ISSN 1881-7815 Online ISSN 1881-7823

BST BioScience Trends

Volume 17, Number 3
June, 2023

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

CONTENTS

Volume 17, Number 3, 2023

Editorial

Review

Original Article

Correspondence

CONTENTS (*Continued*)

Editorial

DOI: 10.5582/bst.2023.01138

New possibilities for medical support systems utilizing artificial intelligence (AI) and data platforms

Kenji Karako[1], Peipei Song[2,3,*], Yu Chen[1,*], Wei Tang[2,3,4]

[1] Department of Human and Engineered Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan;
[2] National Center for Global Health and Medicine, Tokyo, Japan;
[3] National College of Nursing, Japan, Tokyo, Japan;
[4] Hepato-Biliary-Pancreatic Surgery Division, Department of Surgery, The University of Tokyo Hospital, Tokyo, Japan.

SUMMARY In Japan, there is a growing initiative to construct centralized databases and platforms that can aggregate and manage a vast range of medical, health, and caregiving data for research and analysis. Recent advancements in artificial intelligence (AI), particularly in general-purpose models like the Segment Anything model and Chat GPT, promise significant progress towards utilizing such data-rich platforms effectively for healthcare. Traditionally, AI has displayed superior performance by learning specific images or languages, but now it is advancing towards creating models capable of learning universal traits from images and languages by training on extensive datasets. The challenge lies in the fact that these general-purpose models are trained on data that does not sufficiently incorporate medical information, making their direct application to healthcare difficult. However, the introduction of data platforms can potentially solve this problem. This would lead to the development of universally applicable models to process medical images and AI assistants that can support both doctors and patients. These medical AI assistants can function as a "sub-doctor" with extensive knowledge, assisting in comprehensive analysis of symptoms, early detection of rare diseases, and more. They can also serve as an intermediary between the doctor and the patient, helping to simplify consultations and enhance patient understanding of medical conditions and treatments. By bridging this gap, the AI assistant can help to reduce doctors' workload, improve the quality of healthcare, and facilitate early detection and prevention in the elderly population.

Keywords artificial intelligence, healthcare, data platform, medical assistant

In recent years, Japan has been promoting "Data-based Health Management Initiatives" in preparation for a super-aged society (*1*). Japan has a national health insurance system, and in order to maintain that system and provide quality medical services to the elderly in an aging society with a declining birthrate, it is promoting the construction and operation of a database platform that can centrally manage not only medical data, which has been separately managed by various medical facilities, but also beneficial information for public health such as nursing care and health information. On this platform, citizens can access their own medical information, health check-ups, and prescription drug information. In addition, with the patient's consent, each medical facility can access the patient's past medical information, electronic medical records, and images from magnetic resonance imaging (MRI) or X-ray examinations, allowing medical facilities to better cooperate with each other. This centralized electronic health record database is similar to Australia's "My Health Record" and the United Kingdom's "NHS Digital" (*2,3*). Japan's data platform focuses on the management of personal healthcare records (PHRs) beyond medical care. The platform will also enable integration with external services for PHRs. Another major point is that the collected data can be used for analysis and research after anonymizing personal information. This is expected to lead to prevention, measures to combat lifestyle-related diseases, and the development of new treatments.

Generic image segmentation models specific to each medical imaging device

Deep learning technology, a form of artificial intelligence (AI), has made significant advances in recent years and is expected to be used in various fields. Deep learning models using convolutional neural

networks (CNNs) specifically for image processing have emerged in the field of image recognition in particular, and high-performance models such as VGGNet (*4*) and ResNet (*5*) have been proposed. These models extract visual patterns of objects from images and predict the type of object. In addition, object detection and segmentation models that mask the area of an object in an image, such as Mask R-CNN (*6*), YOLO (*7*), and U-Net (*8*), have been proposed. These models are used not only to detect typical objects like animals, faces, and human bodies, but they are also used in research to detect tumors and nodules on images from computed tomography (CT) scans (*9*), MRI scans (*10*), X-rays (*11*), and ultrasonography (*12*). These studies anticipate that AI models can help prevent doctors from missing diagnoses, improving the accuracy of diagnoses. AI can serve as an excellent assistant for physicians in diagnostic medical imaging. However, creating such image diagnostic support models requires a vast amount of applicable images from imaging studies, and a model needs to be created individually for a given disease. This requires significant cost and data collection, which is a challenge for AI technology.

Amid these issues, Meta has released a model for segmentation called the Segment Anything model (SAM) (*13*). SAM is a model that can generically detect objects by learning about various objects contained in about 11 million vast images. Therefore, it can detect objects from various images without additional learning. The conventional approach has been to collect specific images and focus on detecting specific objects, but the creation of a general-purpose model using a large amount of data represents a new approach. Attempts are being made to apply this SAM to medical image data (*14,15*). However, SAM was trained based on images obtained in general situations, so its use as it is has been difficult. Nonetheless, SAM could be a promising model for use in medicine. SAM could be used as a foundation and further trained with diagnostic medical images obtained from specific examination devices to create a general-purpose detection model specifically for images obtained with specific devices. Images obtained from CT scans, MRI scans, X-rays, and ultrasonography each have unique features that differ significantly from images taken with normal cameras, so a general model could be created as the foundation for detecting various diseases without specializing in a specific disease. However, a significant issue in creating a general-purpose medical model is the collection of vast amounts of data. Centralized medical databases can be effectively utilized to solve this problem. Images obtained from regular health check-ups and screening tests for the public, including those with diseases, can be collected and managed, and these can be used to create a general-purpose model to detect objects and diseases in medical images. If such a general-purpose model were integrated into a data platform for effective use, then AI would be able to automatically examine diagnostic images from each patient. As a result, the data platform could become not just a place to manage data but also a valuable support service for diagnosis of disease, thereby providing support to doctors.

