3, 2025.

Original Article

DOI: 10.5582/bst.2024.01393

Comparative analysis of human gut bacterial microbiota between shallow shotgun metagenomic sequencing and full-length 16S rDNA amplicon sequencing

Suwalak Chitcharoen[1,2], Vorthon Sawaswong[3], Pavit Klomkliew[4], Prangwalai Chanchaem[4], Sunchai Payungporn[4,*]

[1] Department of Microbiology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand;
[2] Research and Diagnostic Center for Emerging Infectious Diseases, Khon Kaen University, Khon Kaen, Thailand;
[3] Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok, Thailand;
[4] Center of Excellence in Systems Microbiology, Department of Biochemistry, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand.

SUMMARY: The human gut microbiome is increasingly recognized as important to health and disease, influencing immune function, metabolism, mental health, and chronic illnesses. Two widely used, cost-effective, and fast approaches for analyzing gut microbial communities are shallow shotgun metagenomic sequencing (SSMS) and full-length 16S rDNA sequencing. This study compares these methods across 43 stool samples, revealing notable differences in taxonomic and species-level detection. At the genus level, *Bacteroides* was most abundant in both methods, with *Faecalibacterium* showing similar trends but *Prevotella* was more abundant in full-length 16S rDNA. Genera such as *Alistipes* and *Akkermansia* were more frequently detected by full-length 16S rDNA, whereas *Eubacterium* and *Roseburia* were more prevalent in SSMS. At the species level, *Faecalibacterium prausnitzii*, a key indicator of gut health, was abundant across both datasets, while *Bacteroides vulgatus* was more frequently detected by SSMS. Species within *Parabacteroides* and *Bacteroides* were primarily detected by 16S rDNA, contrasting with higher SSMS detection of *Prevotella copri* and *Oscillibacter valericigenes*. LEfSe analysis identified 18 species (9 species in each method) with significantly different detection between methods, underscoring the impact of methodological choice on microbial diversity and abundance. Differences in classification databases, such as Ribosomal Database Project (RDP) for 16S rDNA and Kraken2 for SSMS, further highlight the influence of database selection on outcomes. These findings emphasize the importance of carefully selecting sequencing methods and bioinformatics tools in microbiome research, as each approach demonstrates unique strengths and limitations in capturing microbial diversity and relative abundances.

Keywords: bacterial profile, microbiome, oxford nanopore technologies (ONT), ion torrent sequencing

1. Introduction

The gut microbiome is a complex ecosystem consisting of a diverse community of microorganisms residing in the human gastrointestinal tract. Numerous studies have demonstrated that a diverse and balanced population of gut microbiota is crucial for maintaining gut and overall health by facilitating digestion, nutrient absorption, and supporting the immune system. Furthermore, mounting evidence has suggested that dysbiosis, an imbalance in the composition and functionality of gut microbiota, is directly or indirectly associated with the pathogenesis of various diseases, including obesity (*1*), diabetes, chronic kidney disease (CKD) (*2*), liver diseases, colorectal cancer (CRC) or adenoma (*3*), and even mental and neurodegenerative disease (*4-6*). Therefore, modulation of the gut microbiome composition has been proposed as a potential therapeutic target, and dietary interventions have been suggested as a means to achieve this goal (*7*). In particular, patients with autoimmune diseases display reduced levels of beneficial bacteria, such as *Bifidobacterium* spp., *Faecalibacterium* spp., *Roseburia* spp., and *Coprococcus eutactus*, alongside increased levels of pathogenic bacteria like *Escherichia coli*, *Staphylococcus aureus*, and *Clostridioides difficile*, accompanied by microbial-driven TH17/TH1 activation and reduced Regulatory T cells, worsening inflammation (*8-10*). Consequently, modulating gut microbiota composition is proposed as a therapeutic target, and dietary interventions are suggested as a viable approach

for achieving this modulation (*11*).

Identifying microorganisms at the species level offers precise insights that can aid clinicians in designing targeted treatments to promote gut health and mitigate disease risk. For instance, within the genus *Bifidobacterium*, *Bifidobacterium lactis* has been shown to reduce the risk of diarrhea and fever in children and infants (*11*), whereas *Bifidobacterium bifidum* is known to enhance the immune system and combat pathogens (*12*). Similarly, *Faecalibacterium prausnitzii*, a member of the Firmicutes phylum, is positively correlated with gut health and plays a role in reducing inflammation and colorectal cancer risk. However, health effects are not uniform across all species within the Firmicutes phylum (*13,14*). Current methods for microbiota analysis primarily include 16S rDNA amplicon sequencing and shotgun metagenomics. The 16S rDNA amplicon sequencing method is based on amplifying a specific region of the 16S rRNA gene, allowing the identification of distinct taxa through variations in the less-conserved regions (*15*). This approach is relatively cost-effective and straightforward. However, taxa assignments are based on a single genomic region, which can introduce amplification biases and affect taxonomic representation due to primer choice and amplification error (*16,17*). The 16S rDNA contains nine hypervariable regions (V1–V9) surrounded by conserved regions, with species-specific variants that enable community-level identification down to the genus level. Full-length 16S rDNA sequencing facilitates species-level identification(*18,19*). In contrast, shotgun metagenomics sequences the entire microbial community's DNA, necessitating greater sequencing depth, thus increasing costs, analytical complexity, and potential host DNA contamination (*20*). The downstream analysis in shotgun metagenomics relies on reference databases for genome assembly, which can result in false positives (*21*). Despite these limitations, shotgun metagenomics provides comprehensive microbial genomic information, including gene function analysis and insights into other microbiome components, such as fungi and viruses. In clinical settings, balancing accuracy, cost, and processing time is vital for achieving species-level microbial profiling.

Oxford Nanopore Technologies (ONT) offers ultra-long nucleic acid sequencing, with read lengths exceeding 2 million base pairs, which enables full-length 16S rDNA sequencing can improves taxonomic resolution by providing a comprehensive sequence of informative sites. Additionally, ONT's devices offer real-time data acquisition, allowing for immediate insights during sequencing runs. This sequencing approach facilitates faster (sequencing time 1-2 hours) and more accurate microbial community analysis (*22-24*). For metagenomic approaches, studies have shown that ~7 Gb of paired-end sequencing data are necessary to achieve > 20X coverage for microbes at > 1% relative abundance, indicating that shallow shotgun metagenomics sequencing (SSMS) is viable for preliminary screening (*25,26*). Ion Torrent offers short-read sequencing platforms that leverage semiconductor-based technology to deliver high-throughput data with rapid turnaround times. Notably, the GeneStudio™ S5 System enhances efficiency with automated library preparation, enabling high-throughput sequencing data in approximately 2 to 4 hours while maintaining ease of use (*27*). These platforms are cost-effective, making them ideal for time-sensitive applications such as SSMS.

However, comparative studies between SSMS and full-length 16S rDNA sequencing remain limited. Most prior research has focused on comparing hypervariable regions of the 16S rDNA gene, such as V4 or V3-V4, using short-read sequencing. These studies have reported biases in taxonomic detection due to the targeted nature of hypervariable region sequencing, which, despite being cost-effective, can lead to incomplete microbial profiles (*15,28*). While some studies have explored full-length 16S rDNA sequencing, they have primarily compared it with deep shotgun metagenomics rather than shallow shotgun sequencing. These studies indicate that full-length 16S rDNA sequencing provides a more comprehensive representation of dominant microorganisms and offers enhanced taxonomic resolution for low-abundance taxa in food-related matrices (*29*).

This research gap underscores the need for a direct comparison between SSMS and full-length 16S rDNA sequencing, particularly in terms of cost, time efficiency, and taxonomic resolution across diverse microbial communities. To address this, our study employs ONT-based full-length 16S rDNA sequencing alongside SSMS using the Ion GeneStudio S5 System to analyze the gut microbiota of healthy individuals. We compare alpha and beta diversity, identify taxa unique to each method, and discuss the broader implications of these sequencing strategies for gut microbiome research and future applications.

2. Methods

2.1. Sample collection and DNA extraction

In this study, forty-three stool samples were collected from consenting participants using collection tubes that incorporated DNA/RNA Shield (Zymo Research, USA) to preserve microbial specimens. These samples were subsequently stored at -20 °C until DNA extraction was performed. For DNA extraction, each fecal sample was thawed on ice, and 20 mg of material was processed using the ZymoBIOMICS DNA Miniprep kit (Zymo Research, USA) according to the manufacturer's protocol. The extracted DNA was preserved at -20 °C until further processing for sequencing.

2.2. Library construction and sequencing

2.2.1. Full length 16S rDNA nanopore sequencing

The full-length of bacterial 16S rDNA was polymerase chain reaction (PCR) amplified for 20 cycles using primers for targeting regions V1-V9 of the 16S rDNA. Primers were described Forword: 5'-TTTCTGTTGGT GCTGATATTGCAGRGTTYGATYMTGGCTCAG-3' and Reverse: 5'-ACTTGCCTG TCGCTCTATCTTCC GGYTACCTTGTTACGACTT-3' (*30*). The 20 µL PCR reaction contained 1 µg of DNA template, 0.2 µM of each primer, 0.2 mM of dNTPs, and 0.4 U of Phusion DNA Polymerase (Thermo Scientific, USA). The 1 µg of DNA template were used in the total volume (20 µL) of PCR reaction. The barcode sequences were added to PCR products using the PCR Barcoding Expansion Kit (Oxford Nanopore Technologies, UK). The products were checked by 1% agarose gel electrophoresis and purified using the QIAquick® PCR Purification Kit (QIAGEN, Germany). The samples were pooled at equal concentration and purified using AMPure XP beads (Beckman Coulter, USA). The final products of full-length V1-V9 region of 16S rDNA, were sequenced using Ligation Sequencing Kit (Oxford Nanopore Technologies, UK) and flow cell version R10.4 (Oxford Nanopore Technologies, UK).

2.2.2. Shallow shotgun metagenomic sequencing (SSMS)

Library preparation was performed using the Ion Xpress™ Fragment Library Kit (Thermo Fisher Scientific) with 100 ng of DNA as input. Adapter ligation, size selection, nick repair, and amplification followed the manufacturer's protocol. Sequencing was conducted on the Ion GeneStudio S5 System (Thermo Fisher Scientific, USA).

2.3. Bioinformatics analysis

2.3.1. 16S rDNA nanopore sequencing (16S rDNA)

The FAST5 files were base called by using Guppy basecaller software v6.0.7 (*31*) (Oxford Nanopore Technologies, UK) with a super-accuracy model to generate pass reads (FASTQ format) with a minimum acceptable quality score (Q > 10). The quality of reads was examined by MinIONQC (*32*). Then, FASTQ sequences were demultiplexed and adaptor-trimmed using Porechop v0.2.4 (*https://github.com/rrwick/Porechop*). The filtered reads were then clustered, polished, and taxonomically classified by NanoCLUST (*33*) based on the size sequences for the V1-V9 region of 16S rDNA sequences from Ribosomal Database Project (RDP) database (*34*). The abundance taxonomic assignment data were converted into QIIME2 software v2021.2 (*35*) data format for illustrating the richness and evenness of bacterial species based on their taxa abundances.

2.3.2. Shallow shotgun metagenomic sequencing (SSMS)

The taxonomic classification and abundance estimation of the shallow shotgun metagenomic sequencing data obtained from the gut microbiome samples were performed using Kraken2 (*36*) and Bracken2 (*37*), respectively. The raw reads were aligned against the PlusPF database (version 9/19/2020 available at *https://benlangmead.github.io/aws-indexes/k2*), which includes both the NCBI and RefSeq microbial genomes and has been demonstrated to have higher accuracy than other databases. Bracken2 was then used to estimate the taxonomic abundances at different levels of the classification hierarchy by adjusting the classification counts based on the distribution of read lengths. The resulting output was a table of taxonomic abundances at various levels of the classification hierarchy, which provided insight into the composition of the gut microbiome. An overview method used in this study is illustrated in Figure 1.

2.4. Statistical analysis

Statistical analyses were conducted on both 16S rDNA and SSMS datasets. The data for bacteria with a relative abundance greater than 1% were visualized using

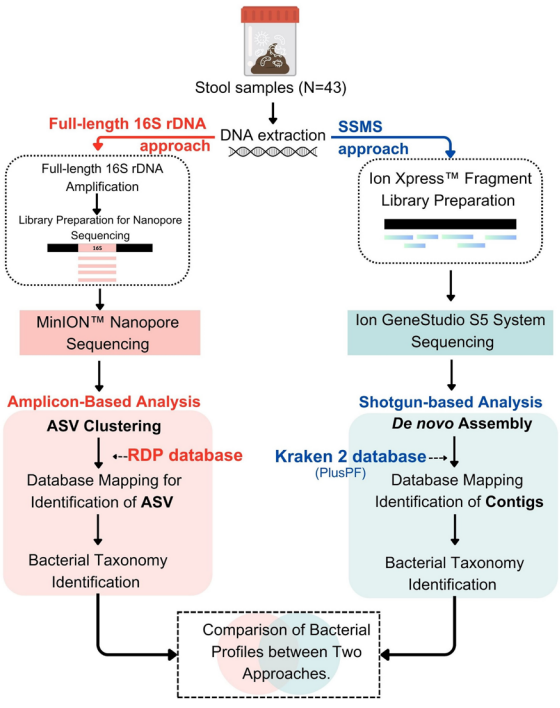


Figure 1. Overview of analytical plan for bacterial taxonomy identification from stool samples using Full-Length 16S rDNA Amplicon Sequencing (Full-length 16S rDNA) and Shallow Shotgun Metagenome sequencing (SSMS).

threshold cut-off values (*38*). Alpha diversity measures, including Observed Species and Chao1, were utilized to assess species richness, while the Shannon and Simpson indices were employed to evaluate both richness and evenness. Each alpha diversity measure was calculated using the R software (version 3.5.0) with the vegan package, aiming to examine microbiota diversity across the datasets. The relative abundance was compared between the two approaches using the Wilcoxon signed-rank test, which was performed in Python using the Pandas and SciPy libraries. Beta-diversity was analyzed using PERMANOVA tests based on Bray-Curtis and Jaccard distances, as implemented in the MicrobiomeAnalyst tools (*39*). Statistical significance was attributed to *P*-values less than 0.01 for ensuring robust statistical interpretation.

2.5. Data availability

The raw sequence reads generated during this study have been submitted to the NCBI Sequence Read Archive database under the BioProject accession number PRJNA1089554. The raw reads for full-length 16S rDNA sequencing and SSMS are available under BioSample accessions SAMN40544624 and SAMN40544783, respectively.

2.6. Ethics statement

The experiments were conducted after obtaining the approval of Ethical Committee of the Khon Kaen University Ethics Committee for Human Research on HE681056. This Research was conducted in accordance with the Declaration of Helsinki.

3. Results

3.1. Sequencing data

In this microbiome sequencing study, we compared two different sequencing methods: 16S rDNA full-length by Oxford Nanopore Technology sequencing (16S rDNA) and shallow shotgun metagenomic sequencing by Ion Torrent System (SSMS). The dataset comprised 43 samples, yielding a range of 3,622 to 89,831 raw reads for the 16s rDNA (mean: 11,677 ± 2,38) and 1,590,861 to 3,200,974 raw reads for the SSMS 1,590,861 to 3,200,974 (mean: 2,449,982 ± 44,489). The mapped reads for 16S rDNA ranged from 1,453 to 62,973 with an average of 8,489 ± 1,705. For SSMS, the mapped reads from 501,452 to 1,846,203, with a mean of 1,167,404 ± 52,511. Percentages of mapped reads were 46.96% (95% CI: 4.51-19.02%) and 71.91% (95% CI: 40.12-87.87%) for 16S rDNA full-length and shallow shotgun metagenomic sequencing, respectively (Table 1). Our findings provide important insights into the performance of these two sequencing methods and their potential application in microbiome studies.

3.2. Diversity of bacterial composition between shallow shotgun metagenomic sequencing and full-length 16S rDNA amplicon sequencing

To investigate gut microbial patterns associated with technical methods, we compared available microbiome data generated by two different approaches (shallow shotgun metagenomic sequencing (SSMS) and full-length 16S rDNA sequencing (full-length 16S rDNA)). Initial analysis without data cut-off parameter revealed a core microbiome of 7 phyla and 81 bacterial species common to both protocols. However, SSMS demonstrated greater sensitivity, identifying an additional 31 phyla, 1,235 genera, and 2,613 bacterial species. The full-length 16S rDNA approach also detected unique microbes, with 181 species belonging to 109 genera not found in the SSMS dataset. Applying a 1% abundance threshold narrowed the focus to a diverse bacterial community composed of 7 distinct phyla, 83 genera, and 205 species. There was significant overlap between the methods 47 genera (37.93%) and 113 species (54.00%) detected by both approach), while each method also demonstrated unique detection capabilities (13 genera (10.13%) and 38 species (16.50%) unique to SSMS, 23 genera (51.90%) and 54 species (29.50%) unique to full-length 16S rDNA sequencing (Figure 2). Furthermore, we observed discrepancies in bacterial nomenclature

Table 1. Sequencing statistic

Types	Statistic value	Sequencing approach	
		full-length 16S rDNA	SSMS
No. of Reads	Minimum	3,622	6,363,443
	Maximum	89,831	12,803,895
	Mean±Std.	11,677 ± 2,387	9,799,926 ± 177,956
	95% CI of median (0.9685)	6,737 - 9,001	9,151,579 - 10,276,515
No. of Mapped Reads	Minimum	1,453	501,452
	Maximum	62,973	1,846,203
	Mean ± Std.	8,489 ± 1,705	1,167,404 ± 52,511
	95% CI of median (0.9685)	4,436	1,000,816
%mapped reads	Mean ± Std.	71.91 ± 1.562	11.99 ± 0.5205
	95% CI of median (0.9685)	40.12 - 87.87	4.51 - 19.02

across databases. For example, *Bacteroides vulgatus* was designated as *Phocaeicola vulgatus, Eubacterium eligens* as *Lachnospira eligens*, and *Clostridium bolteae* as *Enterocloster bolteae* (Supplemental Table S1, *https://www.biosciencetrends.com/supplementaldata/249*). After consolidating taxa names, our analysis identified 200 bacterial species with 79 genera. The full-length 16S rDNA sequencing method detected a higher number of species (161 species) compared to SSMS (96 species) (Figure 2) and will therefore be used for further analysis.