Medical health assistants for doctors and patients

Recent developments in AI are not limited to the field of image recognition but are also in the field of natural language processing. Models that use a structure called Attention, such as Transformers (*16*), have become the foundation for BART (*17*) and GPT (*18*), leading to the creation of large general-purpose language models. Chat GPT (*19*) in particular, which appeared in 2022, has attracted a lot of attention, and its use in various fields is being researched. These language models acquire their capabilities by collecting articles published on the Internet and learning from a large amount of data. While Chat GPT can carry out natural conversations, it does not always output correct information. Combining such a large-scale language model capable of natural conversation with a medical data platform would enable the creation of an assistant that supports both doctors and patients. On a centralized data platform, information such as patients' symptoms, diagnostic information, and results are managed and accessible. Training on this large amount of medical data would allow the creation of an assistant model that supports doctors in their medical practice. The medical AI assistant can assist both patients and doctors and function as a bridge between patients and doctors.

For doctors, the medical AI assistant can function as a "sub-doctor" with a wide range of knowledge. In medical practice, the medical AI assistant can comprehensively indicate additional examinations that should be performed and symptoms that should be checked based on patients' symptoms, thereby preventing doctors from overlooking them. Moreover, the medical AI assistant has learned a wide range of information, so it can provide information even on rare diseases and diseases in fields where the doctor is not a specialist, allowing the identification of early signs of disease in general clinical settings and suggesting more detailed examinations.

Patients can benefit in various ways by accessing the medical AI assistant from the data platform. For example, they could tell the medical AI assistant about their own physical ailment and find out what the early signs of disease are, what measures they need to take, and whether they need to go to a medical facility. In addition, recording and sharing personal health records such as daily food intake, lifestyle habits, and heart rate with the data platform and the medical AI assistant will reduce the effort of recording data and also enable the medical AI assistant to provide early notification if any abnormalities are found and to suggest lifestyle modifications. This helps in preventing disease and

detecting it early.

Moreover, the medical AI assistant not only supports doctors and patients but also serves as a bridge between the two. For example, the medical AI assistant may be able to substitute for the doctor during a patient interview. Patients can consult the medical AI assistant about their symptoms in advance, and having the medical AI assistant ask them pertinent information would simplify interaction between the doctor and the patient and save time. Conversely, patients can ask the medical AI assistant about a disease diagnosed by a doctor, examinations, and surgery to better understand their condition. This could help to reduce the burden on doctors in a super-aged society with a decreasing working population while providing higher quality care and contributing to early detection and prevention.

In summary, Japan is making progress in building databases and platforms that can collect and manage medical, care, and health data for use in research and data analysis. Such a platform, capable of centrally managing a vast amount of data, can be expected to serve as a venue for developing and effectively utilizing general-purpose models such as SAM and Chat GPT for medical purposes. Thus far, AI has performed well by learning specific images and languages, but models that learn universal features of images and languages are becoming feasible through the use of a vast amount of data for extensive learning. However, these general-purpose models are learning from training data that does not contain sufficient medical data, so their direct use in medicine would be difficult. With the advent of data platforms, however, the hope is that these issues can be resolved and that the creation of a universal model to process medical images and an AI assistant that supports both doctors and patients will become possible.

Funding: None

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

References

1. Ministry of Health, Labour and Welfare, Data-based Health Management Initiatives *https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000148743.html* (Accessed June 1, 2023). (in Japanese)
2. Australian Digital Health Agency, My Health Record, Australian Digital Health Agency *https://www.digitalhealth.gov.au/initiatives-and-programs/my-health-record* (Accessed June 1, 2023).
3. NHS Digital, Home - NHS Digital *https://digital.nhs.uk/* (Accessed June 1, 2023).
4. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556. 2014; https://doi.org/10.48550/arXiv.1409.1556.
5. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. IEEE Conf Comp Vis Pattern Recog. IEEE; 2016; 770-778.
6. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. IEEE Trans Pattern Anal Machine Intel 2017: 386-397.
7. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. IEEE Conf Comp Vis Pattern Recog 2016:7 79-788.
8. Frangi AF, Hornegger J, Navab N, Wells WM. U-net: Convolutional networks for biomedical image segmentation. Med Image Computing Computer-Assisted Interven 2015. Vol 9351. Switzerland: Springer International Publishing AG; 2015: 234-241.
9. Sahoo PK, Mishra S, Panigrahi R, Bhoi AK, Barsocchi P. An improvised deep-learning-based mask R-CNN model for laryngeal cancer detection using CT images. Sensors (Basel). 2022; 22:8834.
10. Jing X, Wielema M, Cornelissen LJ, van Gent M, Iwema WM, Zheng S, Sijens PE, Oudkerk M, Dorrius MD, van Ooijen PMA. Using deep learning to safely exclude lesions with only ultrafast breast MRI to shorten acquisition and reading time. Eur Radiol. 2022; 32:8706-8715.
11. Khan E, Rehman MZU, Ahmed F, Alfouzan FA, Alzahrani NM, Ahmad J. Chest X-ray classification for the detection of COVID-19 using deep learning techniques. Sensors (Basel). 2022; 22:1211.
12. Karako K, Mihara Y, Arita J, Ichida A, Bae SK, Kawaguchi Y, Ishizawa T, Akamatsu N, Kaneko J, Hasegawa K, Chen Y. Automated liver tumor detection in abdominal ultrasonography with a modified faster region-based convolutional neural networks (Faster R-CNN) architecture. Hepatobiliary Surg Nutr. 2022; 11:675-683.
13. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg A, Lo W, Dollár P, Girshick R. Segment Anything. ArXiv abs/2304.02643. 2023; https://doi.org/10.48550/arXiv.2304.02643
14. He S, Bao R, Li J, Stout JN, Bjørnerud A, Grant PE, Ou Y. Computer-vision benchmark Segment-Anything Model (SAM) in medical images: Accuracy in 12 datasets. ArXiv abs/2304.09324, 2023; https://doi.org/10.48550/arXiv.2304.09324
15. Deng R, Cui C, Liu Q, Yao T, Remedios LW, Bao S, Landman BA, Wheless L, Coburn LL, Wilson KT, Wang Y, Zhao S, Fogo AB, Yang H, Tang Y, Huo Y. Segment Anything Model (SAM) for digital pathology: Assess zero-shot segmentation on whole slide imaging. ArXiv abs/2304.04155, 2023; https://doi.org/10.48550/arXiv.2304.04155
16. Vaswani A, Shazeer NM, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. NIPS. *https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf* (Accessed June 1, 2023).
17. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. ArXiv abs/1810.04805. 2019; https://doi.org/10.48550/arXiv.1810.04805
18. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-Training. *https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf* (Accessed June 1, 2023).