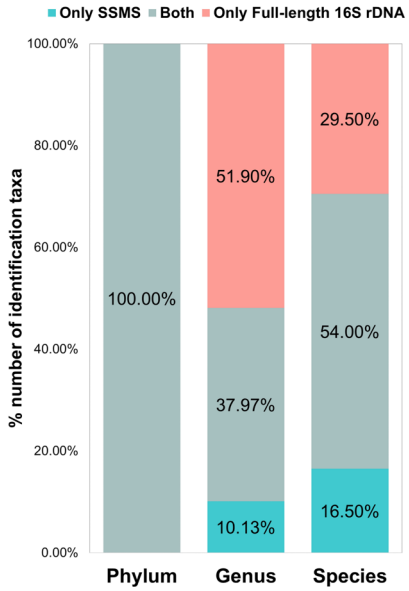


Figure 2. Bacterial taxonomic identification counts by sequencing approach (phyla, genera, and species).

Alpha diversity was quantified by observed richness (Figure 3A), Chao1 index (Figure 2B), Shannon's diversity (Figure 3C), and Simpson's diversity (Figure 3D), to evaluate bacterial richness and evenness across the two identification approaches. Analysis of full-length 16S rDNA sequencing revealed significantly higher bacterial diversity compared to the SSMS method, as demonstrated by all four diversity indices (Wilcoxon test, $p < 0.01$). Beta diversity analysis of the gut microbiome, assessed using Bray-Curtis and Jaccard dissimilarity indices, revealed significant separation between datasets generated by the two sequencing approaches ($p < 0.001$, Figure 3E and 3F).

3.3. Relative abundance and core species of gut microbiome from shallow shotgun metagenomic sequencing and full-Length 16S rDNA amplicon sequencing

At the phylum level, Bacteroidetes predominated in the SSMS method, accounting for 57.60% of the total abundance, whereas it was the second most abundant phylum in the full-length 16S rDNA sequencing method, constituting 30.93%. Conversely, Firmicutes was the most abundant phylum detected by the full-length 16S rDNA method, representing 57.40% of the observed microbiota, and was observed as the second most abundant phylum in the SSMS method, with a relative abundance of 28.93%. Additionally, Proteobacteria exhibited a higher prevalence in the

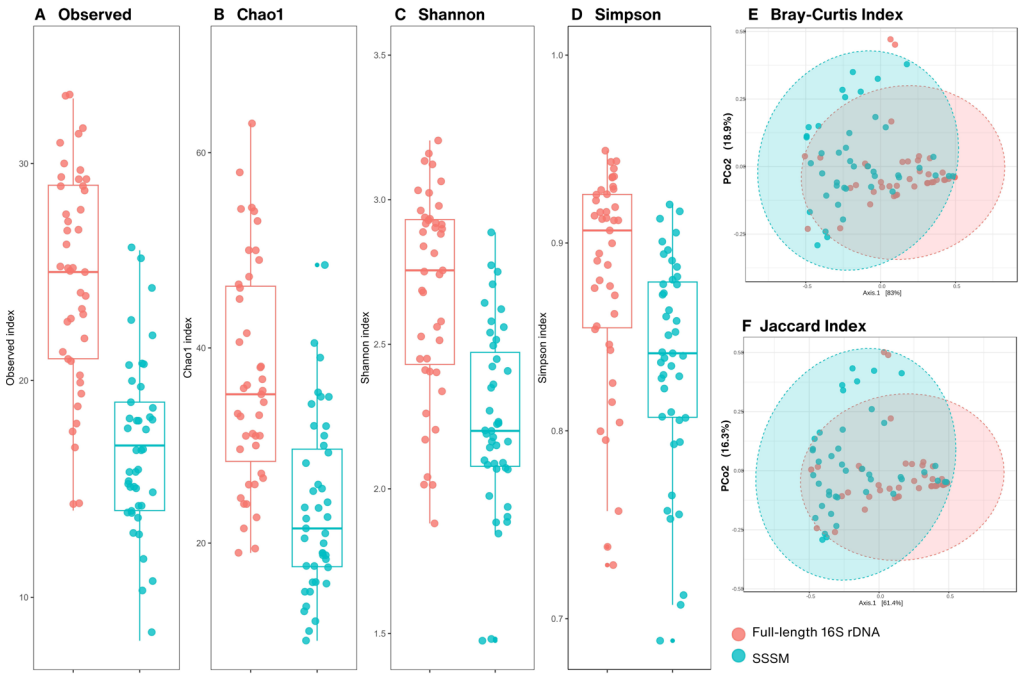


Figure 3. Comparison of gut microbiome diversity measures between sequencing approaches. Alpha diversity is represented by Observed species **(A)**, Chao1 **(B)**, Shannon index **(C)**, and Simpson index **(D)**, with significant differences determined by the Wilcoxon rank-sum test ($p < 0.001$). Beta diversity is visualized using Principal Coordinate Analysis (PCoA) with Bray-Curtis **(E)** and Jaccard **(F)** dissimilarity indices, with statistical significance determined by the PERMANOVA test ($p < 0.001$).

SSMS dataset, with a relative abundance of 7.17%, compared to 5.62% in the full-length 16S rDNA dataset (Supplemental Table S2, *https://www.biosciencetrends.com/supplementaldata/249*). The phyla Actinobacteria, Fusobacteria, Lentisphaerae, and Verrucomicrobia displayed low abundance in both methodologies (Figure 4A). However, the statistical analysis using the Wilcoxon signed-rank test showed that the total abundance at the phylum level is not significantly different between the two approaches ($p = 0.974$).

At the genus level, *Bacteroides* emerged as the most abundant genus within both datasets, although its presence was significantly greater in the SSMS dataset (47.18%) compared to the full-length 16S rDNA method (Figure 4C). Conversely, *Faecalibacterium* ranked as the second most abundant genus in the SSMS dataset (10.10%) but demonstrated markedly lower abundance in the full-length 16S rDNA dataset (1.11%). In contrast, *Prevotella* exhibited a high relative abundance of 8.36% in the full-length 16S rDNA dataset, significantly exceeding its presence in the SSMS dataset (1.83%). Other genera, including *Alistipes*, *Escherichia*, *Parabacteroides,* and *Akkermansia*, also showed higher relative abundances in the full-length 16S rDNA dataset compared to SSMS. Meanwhile, *Eubacterium*, *Roseburia*, *Bifidobacterium*, *Prevotella*, *Oscillibacter, Clostridium*, *Blautia*, and *Ruminococcus* were more prevalent in the full-length 16S dataset than in SSMS. Despite these variances at the genus level, there was no significant difference in the overall abundance of bacterial communities between the two methods ($P = 0.443$), as shown in Figure 4B and

Figure 4. The relative abundance of gut microbiota between sequencing approaches is shown across different taxonomic levels, including phylum (A), genus (B), and species (C). The percentages of relative abundance at each level are displayed for individual samples and group averages for both full-length 16S rRNA and SSMS sequencing approaches.

Supplemental Table S2 (*https://www.biosciencetrends.com/supplementaldata/249*).

Notably, at the species level, *Faecalibacterium prausnitzii* maintained a consistent dominance in both the SSMS (11.07%) and full-length 16S rDNA methodologies (9.94%). In stark contrast, *Bacteroides vulgatus* was significantly more dominant in the SSMS dataset, with 15.71%, compared to a considerably lower prevalence of 4.66% in the full-length 16S rDNA dataset. Additionally, a suite of species within the *Parabacteroides* and *Bacteroides* genera exhibited higher abundances solely in the full-length 16S rDNA dataset, including *Parabacteroides distasonis, Bacteroides distasonis, Bacteroides dorei, Bacteroides uniformis, Bacteroides fragilis, Bacteroides thetaiotaomicron, Bacteroides xylanisolvens, Bacteroides caccae, and Bacteroides ovatus. Conversely, Prevotella copri and Oscillibacter valericigenes* showed a notably higher prevalence in the SSMS dataset, with relative abundances of 6.90% and 5.16%, respectively. These data reveal disparities at the species level, indicating a statistically significant difference in the total relative abundance of species between the two datasets, with an extremely low *P*-value ($P = 2.27e{-}13$) as shown in Figure 4C and Supplemental Table S2 (*https://www.biosciencetrends.com/supplementaldata/249*).