19. OpenAI, ChatGPT: Optimizing language models for dialogue. *https://openai.com/blog/chatgpt/* (Accessed June 1, 2023).

Received June 3, 2023; Accepted June 21, 2023.

**Address correspondence to:*
Peipei Song, Center for Clinical Sciences, National Center for Global Health and Medicine, 1-21-1 Toyama, Shinjuku-ku, Tokyo 162-8655, Japan.
E-mail: psong@it.ncgm.go.jp

Yu Chen, Department of Human and Engineered Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 227-8568, Japan.
E-mail: chen@edu.k.u-tokyo.ac.jp

Released online in J-STAGE as advance publication June 26, 2023.

DOI: 10.5582/bst.2023.01148

Editorial

Advances in deep learning: From diagnosis to treatment

Tianqi Huang[1], Longfei Ma[1], Boyu Zhang[2], Hongen Liao[1,*]

[1] Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China;
[2] Research Center for Industries of the Future, and Key Laboratory of 3D Micro/Nano Fabrication and Characterization of Zhejiang Province, School of Engineering, Westlake University, Hangzhou, Zhejiang, China.

SUMMARY Deep learning has brought about a revolution in the field of medical diagnosis and treatment. The use of deep learning in healthcare has grown exponentially in recent years, achieving physician-level accuracy in various diagnostic tasks and supporting applications such as electronic health records and clinical voice assistants. The emergence of medical foundation models, as a new approach to deep learning, has greatly improved the reasoning ability of machines. Characterized by large training datasets, context awareness, and multi-domain applications, medical foundation models can integrate various forms of medical data to provide user-friendly outputs based on a patien's information. Medical foundation models have the potential to integrate current diagnostic and treatment systems, providing the ability to understand multi-modal diagnostic information and real-time reasoning ability in complex surgical scenarios. Future research on foundation model-based deep learning methods will focus more on the collaboration between physicians and machines. On the one hand, developing new deep learning methods will reduce the repetitive labor of physicians and compensate for shortcomings in their diagnostic and treatment capabilities. On the other hand, physicians need to embrace new deep learning technologies, comprehend the principles and technical risks of deep learning methods, and master the procedures for integrating them into clinical practice. Ultimately, the integration of artificial intelligence analysis with human decision-making will facilitate accurate personalized medical care and enhance the efficiency of physicians.

Keywords deep learning, foundation model, integrative diagnosis and treatment, human-machine collaboration

Deep learning (DL) is a novel research direction in the field of machine learning. Deep learning involves inputting raw data into a machine, which then develops the representations required for pattern recognition through multiple network layers. By combining these multilayer neural networks, computers can learn highly complex functions to perform tasks (*1*). DL methods have achieved significant success in various medical imaging tasks (*2*) and medical natural language processing (NLP) tasks (*3*). In medical image analysis, deep learning methods such as R-CNN and U-Net have achieved physician-level accuracy in various diagnostic tasks, including detection of diabetic retinopathy, classification of liver lesions on CT, detection of breast lesions on X-rays, assessment of hepatic diseases on ultrasound, and analysis of the spine on MRI (*4,5*). In these tasks, deep learning technologies can assist physicians with diagnosis by providing second opinions and annotating regions in images. For medical NLP tasks, deep learning methods have been used to support various applications, such as electronic health records and clinical voice assistants. Recurrent neural networks, for instance, can model the time series of structured events cited in patient records to predict future clinical events (*6*). Automatic speech recognition technology can convert patient-doctor conversations into transcribed text records and clinical voice assistants can be developed to accurately record patient visits (*7*).

Technological advancements in medical foundation models

Previous research on deep learning has primarily focused on low-level tasks in image processing and NLP, such as segmentation of medical images, identification of lesion features, classification of disease, and semantic analysis of medical information. Advancements in deep learning technology have enabled these methods to achieve the accuracy level of professional physicians in these low-level tasks. These deep learning methods can effectively assist physicians in their work by reducing the time spent on repetitive labor in these tasks and improving medical efficiency.