3.4. Differential detection of microbes using SSMS and full-length 16S rDNA sequencing

The LEfSE analysis revealed marked significant differences ($p < 0.01$, LDA > 5, LDA < -5) across 18 bacterial species, as depicted in Figure 4A. The nine species—*Escherichia coli, Bacteroides ovatus, Bacteroides caccae, Parabacteroides distasonis, Bacteroides thetaiotaomicron, Bacteroides fragilis, Bacteroides dorei, Bacteroides uniformis,* and *Bacteroides vulgatus* exhibited elevated detection rates when analyzed using SSMS, as shown in Figure 5B. In contrast, an equivalent number of species, including *Oscillibacter valericigenes, Bacteroides plebeius, Lachnospira pectinoschiza, Blautia obeum, Gemmiger formicilis, Ruminococcus torques, Bacteroides massiliensis, Bacteroides stercoris,* and *Megamonas rupellensis*, demonstrated 50 to 0 percent relative abundance when sequenced using the full-length 16S rDNA approach but were not detectable *via* SSMS, as illustrated in Figure 5C. These findings underscore significant disparities in the detection of relative abundances of gut microbiota attributable to the two sequencing methodologies employed.

4. Discussion

To comprehensively evaluate the bacteria taxa detection from pair samples using two methods: 16S full-length rDNA sequencing from oxford nanopore technology with classification by the RDP database and shallow shotgun metagenomic sequencing by ion torrent with classification by Kraken2. Although the SSMS was applied in this study, the number of bacterial reads was identified as 1,167,404 reads (the average of 43 samples), consistent with previous reports that SSMS data were assigned accuracy taxonomic in species levels (*40*). Despite the similar time and cost requirements of the two approaches, there is a substantial difference in their data output, with one method yielding approximately 839 times more data than the other. This divergence is primarily attributed to the specific gene amplification and random sequencing of all nucleotides, which results in a low mapping ratio for the SSMS approach, recorded at only 11.91%. Consequently, the SSMS method required a significantly higher number of reads compared to the full-length 16S rDNA sequencing approach, which achieved a much higher mapping percentage of approximately 71.91%. These findings align with previous studies using V4 region 16S rDNA sequencing, which reported mapping percentages of 94.4% (*41*).

Numerous studies have compared V4 or V3-V4 16S rDNA sequencing with shotgun and shallow shotgun metagenomic sequencing, with most findings suggesting that SSMS is more effective for identifying bacterial species than V4 or V3-V4 16S rDNA sequencing (*25,28,42*). However, our results using full-length 16S rDNA sequencing indicate the opposite. Our findings show that 29.5% of the identified species were detected exclusively by the full-length 16S rDNA method, while only 16.5% were identified solely by the SSMS method. Although without applying a cutoff to exclude species with a relative abundance lower than 1%, SSMS identified a larger number of species (2,613 species). Many of these identifications were at very low abundance, suggesting that they may be artifact reads. After applying the 1% cutoff, only 94 species remained. This discrepancy may be due to the different methodological sensitivities and biases inherent in each approach. Full-length 16S rDNA sequencing provides more comprehensive coverage of the rRNA gene, which may lead to more accurate species identification, particularly for low-abundance or rare taxa. In contrast, SSMS, while effective at capturing a broad range of species, may include a higher number of false positives, especially when low-abundance thresholds are not applied. We observed significant differences in the gut microbial profiles between the two approaches, from alpha diversity (richness) to beta diversity, even though the Shannon index showed no significant difference. This suggests that the methods differ in their ability to capture species diversity and community composition. Despite these differences, both methods consistently identified Bacteroidetes and Firmicutes as the predominant phyla, which aligns with previous studies of the gut microbiome in healthy individuals (*43,44*).

The comparative analysis of shallow shotgun

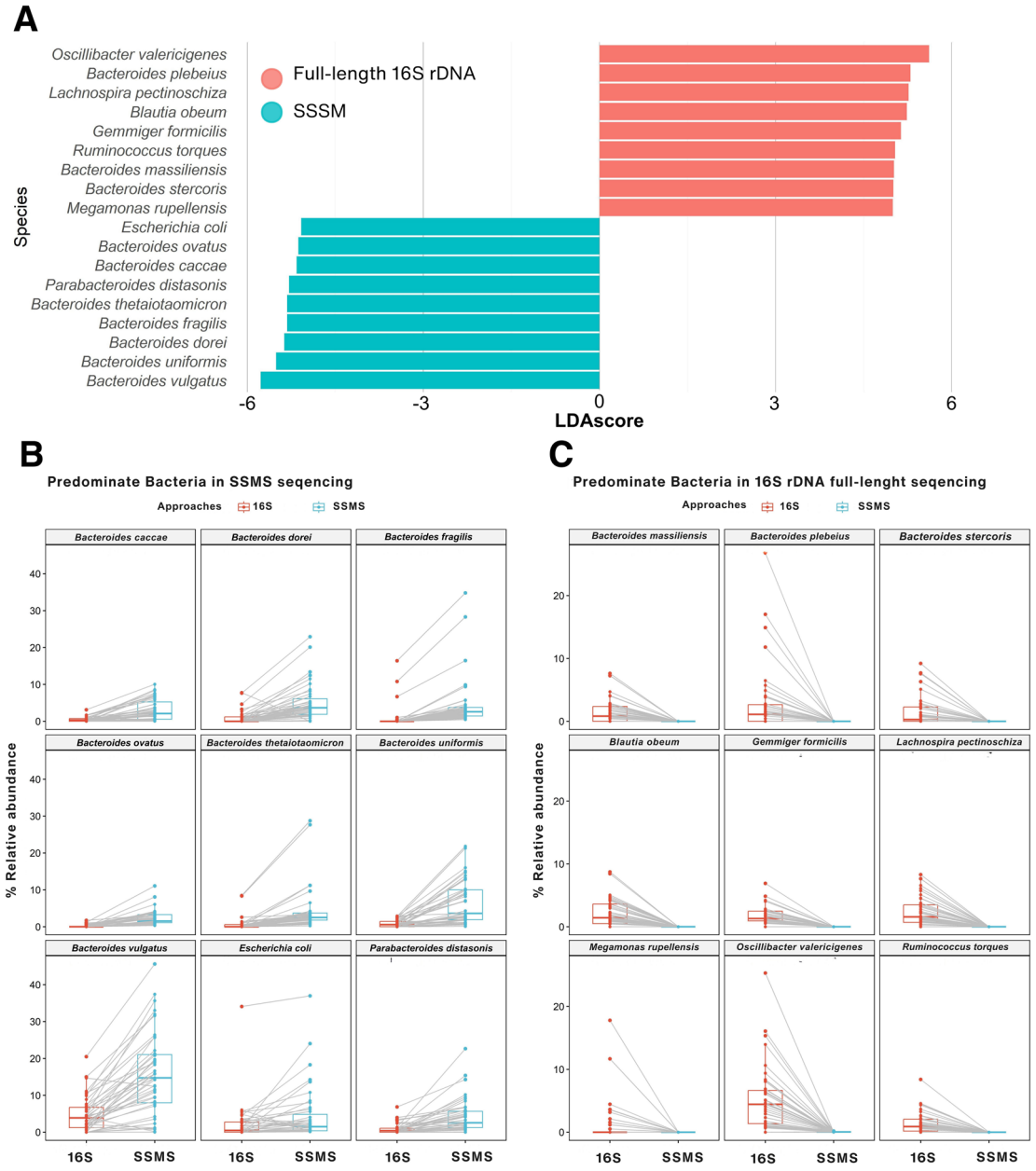


Figure 5. Identification of differentially abundant bacterial species. (A) Linear discriminant analysis (LDA) scores from LEfSe analysis reveal species with differential abundance ($p < 0.01$, LDA > 5 or LDA < -5). SSMS **(B)** and full-length 16S rDNA sequencing **(C)** each show species with significantly higher abundance as detected by their respective approaches, with significant differences determined by the Wilcoxon rank-sum test ($p < 0.001$).

metagenomic sequencing (SSMS) and full-length 16S rDNA sequencing highlights distinct discrepancies in the relative abundances of bacterial genera and species within the gut microbiome. Both methods efficiently capture major microbial groups; however, they demonstrate significant variation in detecting less abundant taxa. At the phylum level, Bacteroides and Faecalibacterium were more prevalent in the SSMS dataset, potentially reflecting the method's increased sensitivity to these groups due to broader genomic coverage and a more extensive database (*45*). This observation aligns with prior studies that indicate SSMS method efficacy in detecting a wide range of taxa, particularly those with greater genomic diversity.

At the species level, *Faecalibacterium prausnitzii* exhibited stable abundance across both methods, underscoring its role as a resilient and central component of the gut microbiome. *F. prausnitzii*, recognized for its high prevalence within the human gut, has been consistently linked to beneficial gut health effects, with decreased levels associated with inflammatory diseases such as Crohn's disease and ulcerative colitis (*13,46-48*). In contrast, notable differences in the abundance of *Bacteroides vulgatus* and other *Bacteroides* species between the SSMS and full-length 16S rDNA datasets suggest that SSMS may either overestimate or capture

strain-level variations not detected by full-length 16S rDNA sequencing (*49*). Additionally, the prominence of *Prevotella copri* and *Oscillibacter valericigenes* in the SSMS dataset suggests that SSMS may better capture specific species; however, this observation could be influenced by low-abundance artifacts(*50*).