In recent years, research on foundation models has brought new breakthroughs. The characteristics of foundation models include large training datasets, context awareness, multi-domain applications, and the ability to perform more advanced tasks by integrating information and performing reasoning (*8*). In the medical field, ChatGPT has approached the passing threshold for the United States Medical Licensing Examination without any specialized training, demonstrating its potential in medical diagnosis and clinical decision support (*9*).

How can medical foundation models benefit treatment?

Medical foundation models possess significant value due to their exceptional reasoning ability, enabling the transition from low-level to high-level tasks. Compared to current deep learning methods, medical foundation models have the potential to perform more diverse and challenging tasks. Current foundation models, such as GPT-4, are now capable of comprehending multiple forms of information, including language and images (*10*). Future research can utilize diverse and extensive datasets to enable medical foundational models to flexibly integrate various forms of medical data, including medical images, electronic health records, genomics, charts, and medical texts. These models can provide user-friendly outputs based on multiple forms of patient information, such as diagnostic suggestions or image annotations. By utilizing these capabilities, medical professionals can make more informed decisions and provide better care to patients.

Within treatment scenarios, deep learning technology is primarily used for surgical planning and decision-making. However, there is a growing interest in automated operations by surgical robots. The autonomous operation of robots during surgery using deep learning presents a challenging yet promising direction. Reinforcement learning methods in particular have shown potential in enabling robots to autonomously perform simple surgical tasks (*11,12*). However, more complex surgical scenarios necessitate more complex reasoning ability so that robot operations can be modified based on intraoperative diagnostic information. Researchers have developed various surgical devices to integrate diagnosis and treatment (*13,14*). Medical foundation model technology can provide such devices with the capabilities to effectively analyze medical information and perform diagnostic decision-making. Integrating different diagnostic modes and various forms of medical information in medical foundation models can ensure reliable, timely, and intelligent diagnostic conclusions, thereby improving the efficiency and quality of treatment.

Collaboration between physicians and machines

Currently, research on foundation models in the medical field is still in its early stages. However, we believe that these models have great potential for future applications. With advancements in technology and computing power, the performance of foundation models, including their understanding of information and task accuracy, will continue to improve. Ultimately, creating a fully autonomous diagnosis and treatment system that only requires patients to lie down on the machine for automatic diagnosis and accurate treatment may be possible. At this stage of research, however, autonomous robot systems that can be used in clinical practice without the involvement of physicians is still a challenge. The primary reasons for this include: 1). the complexity of medical data, which are often unstructured, in multiple forms, and individualized; 2). the non-repeatability and safety requirements of medical procedures; and 3. the diagnosis and treatment of diseases involves social and humanistic factors that machines alone cannot fully incorporate.

Researchers should prioritize enhancing physicians' diagnostic and treatment capabilities and reducing physicians' repetitive labor through deep learning technology, rather than replacing them. The development of deep learning technology should aim to address the inadequacies of human skills by assisting with diagnosis and treatment. At the same time, physicians must embrace artificial intelligence and deep learning technology, as they have with other tools such as medical imaging and computer-assisted surgical navigation over the past few decades. To do so, they must comprehend the principles and technical risks of deep learning methods and they must master the procedures for integrating them into clinical practice. Only then can they adapt to the development of deep learning technologies in this era, maintain their independence amidst technological progress, and promote personalized healthcare. Ultimately, the integration of artificial intelligence analysis with human decision-making will facilitate accurate personalized medical care and enhance the efficiency of physicians.

Funding: None.

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

References

1. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015; 521:436-444.
2. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Med Image Anal. 2017; 42:60-88.
3. Hahn U, Oleynik M. Medical information extraction in the age of deep learning. Yearb Med Inform. 2020; 29:208-220.
4. Chen X, Wang X, Zhang K, Fung KM, Thai TC, Moore K,

Mannel RS, Liu H, Zheng B, Qiu Y. Recent advances and clinical applications of deep learning in medical image analysis. Med Image Anal. 2022; 79:102444.
5. Ma LF, Wang R, He Q, Huang LJ, Wei XY, Lu X, Du YN, Luo JW, Liao HE. Artificial intelligence-based ultrasound imaging technologies for hepatic diseases. iLIVER. 2022; 1:252-264.
6. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. JMLR Workshop Conf Proc. 2016 ; 56:301-318.
7. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE J Biomed Health Inform. 2017; 22:1589-1604.
8. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, Rajpurkar P. Foundation models for generalist medical artificial intelligence. Nature. 2023; 616:259-265.
9. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, Madriaga M, Aggabao R, Diaz-Candido G, Maningo J, Tseng V. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023; 2:e0000198.
10. Bubeck S, Chandrasekaran V, Eldan R, *et al*. Sparks of artificial general intelligence: Early experiments with gpt-4. ArXiv230312712. 2023. https://doi.org/10.48550/arXiv.2303.12712
11. Ning G, Zhang X, Liao H. Autonomic robotic ultrasound imaging system based on reinforcement learning. IEEE Trans Biomed Engineer. 2021; 68: 2787-2797.
12. Saeidi H, Opfermann JD, Kam M, Wei S, Leonard S, Hsieh MH, Kang JU, Krieger A. Autonomous robotic laparoscopic surgery for intestinal anastomosis. Sci Robot. 2022; 7:eabj2908.
13. Liao H. Integrated diagnostic and therapeutic techniques: Toward an intelligent medical system. Computerized Med Imaging Graphics. 2014; 38:421-422.
14. Liao H, Noguchi M, Maruyama T, Muragaki Y, Kobayashi E, Iseki H, Sakuma I. An integrated diagnosis and therapeutic system using intra-operative 5-aminolevulinic-acid-induced fluorescence guided robotic laser ablation for precision neurosurgery. Med Image Anal. 2012; 16:754-66.