The SSMS method identified *P. vulgatus*, a bacterium associated with gastrointestinal diseases such as inflammatory bowel disease (IBD), colorectal cancer, and obesity (*50*). Interestingly, *Prevotella copri*, frequently associated with both beneficial and detrimental health effects, was predominantly detected in the SSMS dataset, while *Oscillibacter valericigenes*, a challenging bacterium to culture linked to bacteremia, showed low abundance in SSMS, suggesting a potential advantage of the full-length 16S rDNA method for profiling low-abundance taxa (*51*). Moreover, *Lachnospira eligens*, also referred to by its basionym *Eubacterium eligens*, was detected by SSMS, while *E. eligens* was primarily identified through full-length 16S rDNA sequencing, illustrating taxonomic discrepancies between the two methods due to database differences. The consistent detection of *Bacteroides dorei* by SSMS, known for promoting the proliferation of gut probiotics, highlights SSMS's potential utility in identifying functionally significant species (*52*).

The taxonomic naming discrepancies observed between databases underscore the critical role of database choice in microbiome research. The RDP database, commonly used for full-length 16S rDNA classification, contrasts with Kraken2, which efficiently processes large datasets from high-throughput sequencing platforms like Illumina and Ion Torrent Torrent (*53,54*). Previous studies affirm that database selection significantly affects the detection and classification of microbiota, further complicating comparisons across sequencing techniques (*45*).

Overall, both full-length 16S rDNA sequencing and shallow shotgun metagenomic sequencing (SSMS) demonstrated time- and cost-efficiency, making them suitable for clinical applications. However, method and database selection significantly impact the detection of low-abundance gut microbiome species, emphasizing the need for careful evaluation. The findings highlight the need for critical evaluation of these methodologies, as each offers unique benefits and limitations regarding microbial diversity and relative abundance resolution. A major strength of this study is the first comparative analysis of full-length 16S rDNA sequencing and SSMS within the same sample set, minimizing inter-sample variability while providing a cost-effective, species-level microbiome characterization. 16S rDNA sequencing offers higher taxonomic resolution, particularly for dominant bacterial species, whereas SSMS captures some broader genomic insights detection the functional genes as antibiotic resistance and virulence factors, making it valuable for infection control. However, SSMS requires higher data and cost compared to 16S rDNA sequencing, which is approximately two times more cost-effective, making it more practical for routine clinical microbiome profiling. Despite these advantages, certain limitations must be acknowledged. The small sample size (n = 43) may impact generalizability, and database-dependent taxonomic biases could influence microbial classification. The observed method-dependent differences suggest that an integrative approach combining SSMS and full-length 16S rDNA sequencing could provide a more comprehensive microbiome profile. To advance microbiome research, standardized classification pipelines are needed to reduce inter-study variability. Expanding sample sizes and diversifying study populations will enhance the robustness and clinical relevance of findings. This approach will enhance considerations for selecting gut microbiome detection methods, facilitating its integration into clinical diagnostics.

5. Conclusion

The comparative study of SSMS and full-length 16S rDNA sequencing highlights the impact of sequencing method and database choice on gut microbiome analysis. Despite comparable time and cost requirements, SSMS yielded significantly more data, primarily due to its broad genomic coverage. However, the full-length 16S rDNA approach offered higher mapping accuracy and identified unique bacterial taxa, particularly at low abundances. Differences in taxonomic classification between RDP and Kraken2 further emphasize the influence of database selection on identification accuracy. Notably, *Bacteroides vulgatus*, *Prevotella copri* and *Oscillibacter valericigenes* exhibited method-dependent detection patterns, underscoring the critical role of methodological choice in microbial analysis. Given these differences, integrating SSMS and full-length 16S rDNA sequencing may provide a more comprehensive relevant representation of gut microbiota. To advance microbiome research and its clinical applications, the development of standardized classification pipelines and expansion of study cohorts with diverse populations are essential. These efforts will enhance the accuracy, consistency, and clinical relevance of microbial community profiling, ultimately deepening our understanding of the gut microbiome's role in health and disease.

Funding: This study was supported by a grant from the by the Fundamental Fund which has received funding support from the National Science Research and Innovation Fund, Thailand (project number 67A103000113).

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Spyrou N, Vallianou N, Kadillari J, Dalamaga M. The interplay of obesity, gut microbiome and diet in the immune check point inhibitors therapy era. Seminars in Cancer Biology. 2021; 73:356-376.
2. Hu X, Ouyang S, Xie Y, Gong Z, Du J. Characterizing the gut microbiota in patients with chronic kidney disease. Postgraduate Medicine. 2020; 132:495-505.
3. Chen F, Dai X, Zhou CC, Li KX, Zhang YJ, Lou XY, Zhu YM, Sun YL, Peng BX, Cui W. Integrated analysis of the faecal metagenome and serum metabolome reveals the role of gut microbiome-associated metabolites in the detection of colorectal cancer and adenoma. Gut. 2022; 71:1315-1325.
4. Onywera H, Meiring TL. Comparative analyses of Ion Torrent V4 and Illumina V3-V4 16S rRNA gene metabarcoding methods for characterization of cervical microbiota: taxonomic and functional profiling. Scientific African. 2020; 7:e00278.
5. Valdes AM, Walter J, Segal E, Spector TD. Role of the gut microbiota in nutrition and health. Bmj. 2018; 361.
6. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. Cell research. 2020; 30:492-506.
7. Cani PD. Human gut microbiome: hopes, threats and promises. Gut. 2018; 67:1716-1725.
8. Sadeghpour Heravi F. Gut Microbiota and Autoimmune Diseases: Mechanisms, Treatment, Challenges, and Future Recommendations. Current Clinical Microbiology Reports. 2024; 11:18-33.
9. Wang X, Yuan W, Yang C, Wang Z, Zhang J, Xu D, Sun X, Sun W. Emerging role of gut microbiota in autoimmune diseases. Frontiers in Immunology. 2024; 15.
10. Miyauchi E, Shimokawa C, Steimle A, Desai MS, Ohno H. The impact of the gut microbiome on extra-intestinal autoimmune diseases. Nature Reviews Immunology. 2023; 23:9-23.
11. Shaheen WA, Quraishi MN, Iqbal TH. Gut microbiome and autoimmune disorders. Clinical and Experimental Immunology. 2022; 209:161-174.
12. Turroni F, Taverniti V, Ruas-Madiedo P, Duranti S, Guglielmetti S, Lugli GA, Gioiosa L, Palanza P, Margolles A, Sinderen Dv, Ventura M. Bifidobacterium bifidum PRL2010 Modulates the Host Innate Immune Response. Applied and Environmental Microbiology. 2014; 80:730-740.
13. Lopez-Siles M, Duncan SH, Garcia-Gil LJ, Martinez-Medina M. Faecalibacterium prausnitzii: from microbiology to diagnostics and prognostics. The ISME Journal. 2017; 11:841-852.
14. Huang Y, Wang Y, Huang X, Yu X. Unveiling the overlooked fungi: the vital of gut fungi in inflammatory bowel disease and colorectal cancer. Gut Pathogens. 2024; 16:59.
15. Bars-Cortina D, Ramon E, Rius-Sansalvador B, Guinó E, Garcia-Serrano A, Mach N, Khannous-Lleiffe O, Saus E, Gabaldón T, Ibáñez-Sanz G. Comparison between 16S rRNA and shotgun sequencing in colorectal cancer, advanced colorectal lesions, and healthy human gut microbiota. BMC genomics. 2024; 25:730.
16. Tessler M, Neumann JS, Afshinnekoo E, Pineda M, Hersch R, Velho LFM, Segovia BT, Lansac-Toha FA, Lemke M, DeSalle R. Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. Scientific reports. 2017; 7:6589.
17. Abellan-Schneyder I, Matchado MS, Reitmeier S, Sommer A, Sewald Z, Baumbach J, List M, Neuhaus K. Primer, pipelines, parameters: issues in 16S rRNA gene sequencing. Msphere. 2021; 6:10.1128/msphere. 01202-01220.
18. Olivier SA, Bull MK, Strube ML, Murphy R, Ross T, Bowman JP, Chapman B. Long-read MinION™ sequencing of 16S and 16S-ITS-23S rRNA genes provides species-level resolution of Lactobacillaceae in mixed communities. Frontiers in Microbiology. 2023; 14:1290756.
19. Qing W, Shi Y, Chen R, Zou Ya, Qi C, Zhang Y, Zhou Z, Li S, Hou Y, Zhou H. Species-level resolution for the vaginal microbiota with short amplicons. Msystems. 2024; 9:e01039-01023.
20. Zhou Y, Liu M, Yang J. Recovering metagenome-assembled genomes from shotgun metagenomic sequencing data: methods, applications, challenges, and opportunities. Microbiological research. 2022; 260:127023.
21. Garrido-Sanz L, Àngel Senar M, Piñol J. Drastic reduction of false positive species in samples of insects by intersecting the default output of two popular metagenomic classifiers. PLoS One. 2022; 17:e0275790.
22. Dommann J, Kerbl-Knapp J, Albertos Torres D, Egli A, Keiser J, Schneeberger PH. A novel barcoded nanopore sequencing workflow of high-quality, full-length bacterial 16S amplicons for taxonomic annotation of bacterial isolates and complex microbial communities. Msystems. 2024; e00859-00824.
23. Curry KD, Wang Q, Nute MG, Tyshaieva A, Reeves E, Soriano S, Wu Q, Graeber E, Finzer P, Mendling W. Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. Nature methods. 2022; 19:845-853.
24. Zhang T, Li H, Ma S, Cao J, Liao H, Huang Q, Chen W. The newest Oxford Nanopore R10. 4.1 full-length 16S rRNA sequencing enables the accurate resolution of species-level microbial community profiling. Applied and Environmental Microbiology. 2023; 89:e00605-00623.
25. La Reau AJ, Strom NB, Filvaroff E, Mavrommatis K, Ward TL, Knights D. Shallow shotgun sequencing reduces technical variation in microbiome analysis. Scientific Reports. 2023; 13:7668.
26. Hillmann B, Al-Ghalith Gabriel A, Shields-Cutler Robin R, Zhu Q, Gohl Daryl M, Beckman Kenneth B, Knight R, Knights D. Evaluating the Information Content of Shallow Shotgun Metagenomics. mSystems. 2018; 3:10.1128/msystems.00069-00018.
27. Marine RL, Magaña LC, Castro CJ, Zhao K, Montmayeur AM, Schmidt A, Diez-Valcarce M, Ng TFF, Vinjé J, Burns CC. Comparison of Illumina MiSeq and the Ion Torrent PGM and S5 platforms for whole-genome sequencing of picornaviruses and caliciviruses. Journal of virological methods. 2020; 280:113865.
28. Durazzi F, Sala C, Castellani G, Manfreda G, Remondini D, De Cesare A. Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. Scientific Reports. 2021; 11:3030.
29. Rubiola S, Macori G, Civera T, Fanning S, Mitchell M, Chiesa F. Comparison Between Full-Length 16S rRNA Metabarcoding and Whole Metagenome Sequencing Suggests the Use of Either Is Suitable for Large-Scale

Microbiome Studies. Foodborne Pathogens and Disease. 2022; 19:495-504.
30. Jitvaropas R, Mayuramart O, Sawaswong V, Kaewsapsak P, Payungporn S. Classification of salivary bacteriome in asymptomatic COVID-19 cases based on long-read nanopore sequencing. Exp Biol Med (Maywood). 2022; 247:1937-1946.
31. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biology. 2019; 20:129.
32. Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B. MinIONQC: fast and simple quality control for MinION sequencing data. Bioinformatics. 2018; 35:523-525.
33. Rodríguez-Pérez H, Ciuffreda L, Flores C. NanoCLUST: a species-level analysis of 16S rRNA nanopore sequencing data. Bioinformatics. 2020; 37:1600-1601.
34. Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA, Chandra S, McGarrell DM, Schmidt TM, Garrity GM, Tiedje JM. The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucleic Acids Research. 2003; 31:442-443.
35. Bolyen E, Rideout JR, Dillon MR, *et al*. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology. 2019; 37:852-857.
36. Garcia BJ, Simha R, Garvin M, Furches A, Jones P, Gazolla JGFM, Hyatt PD, Schadt CW, Pelletier D, Jacobson D. A k-mer based approach for classifying viruses without taxonomy identifies viral associations in human autism and plant microbiomes. Computational and Structural Biotechnology Journal. 2021; 19:5911-5919.
37. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. PeerJ Computer Science. 2017; 3:e104.
38. de Cena JA, Zhang J, Deng D, Damé-Teixeira N, Do T. Low-Abundant Microorganisms: The Human Microbiome's Dark Matter, a Scoping Review. Front Cell Infect Microbiol. 2021; 11:689197.
39. Chong J, Liu P, Zhou G, Xia J. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. Nature Protocols. 2020; 15:799-821.
40. Xu W, Chen T, Pei Y, Guo H, Li Z, Yang Y, Zhang F, Yu J, Li X, Yang Y. Characterization of shallow whole-metagenome shotgun sequencing as a high-accuracy and low-cost method by complicated mock microbiomes. Frontiers in Microbiology. 2021; 12:678319.
41. Usyk M, Peters BA, Karthikeyan S, *et al*. Comprehensive evaluation of shotgun metagenomics, amplicon sequencing, and harmonization of these platforms for epidemiological studies. Cell Reports Methods. 2023; 3.
42. Lamoureux C, Surgers L, Fihman V, Gricourt G, Demontant V, Trawinski E, N'debi M, Gomart C, Royer G, Launay N. Prospective comparison between shotgun metagenomics and sanger sequencing of the 16S rRNA gene for the etiological diagnosis of infections. Frontiers in Microbiology. 2022; 13:761873.
43. Vaiserman A, Romanenko M, Piven L, Moseiko V, Lushchak O, Kryzhanovska N, Guryanov V, Koliada A. Differences in the gut Firmicutes to Bacteroidetes ratio across age groups in healthy Ukrainian population. BMC microbiology. 2020; 20:1-8.
44. Singh P, Rawat A, Alwakeel M, Sharif E, Al Khodor S. The potential role of vitamin D supplementation as a gut microbiota modifier in healthy individuals. Scientific reports. 2020; 10:1-14.
45. Nearing JT, Douglas GM, Hayes MG, MacDonald J, Desai DK, Allward N, Jones CMA, Wright RJ, Dhanani AS, Comeau AM, Langille MGI. Microbiome differential abundance methods produce different results across 38 datasets. Nature Communications. 2022; 13:342.
46. Maioli TU, Borras-Nogues E, Torres L, Barbosa SC, Martins VD, Langella P, Azevedo VA, Chatel J-M. Possible Benefits of Faecalibacterium prausnitzii for Obesity-Associated Gut Disorders. Frontiers in Pharmacology. 2021; 12.
47. Fujimoto T, Imaeda H, Takahashi K, Kasumi E, Bamba S, Fujiyama Y, Andoh A. Decreased abundance of F aecalibacterium prausnitzii in the gut microbiota of C rohn's disease. Journal of gastroenterology and hepatology. 2013; 28:613-619.
48. Lopez-Siles M, Enrich-Capó N, Aldeguer X, Sabat-Mir M, Duncan SH, Garcia-Gil LJ, Martinez-Medina M. Alterations in the abundance and co-occurrence of Akkermansia muciniphila and Faecalibacterium prausnitzii in the colonic mucosa of inflammatory bowel disease subjects. Frontiers in cellular and infection microbiology. 2018; 8:281.
49. Bank NC, Singh V, McCourt B, Burberry A, Roberts KD, Grubb B, Rodriguez-Palacios A. Antigenic operon fragmentation and diversification mechanism in Bacteroidota impacts gut metagenomics and pathobionts in Crohn's disease microlesions. Gut Microbes. 2024; 16:2350150.
50. Verbrugghe P, Van Aken O, Hållenius F, Nilsson A. Development of a real-time quantitative PCR method for detection and quantification of Prevotella copri. BMC Microbiology. 2021; 21:23.
51. Sydenham TV, Arpi M, Klein K, Justesen US. Four Cases of Bacteremia Caused by Oscillibacter ruminantium, a Newly Described Species. Journal of Clinical Microbiology. 2014; 52:1304-1307.
52. Tufail MA, Schmitz RA. Exploring the Probiotic Potential of Bacteroides spp. Within One Health Paradigm. Probiotics and Antimicrobial Proteins. 2024.
53. Lu J, Salzberg SL. Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. Microbiome. 2020; 8:1-11.
54. Terrón-Camero LC, Gordillo-González F, Salas-Espejo E, Andrés-León E. Comparison of Metagenomics and Metatranscriptomics Tools: A Guide to Making the Right Choice. Genes. 2022; 13:2280.

Received December 11, 2024; Revised March 24, 2025; Accepted March 28, 2025.

**Address correspondence to:*
Sunchai Payungporn, Center of Excellence in Systems Microbiology, Department of Biochemistry, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand.
E-mail: sp.medbiochemcu@gmail.com

Released online in J-STAGE as advance publication April 4, 2025.

BST
BioScience Trends

Guide for Authors

1. Scope of Articles

BioScience Trends (Print ISSN 1881-7815, Online ISSN 1881-7823) is an international peer-reviewed journal. *BioScience Trends* devotes to publishing the latest and most exciting advances in scientific research. Articles cover fields of life science such as biochemistry, molecular biology, clinical research, public health, medical care system, and social science in order to encourage cooperation and exchange among scientists and clinical researchers.