Received June 7, 2023; Accepted June 28, 2023.

**Address correspondence to:*
Hongen Liao, Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China.
E-mail: liao@tsinghua.edu.cn

Released online in J-STAGE as advance publication June 30, 2023.

Review

DOI: 10.5582/bst.2023.01086

The application and prospection of augmented reality in hepato-pancreato-biliary surgery

Junlong Dai[1,2], Weili Qi[3], Zhancheng Qiu[3], Chuan Li[3,*]

[1] Anatomy Facility, School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom;
[2] The School of Simulation and Visualisation, The Glasgow School of Art, Glasgow, United Kingdom;
[3] Division of Liver Surgery, Department of General Surgery, West China Hospital, Sichuan University, Chengdu, Sichuan, China.

SUMMARY Augmented Reality (AR) is one of the main forms of Extended Reality (XR) application in surgery. hepato-pancreato-biliary (HPB) surgeons could benefit from AR as an efficient tool for making surgical plans, providing intraoperative navigation, and enhancing surgical skills. The introduction of AR to HPB surgery is less than 30 years but brings profound influence. From the early days of projecting liver models on patients' surfaces for locating a better puncture point to today's assisting surgeons to perform live donor liver transplantation, a series of successful clinical practices have proved that AR can play a constructive role in HPB surgery and has great potential. Thus far, AR has been shown to increase efficiency and safety in surgical resection, and, at the same time, can improve oncological outcomes and reduce surgical risk. Although AR has presented admitted advantages in surgery, AR's application is still immature as an emerging technique and needs more exploration. In this paper, we reviewed the principles of AR and its developing history in HPB surgery, describing its significant practical applications over the past 30 years. Reviewing the past attempts of AR in HPB surgery could make HPB surgeons a better understanding of future surgery and the digital trends in medicine. The routine uses of AR in HPB surgery, as an indication of the operating room entering the new era, is coming soon.

Keywords augmented reality, HPB surgery, medical visualization, artificial intelligence

1. Introduction

Augmented Reality (AR) is a relatively new technical field that promotes the integration of real-world and virtual-world information by simulation on the basis of computer science (*1*). It can superimpose virtual content which mainly are 3-dimensional models in the real world and can be captured by human senses in the whole process, so as to achieve a sensory experience beyond reality. As opposed to showing the content simply, the objects generated by the AR system could obey the physical rules like casting the shadow and the perspective. The real environment and virtual objects coexist in the same space more naturally, which allows users to complete human-computer interaction.

AR is an important branch of Extended Reality (XR). XR is human-oriented and emphasizes Human-Computer Interaction (HCI). Depending on how virtual and reality are combined, XR can be further divided into Augment Reality (AR), Virtual Reality (VR), and Mixed Reality (MR) (*2*). AR is a combination of real-time models and actual occasions, which strengthens interaction and perception and could be seen as a complement to human information acquisition. While VR would create a new three-dimensional space that simulates all human feelings. It has three characteristics (3Rs), namely, Real-time rendering, Real space, and Real interaction. The main difference between AR and VR could be summarized simply as whether an unreal scene is established to replace the real environment. MR is more like a bridge to connect the real and the unreal. But unlike the AR would respect the objective existence, the MR would modify some components of the real scene to emphasize some specific information resource of the real world.

Although AR, VR, and MR are distinct in definition, methodology, and final effect, all these visualization technologies bring significant innovations to medicine. Especially for the surgery, XR has a series of attempts to achieve inspirational consequences. In a large-scale meta-analysis, Zhang J *et al*. included one hundred and sixty-eight studies and summarized the present status of the application of

the XR to surgery (*3*). Among these research, thirty-one studies used XR as a tool for making surgical plans beforehand and had been proven to have a great advantage in identifying anatomical structures in advance. Meantime, forty-nine studies applied AR to surgery to test its safety and effect. The result showed the participation of AR could help the surgeon to avoid unintentional damage. And another interesting figure is that eleven studies found VR could release post-surgery pain reducing their need for analgesics. The rest of the studies focused on surgical skills training and medical education. By and large, AR has a wider range of utilization in the surgery itself compared to VR and MR. So, this literature review will focus on the exploration and attempts of AR in hepato-pancreato-biliary (HPB) surgery.

HPB surgery is one of the most complicated portions of abdominal surgery. Abundant blood supply increased the risk of haemorrhage. Complicated anatomical structures and variations also led to the uncertainty of the surgical treatment. So, the mortality and post-surgery complications rate of HPB surgery were high in the past. However, with the application of lots of new inventions like fluorescence-guided technology, venovenous bypass, intraoperative ultrasonic monitoring, and so forth, HPB surgery has made great strides (*4-6*). New technology allowed surgeons to explore many exclusion zones which were hard to imagine before. With accumulating experience and technological improvement, from living donor liver transplantation to laparoscopic pancreatoduodenectomy, surgical technology has no big challenge for HPB surgeons anymore (*7,8*). But in recent years, the development of HPB surgery seems to hit a bottleneck. The lack of innovation, contentment with the status quo, and the dearth of groundbreaking technology advancements are troubling HPB surgeons (*9*). In a word, the HPB surgery appeals to new elicitation. This is why the academic community of surgery was so excited when AI appeared. Surgeries require high-level observative ability and deciding ability according to actual intra-operation conditions. And AR is more suitable to deal with such an intricate task based on fact. Hence the AR is no doubt to be the linchpin to thrust further progress in surgery (*10*).

2. Technical Path for Augmented Reality

The term Augment Reality was first proposed by staff Tom Caudell and David Mizell of Boeing Co. in 1990 to describe the process of adding virtual elements made by the computer to the real world (*11*). Meantime, they also gave a detailed depiction of the features of AR: fewer rendering elements and higher requirements for registration. But it was not a new invention at that time. In fact, the history of AR is much longer than the proposal of the AR concept. The first AR system could be traced back to 1968. Ivan Sutherland, the Turing Award winner from Harvard University, developed a head-mounted display that was named after the Sword of Damokles and was considered to be the prototype of later AR (*12*). The system used an optical perspective head-mounted display with two trackers, one mechanical and the other ultrasonic. Due to the limited processing power of computers at that time, this system was hung on the ceiling above the user's head and was able to convert simple wireframed images into 3D images by a connecting rod to a helmet.

A typical AR system consists of a virtual scene generation unit, displays, and interactive devices (*13*). Among them, the virtual scene generation unit is responsible for modeling, adjusting, and managing the virtual objects; the display presents the signal after the fusion of virtual and reality; and the interactive devices realize the input and output of sensory signals and environmental control operation signals. The whole working flow is modular and cooperative (Figure 1).

In the whole process, there are three core technologies that are crucial for AR: three-dimensional (3D) registration, fused signal display, and real-time computer-human interaction.

2.1. Three-dimensional registration

The 3D registration technique is the core of AR, which involves accurately placing the rendering of virtual objects or special effects, thereby ensuring mapping consistency (*14*). The mainstream of 3D registration has 3 methods

2.1.1. The method based on computer vision

Computer vision depends on images caught by cameras. The core algorithm is divided into 2 ways: marker tracking and natural feature tracking. These detective points will be recognized and re-located in a new coordinate system so that rendering models could be placed in proper 3D locations and be reflected on the screen. This method produces exquisite scenes and needs less equipment but requires high-performance central processors and complicated computation. Prolonged processing time impedes its dexterity.

2.1.2. The method based on hardware

The method based on hardware collects information from sensors. This method has a better performance in transferring between the actual and virtual coordinate systems owing to the sensors' relatively fixed position. But it is also bedeviled by equipment accuracy and environmental intervention limiting its wider utilization.