2. Submission Types

Original Articles should be well-documented, novel, and significant to the field as a whole. An Original Article should be arranged into the following sections: Title page, Abstract, Introduction, Materials and Methods, Results, Discussion, Acknowledgments, and References. Original articles should not exceed 5,000 words in length (excluding references) and should be limited to a maximum of 50 references. Articles may contain a maximum of 10 figures and/or tables. Supplementary Data are permitted but should be limited to information that is not essential to the general understanding of the research presented in the main text, such as unaltered blots and source data as well as other file types.

Brief Reports definitively documenting either experimental results or informative clinical observations will be considered for publication in this category. Brief Reports are not intended for publication of incomplete or preliminary findings. Brief Reports should not exceed 3,000 words in length (excluding references) and should be limited to a maximum of 4 figures and/or tables and 30 references. A Brief Report contains the same sections as an Original Article, but the Results and Discussion sections should be combined.

Reviews should present a full and up-to-date account of recent developments within an area of research. Normally, reviews should not exceed 8,000 words in length (excluding references) and should be limited to a maximum of 10 figures and/or tables and 100 references. Mini reviews are also accepted, which should not exceed 4,000 words in length (excluding references) and should be limited to a maximum of 5 figures and/or tables and 50 references.

Policy Forum articles discuss research and policy issues in areas related to life science such as public health, the medical care system, and social science and may address governmental issues at district, national, and international levels of discourse. Policy Forum articles should not exceed 3,000 words in length (excluding references) and should be limited to a maximum of 5 figures and/or tables and 30 references.

Communications are short, timely pieces that spotlight new research findings or policy issues of interest to the field of global health and medical practice that are of immediate importance. Depending on their content, Communications will be published as "Comments" or "Correspondence". Communications should not exceed 1,500 words in length (excluding references) and should be limited to a maximum of 2 figures and/or tables and 20 references.

Editorials are short, invited opinion pieces that discuss an issue of immediate importance to the fields of global health, medical practice, and basic science oriented for clinical application. Editorials should not exceed 1,000 words in length (excluding references) and should be limited to a maximum of 10 references. Editorials may contain one figure or table.

News articles should report the latest events in health sciences and medical research from around the world. News should not exceed 500 words in length.

Letters should present considered opinions in response to articles published in *BioScience Trends* in the last 6 months or issues of general interest. Letters should not exceed 800 words in length and may contain a maximum of 10 references. Letters may contain one figure or table.

3. Editorial Policies

For publishing and ethical standards, *BioScience Trends* follows the Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals issued by the International Committee of Medical Journal Editors (ICMJE, *https://icmje.org/recommendations*), and the Principles of Transparency and Best Practice in Scholarly Publishing jointly issued by the Committee on Publication Ethics (COPE, *https://publicationethics.org/resources/guidelines-new/principles-transparency-and-best-practice-scholarly-publishing*), the Directory of Open Access Journals (DOAJ, *https://doaj.org/apply/transparency*), the Open Access Scholarly Publishers Association (OASPA, *https://oaspa.org/principles-of-transparency-and-best-practice-in-scholarly-publishing-4*), and the World Association of Medical Editors (WAME, *https://wame.org/principles-of-transparency-and-best-practice-in-scholarly-publishing*).

BioScience Trends will perform an especially prompt review to encourage innovative work. All original research will be subjected to a rigorous standard of peer review and will be edited by experienced copy editors to the highest standards.

Ethical Approval of Studies and Informed Consent: For all manuscripts reporting data from studies involving human participants or animals, formal review and approval, or formal review and waiver, by an appropriate institutional review board or ethics committee is required and should be described in the Methods section. When your manuscript contains any case details, personal information and/or images of patients or other individuals, authors must obtain appropriate written consent, permission and release in order to comply with all applicable laws and regulations concerning privacy and/or security of personal information. The consent form needs to comply with the relevant legal requirements of your particular jurisdiction, and please do not send signed consent form to *BioScience Trends* to respect your patient's and any other individual's privacy. Please instead describe the information clearly in the Methods (patient consent) section of your manuscript while retaining copies of the signed forms in the event they should be needed. Authors should also state that the study conformed to the provisions of the Declaration of Helsinki (as revised in 2013, *https://wma.net/what-we-do/medical-ethics/declaration-of-helsinki*). When reporting experiments on animals, authors should indicate whether the institutional and national guide for the care and use of laboratory animals was followed.

Reporting Clinical Trials: The ICMJE (*https:// icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html*) defines a clinical trial as any research project that prospectively assigns people or a group of people to an intervention, with or without concurrent comparison or control groups, to study the relationship between a health-related intervention and a health outcome. Registration of clinical trials in a public trial registry at or before the time of first patient enrollment is a condition of consideration for publication in *BioScience Trends*, and the trial registration number will be published at the end of the Abstract. The registry must be independent of for-profit interest and publicly accessible. Reports of trials must conform to CONSORT 2010 guidelines (*https://consort-statement.org/consort-2010*). Articles reporting the results of randomized trials must include the CONSORT flow diagram showing the progress of patients throughout the trial.

Conflict of Interest: All authors are required to disclose any actual or potential conflict of interest including financial interests or relationships with other people or organizations that might raise questions of bias

in the work reported. If no conflict of interest exists for each author, please state "There is no conflict of interest to disclose".

Submission Declaration: When a manuscript is considered for submission to *BioScience Trends*, the authors should confirm that 1) no part of this manuscript is currently under consideration for publication elsewhere; 2) this manuscript does not contain the same information in whole or in part as manuscripts that have been published, accepted, or are under review elsewhere, except in the form of an abstract, a letter to the editor, or part of a published lecture or academic thesis; 3) authorization for publication has been obtained from the authors' employer or institution; and 4) all contributing authors have agreed to submit this manuscript.

Initial Editorial Check: Immediately after submission, the journal's managing editor will perform an initial check of the manuscript. A suitable academic editor will be notified of the submission and invited to check the manuscript and recommend reviewers. Academic editors will check for plagiarism and duplicate publication at this stage. The journal has a formal recusal process in place to help manage potential conflicts of interest of editors. In the event that an editor has a conflict of interest with a submitted manuscript or with the authors, the manuscript, review, and editorial decisions are managed by another designated editor without a conflict of interest related to the manuscript.

Peer Review: *BioScience Trends* operates a single-anonymized review process, which means that reviewers know the names of the authors, but the authors do not know who reviewed their manuscript. All articles are evaluated objectively based on academic content. External peer review of research articles is performed by at least two reviewers, and sometimes the opinions of more reviewers are sought. Peer reviewers are selected based on their expertise and ability to provide quality, constructive, and fair reviews. For research manuscripts, the editors may, in addition, seek the opinion of a statistical reviewer. Every reviewer is expected to evaluate the manuscript in a timely, transparent, and ethical manner, following the COPE guidelines (*https://publicationethics.org/files/cope-ethical-guidelines-peer-reviewers-v2_0.pdf*). We ask authors for sufficient revisions (with a second round of peer review, when necessary) before a final decision is made. Consideration for publication is based on the article's originality, novelty, and scientific soundness, and the appropriateness of its analysis.

Suggested Reviewers: A list of up to 3 reviewers who are qualified to assess the scientific merit of the study is welcomed. Reviewer information including names, affiliations, addresses, and e-mail should be provided at the same time the manuscript is submitted online. Please do not suggest reviewers with known conflicts of interest, including participants or anyone with a stake in the proposed research; anyone from the same institution; former students, advisors, or research collaborators (within the last three years); or close personal contacts. Please note that the Editor-in-Chief may accept one or more of the proposed reviewers or may request a review by other qualified persons.

Language Editing: Manuscripts prepared by authors whose native language is not English should have their work proofread by a native English speaker before submission. If not, this might delay the publication of your manuscript in *BioScience Trends*.

The Editing Support Organization can provide English proofreading, Japanese-English translation, and Chinese-English translation services to authors who want to publish in *BioScience Trends* and need assistance before submitting a manuscript. Authors can visit this organization directly at *https://www.iacmhr.com/iac-eso/support.php?lang=en*. IAC-ESO was established to facilitate manuscript preparation by researchers whose native language is not English and to help edit works intended for international academic journals.

Copyright and Reuse: Before a manuscript is accepted for publication in *BioScience Trends*, authors will be asked to sign a transfer of copyright agreement, which recognizes the common interest that both the journal and author(s) have in the protection of copyright. We accept that some authors (*e.g.*, government employees in some countries) are unable to transfer copyright. A JOURNAL PUBLISHING AGREEMENT (JPA) form will be e-mailed to the authors by the Editorial Office and must be returned by the authors by mail, fax, or as a scan. Only forms with a hand-written signature from the corresponding author are accepted. This copyright will ensure the widest possible dissemination of information. Please note that the manuscript will not proceed to the next step in publication until the JPA Form is received. In addition, if excerpts from other copyrighted works are included, the author(s) must obtain written permission from the copyright owners and credit the source(s) in the article.

4. Cover Letter

The manuscript must be accompanied by a cover letter prepared by the corresponding author on behalf of all authors. The letter should indicate the basic findings of the work and their significance. The letter should also include a statement affirming that all authors concur with the submission and that the material submitted for publication has not been published previously or is not under consideration for publication elsewhere. The cover letter should be submitted in PDF format. For an example of Cover Letter, please visit: *https://www.biosciencetrends.com/downcentre* (Download Centre).

5. Submission Checklist

The Submission Checklist should be submitted when submitting a manuscript through the Online Submission System. Please visit Download Centre (*https://www.biosciencetrends.com/downcentre*) and download the Submission Checklist file. We recommend that authors use this checklist when preparing your manuscript to check that all the necessary information is included in your article (if applicable), especially with regard to Ethics Statements.

6. Manuscript Preparation

Manuscripts are suggested to be prepared in accordance with the "Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals", as presented at *https://www.ICMJE.org*.

Manuscripts should be written in clear, grammatically correct English and submitted as a Microsoft Word file in a single-column format. Manuscripts must be paginated and typed in 12-point Times New Roman font with 24-point line spacing. Please do not embed figures in the text. Abbreviations should be used as little as possible and should be explained at first mention unless the term is a well-known abbreviation (*e.g.* DNA). Single words should not be abbreviated.

Title page: The title page must include 1) the title of the paper (Please note the title should be short, informative, and contain the major key words); 2) full name(s) and affiliation(s) of the author(s), 3) abbreviated names of the author(s), 4) full name, mailing address, telephone/fax numbers, and e-mail address of the corresponding author; 5) author contribution statements to specify the individual contributions of all authors to this manuscript, and 6) conflicts of interest (if you have an actual or potential conflict of interest to disclose, it must be included as a footnote on the title page of the manuscript; if no conflict of interest exists for each author, please state "There is no conflict of interest to disclose").

Abstract: The abstract should briefly state the purpose of the study, methods, main findings, and conclusions. For articles that are Original Articles, Brief Reports, Reviews, or Policy Forum articles, a one-paragraph abstract consisting of no more than 250 words must be included in the manuscript. For Communications, Editorials, News, or Letters, a brief summary of main content in 150 words or fewer should be included in the manuscript. For articles reporting clinical trials, the trial registration number should be stated at the end of the Abstract. Abbreviations must be kept to a minimum and non-standard

abbreviations explained in brackets at first mention. References should be avoided in the abstract. Three to six key words or phrases that do not occur in the title should be included in the Abstract page.

Introduction: The introduction should provide sufficient background information to make the article intelligible to readers in other disciplines and sufficient context clarifying the significance of the experimental findings

Materials/Patients and Methods: The description should be brief but with sufficient detail to enable others to reproduce the experiments. Procedures that have been published previously should not be described in detail but appropriate references should simply be cited. Only new and significant modifications of previously published procedures require complete description. Names of products and manufacturers with their locations (city and state/country) should be given and sources of animals and cell lines should always be indicated. All clinical investigations must have been conducted in accordance Materials/ Patients and Methods.

Results: The description of the experimental results should be succinct but in sufficient detail to allow the experiments to be analyzed and interpreted by an independent reader. If necessary, subheadings may be used for an orderly presentation. All Figures and Tables should be referred to in the text in order, including those in the Supplementary Data.

Discussion: The data should be interpreted concisely without repeating material already presented in the Results section. Speculation is permissible, but it must be well-founded, and discussion of the wider implications of the findings is encouraged. Conclusions derived from the study should be included in this section.

Acknowledgments: All funding sources (including grant identification) should be credited in the Acknowledgments section. Authors should also describe the role of the study sponsor(s), if any, in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication. If the funding source had no such involvement, the authors should so state.

In addition, people who contributed to the work but who do not meet the criteria for authors should be listed along with their contributions.

References: References should be numbered in the order in which they appear in the text. Citing of unpublished results, personal communications, conference abstracts, and theses in the reference list is not recommended but these sources may be mentioned in the text. In the reference list, cite the names of all authors when there are fifteen or fewer authors; if there are sixteen or more authors, list the first three followed by *et al*. Names of journals should be abbreviated in the style used in PubMed. Authors are responsible for the accuracy of the references. The EndNote Style of *BioScience Trends* could be downloaded at **EndNote** (*https://ircabssagroup.com/examples/ BioScience_Trends.ens*).

Examples are given below:

Example 1 (Sample journal reference):

Inagaki Y, Tang W, Zhang L, Du GH, Xu WF, Kokudo N. Novel aminopeptidase N (APN/CD13) inhibitor 24F can suppress invasion of hepatocellular carcinoma cells as well as angiogenesis. Biosci Trends. 2010; 4:56-60.

Example 2 (Sample journal reference with more than 15 authors):

Darby S, Hill D, Auvinen A, *et al*. Radon in homes and risk of lung cancer: Collaborative analysis of individual data from 13 European case-control studies. BMJ. 2005; 330:223.

Example 3 (Sample book reference):

Shalev AY. Post-traumatic stress disorder: Diagnosis, history and life course. In: Post-traumatic Stress Disorder, Diagnosis, Management and Treatment (Nutt DJ, Davidson JR, Zohar J, eds.). Martin Dunitz, London, UK, 2000; pp. 1-15.

Example 4 (Sample web page reference):

World Health Organization. The World Health Report 2008 – primary health care: Now more than ever. *http://www.who.int/whr/2008/whr08_ en.pdf* (accessed September 23, 2022).

Tables: All tables should be prepared in Microsoft Word or Excel and should be arranged at the end of the manuscript after the References section. Please note that tables should not in image format. All tables should have a concise title and should be numbered consecutively with Arabic numerals. If necessary, additional information should be given below the table.

Figure Legend: The figure legend should be typed on a separate page of the main manuscript and should include a short title and explanation. The legend should be concise but comprehensive and should be understood without referring to the text. Symbols used in figures must be explained. Any individually labeled figure parts or panels (A, B, *etc*.) should be specifically described by part name within the legend.

Figure Preparation: All figures should be clear and cited in numerical order in the text. Figures must fit a one- or two-column format on the journal page: 8.3 cm (3.3 in.) wide for a single column, 17.3 cm (6.8 in.) wide for a double column; maximum height: 24.0 cm (9.5 in.). Please make sure that the symbols and numbers appeared in the figures should be clear. Please make sure that artwork files are in an acceptable format (TIFF or JPEG) at minimum resolution (600 dpi for illustrations, graphs, and annotated artwork, and 300 dpi for micrographs and photographs). Please provide all figures as separate files. Please note that low-resolution images are one of the leading causes of article resubmission and schedule delays.

Units and Symbols: Units and symbols conforming to the International System of Units (SI) should be used for physicochemical quantities. Solidus notation (*e.g.* mg/kg, mg/mL, mol/mm^2/min) should be used. Please refer to the SI Guide *www.bipm.org/en/si/* for standard units.

Supplemental data: Supplemental data might be useful for supporting and enhancing your scientific research and *BioScience Trends* accepts the submission of these materials which will be only published online alongside the electronic version of your article. Supplemental files (figures, tables, and other text materials) should be prepared according to the above guidelines, numbered in Arabic numerals (*e.g.*, Figure S1, Figure S2, and Table S1, Table S2) and referred to in the text. All figures and tables should have titles and legends. All figure legends, tables and supplemental text materials should be placed at the end of the paper. Please note all of these supplemental data should be provided at the time of initial submission and note that the editors reserve the right to limit the size and length of Supplemental Data.

5. Submission Checklist

The Submission Checklist will be useful during the final checking of a manuscript prior to sending it to *BioScience Trends* for review. Please visit Download Centre and download the Submission Checklist file.

6. Online Submission

Manuscripts should be submitted to *BioScience Trends* online at *https://www.biosciencetrends.com/login*. Receipt of your manuscripts submitted online will be acknowledged by an e-mail from Editorial Office containing a reference number, which should be used in all future communications. If for any reason you are unable to submit a file online, please contact the Editorial Office by e-mail at *office@ biosciencetrends.com*

8. Accepted Manuscripts

Page Charge: Page charges will be levied on all manuscripts accepted for publication in *BioScience Trends* (Original Articles / Brief Reports / Reviews / Policy Forum / Communications: $140 per page for black white pages, $340 per page for color pages; News / Letters: a total cost of $600). Under exceptional circumstances, the author(s) may apply to the editorial office for a waiver of the publication charges by stating the reason in the Cover Letter when the manuscript online.

Misconduct: *BioScience Trends* takes seriously all allegations of potential misconduct and adhere to the ICMJE Guideline (*https://icmje.org/recommendations*) and COPE Guideline (*https://publicationethics.org/files/Code_of_conduct_for_journal_editors.pdf*). In cases of suspected research or publication misconduct, it may be necessary for the Editor or Publisher to contact and share submission details with third parties including authors' institutions and ethics committees. The corrections, retractions, or editorial expressions of concern will be performed in line with above guidelines.

(As of December 2022)

BioScience Trends
Editorial and Head Office
Pearl City Koishikawa 603,
2-4-5 Kasuga, Bunkyo-ku,
Tokyo 112-0003, Japan.
E-mail: office@biosciencetrends.com