2.1.3. Mixed Method

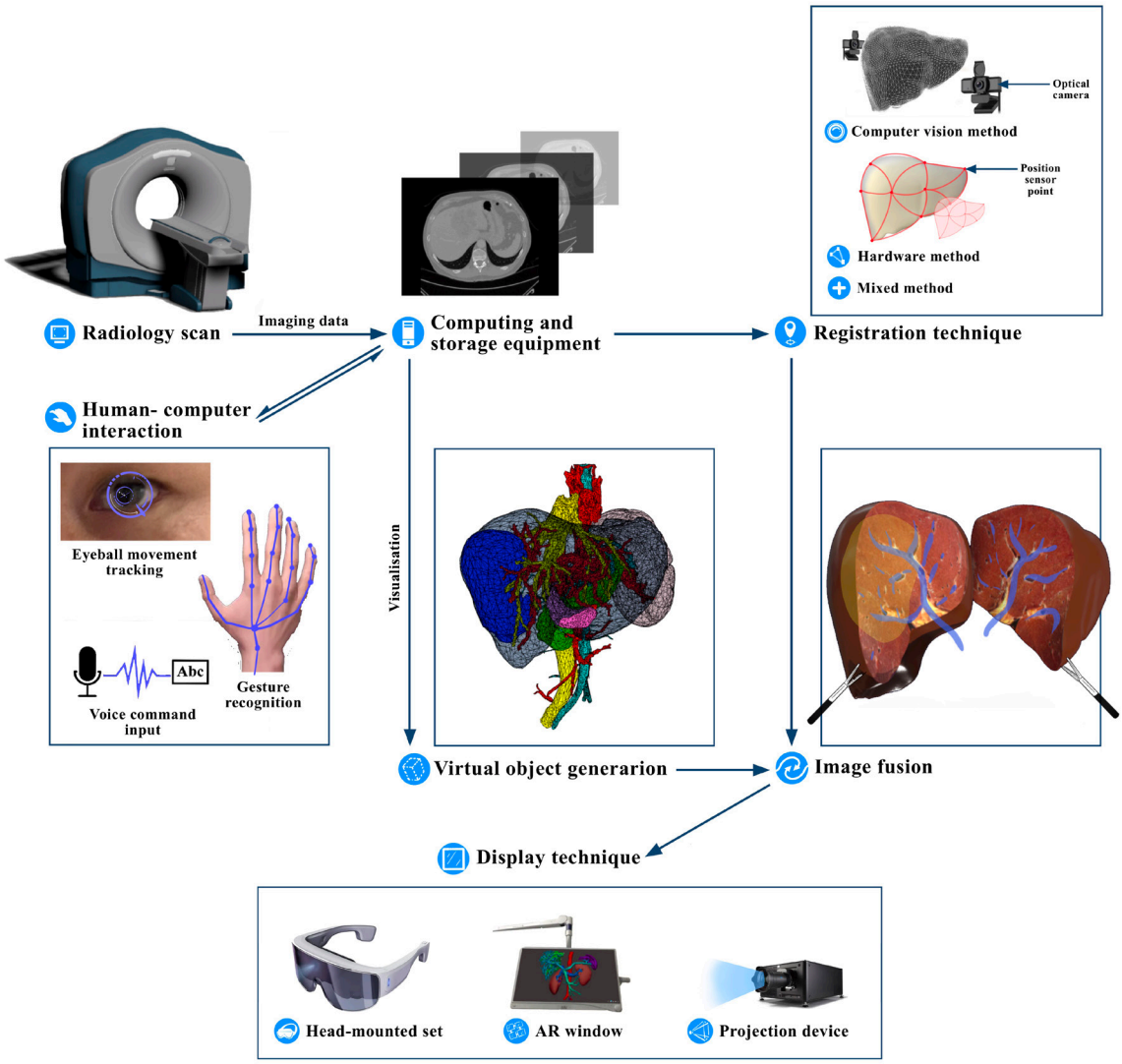

Figure 1. Augmented reality working flow chart. First, the video of the real scene is collected by the camera or sensor (Image input equipment), which is transmitted to the AR's processing unit for analysis (Computing and storage equipment). Then, combined with the data from the tracking device (Tracking technique) realizing the alignment of the coordinate system. The virtual model would appear to proper position, this step is key for the fusion calculation of the virtual and real scene (Image fusion technique). At the same time, the interactive device collects external control signals (User interaction technique). Last, the fused information will be displayed on the monitor in real-time (Display technique), namely the human field of vision.

The combination of both could increase calculating efficiency and scene quality at the same time. Many researchers devoted themselves to this kind of combination. However, good compatibility is challenging to achieve, especially when there is a difference between the camera and sensors. Consequently, it may lead to mistakes or conflicts.

2.2. Display technique

The essence of the display technique is visional representation. After registration, the position of the fused signal has been decided. So, the occlusion relationship between virtual models and real things is clear. But the experience process of AR is successive which means the display device has to be adjusted at a very high frequency according to the eye movement. Every frame needs attention on the light and shadow, distance and perspective to keep the users' optical coherence. According to the location of the equipment, all displays could be classified: the head-mounted-based display (*e.g.*, AR glasses), window-based display (*e.g.*, smartphone, tablet computer), and projection-based display (*15,16*).

2.2.1. The head-mounted-based display

The Head-Mounted Display (HMD) has a good function of information immersion, especially for wireless ones (*17*). It can enhance the user's sensory experience and free the user's hands. Besides, it also allows head movements to be tracked momently for better

registration, which makes it possible to offset delays in signal transmitting and processing. But when moving the head quickly, the stability of the image is hard to be kept. This is intolerant when the AR is dealing with an emergency medical situation and may lead to disastrous results.

2.2.2. The window-based display

The AR window, including the screens of portable and fixed devices, is ubiquitous in modern life. Portable devices like smart mobile phones and tablet computers are easy to be carried with users and could be better at protecting privacy while fixed devices like public screens could realize teamwork based on the same vision. In medicine, a good application is that the tumor model could be projected to the screen of the thoracoscope according to the CT or MRI scan before the surgery, which is beneficial for delineating the area of excision for the thoracic surgical team and reducing the loss of normal tissue (*18*). However, the disadvantages are also obvious: lack of immersed feeling, lower rendering quality, and registration degree.

2.2.3. The projection-based display

Projection device could realize a large-scale scene presentation (*19*). It avoids a single focus; hence, it is more suitable when AR is applied to a large number of users at the same time. But if more interaction is needed, it will give rise to mass calculation weakening the user's experience. In addition, it is also easy to be interfered with by environmental factors like sunlight and noise.

2.3. Human-computer interaction

Human-computer interaction (HCI) is a multidisciplinary field of study focusing on the design of computer technology and, in particular, command input. It is based on reciprocal communication between users and one or several components of computer-generated environments. HCI has two principal features: the user can control the viewpoint in the AR on six degrees of freedom (Navigation) and can interact with objects within the AR (interaction) (*20*).

HCI helps virtual models to be better presented after adjusting by users discretionarily. In some cases, HCI could also do some difficult missions like working as a surgical assistant by using mechanical arms (*21*). From traditional mice and keyboards to motion tracking (*e.g.*, partial/full body, gesture), haptics (*e.g.*, force feedback), gaze tracking, and voice command, interactive technology becomes more variable. These emerging interactive modes fitting for the no-touch principle are particularly valuable for surgeries that require aseptic principles and facilitate the development of AR dramatically in surgical fields (*22*).

3. Application of AR in HPB surgery

Since the millennium, the concept of AR-aiding surgery has been wildly applied to clinical practices by more HPB surgeons, with three main applications emerging (*23*): 1 real-time intraoperative navigation, 2 preoperative simulations, and 3 surgical skills training. In addition, AR could also be used in nursing, anesthesia, and intensive care during the perioperative period (*24-26*). In fact, AR has permeated almost every part of surgery and has been transforming the traditional HPB surgical pattern fundamentally (*27*).

3.1. Real-time intraoperative navigation

The surgical navigation system refers to the organic combination of modern imaging technology, stereotactic technology, artificial intelligence, and surgeons (*28*). It makes adequate use of preoperative scan information and intraoperative findings to enable surgeons to deliver safe, precise, and minimally invasive surgical treatment.

In 1988, Marsescaux *et al*. first introduced the 3D concept to HPB, at that time 3D visualization was adopted to learn the complex liver anatomy and simulate simple liver cancer resection (*29*). Their unprecedented achievement was regarded as a major revolution in surgical practice at that time (*30*). But their attempt was limited since their visualization system only used data from the French National Library of Medicine instead of from real clinical data. One of the pioneers who adopted AR for HPB surgery navigation was Nobuhiko Hata. In 2004, his team developed an AR system called "Projected Augmented Reality" for liver surgery, which could project a 3D model of the patient's liver with a real tumor on the surface of the patient during microwave thermocoagulation, allowing surgeons to see real liver in surgery (*31*). After testing, this AR system's average registration accuracy reached 1.13 mm improving the safety of liver puncture. After that, Stüdeli *et al*. developed another AR system to improve accuracy in needle placement during percutaneous radio-frequency ablation of liver tumors also achieving an ideal result (*32*). These successful attempts made more people see a promising future of surgery aided by AR and promoted further application of AR to more complicated HPB surgery.

Then in 2009, Sugimoto *et al*. projected a virtual cholangiogram on the abdominal wall in laparoscopic cholecystectomy and highlighted hidden bile duct structures improving the surgery safety (*33*). Though this approach was only used in three cases, this was the first time that AR had been used in biliary surgery indicating that AR had officially entered the application stage in HPB surgery. However, such projection-based

AR's disadvantages were obvious: non-real-time, no interaction, and visual challenge for strong colors.

In order to achieve a true intraoperative AR, between 2005-2010, researchers explored and developed many new technologies. Hansen *et al*. presented new methods for intraoperative display of vascular structures in liver surgery reducing the visual complexity of vascular structures, and accentuating spatial relations among main branches (*34*). Shekhar *et al*. developed a live AR navigation system for laparoscopic surgery using continuous low-dose volumetric computed tomography (CT) and tested its stability through experiments upon pigs (*35*). Furthermore, Konishi *et al*. developed another navigation system based on intraoperative ultrasound (IOUS) that achieved AR registration in real-time and avoided extra radiation exposure (*36*). This system could finish scanning liver tumor mimics of pigs in 30 seconds and generate the 3D models in 3 minutes on the screen of the laparoscope. Meantime, Gavaghan *et al*. focused on developing a navigation system for open liver surgery by image overlay projection (*37*). Their extraordinary efforts laid firm ground for the application of AR to the complex HPB surgeries performed on real patients (*38*).

In 2013, Okamoto *et al*. applied AR to perform laparotomy for a patient with benign biliary stricture, a patient with gallbladder carcinoma, and a patient with hepatocellular carcinoma (HCC) (*39*). The operative procedures consisted of choledocho-jejunostomy, right hepatectomy, and microwave coagulation. The site of the tumor, preserved organs and resection aspect overlaid onto the operation field images observed by the monitors enriching the surgeons' perception. In the same year, Marzano *et al*. applied AR to a pancreatico-duodenectomy: the dissection of the superior mesenteric artery, and the hanging maneuver was performed under AR guidance along the hanging plane (*40*). A specific technician manually registered virtual and real images in real time aside. In this 360-minute surgery, AR recognized all the important vascular structures at high precision. The surgeries mentioned above were all the most difficult surgeries in abdominal surgery and AR achieved ideal effects in all of them. After then, AR began to flourish in HPB surgery.

In 2014, Kenngott *et al*. realized real-time image guidance in laparoscopic liver surgery firstly (*41*). After that, Katic *et al*. added human-computer interaction to the AR system for laparoscopic HPB surgery to filter unnecessary information display (*42*). One year later, Pessaux *et al*. first combined AR with robotic surgery to perform hepatic segmentectomy (*43,44*). The same year, Okamoto *et al*. performed pancreatectomy in five cases using AR-based navigation (*45*). Later, whether it is hilar cholangiocarcinoma resection, removal of foreign body in the pancreas, or living donor liver transplantation, AR played an increasingly complex role, and its application was becoming more mature (*46-48*). Up to now, it can be said that HPB has no restricted area for AR anymore (*49*).

3.2. Preoperative simulation

As early as 1998, the concept that AR could assist hepatic and endoscopic surgery was proposed by researchers (*29,50*). Driven by this pioneering concept, in 2003, Bornik *et al*. developed a system for liver surgery planning that enables physicians to visualize and refine segmented input liver data sets, as well as to simulate and evaluate different resections plans (*51*). The system supported surgeons in finding the optimal treatment strategy for each patient and was the first time that AR was applied to make a specific HPB surgical plan. In 2004, Reitinger *et al*. designed an AR-based system to make surgical plans for liver cancer patients (*52*). This system could provide precise position relations between the tumor and portal vein tree. They deemed that measurements based on 2D cross-sectional images were inaccurate while 3D visualization could provide more information enhancing the operational flexibility. Next, Scheuering *et al*. developed a more thorough system which consists of two parts: a preoperative planning tool for liver surgery and an intraoperative real-time visualization component (*51*). The planning tool took into account the individual anatomy of the intrahepatic vessels, determined the vascular territories, and provided methods for fast segmentation of the liver parenchyma, the intrahepatic vessels, and liver lesions. Their practical evaluation had shown a good acceptance of this system for HPB surgeons. Except for Open surgery and laparoscopic surgery, AR was also applied to ablation and interventional operation plan (*53,54*).

AR's comprehensive application in preoperative plans is not limited to make treatment strategy. AR could be also used to correct established surgical plans. Bornik *et al*. developed an AR-based liver segmentation refinement tool that aids doctors to correct inaccurate segmentations efficiently in true 3D using head-mounted displays and tracked input devices. This is of great significance for the delineation of the scope of anatomical hepatectomy because it is non-invasion and provides information beforehand so surgeons could make pointed surgical plans than depending on experience only. In addition, AR was also found to have advantages in detecting abdominal vascular variations which were hard to be reported by CT scan and CT-angiography.

Surgical planning has so far been led by the surgeon, and AR's role has been to provide information and simulation. But as AR advances, finally, surgeons may be from surgical procedure designers to approvers (*55*).

3.3. Surgical skills training

The number of patients who have undergone laparoscopic

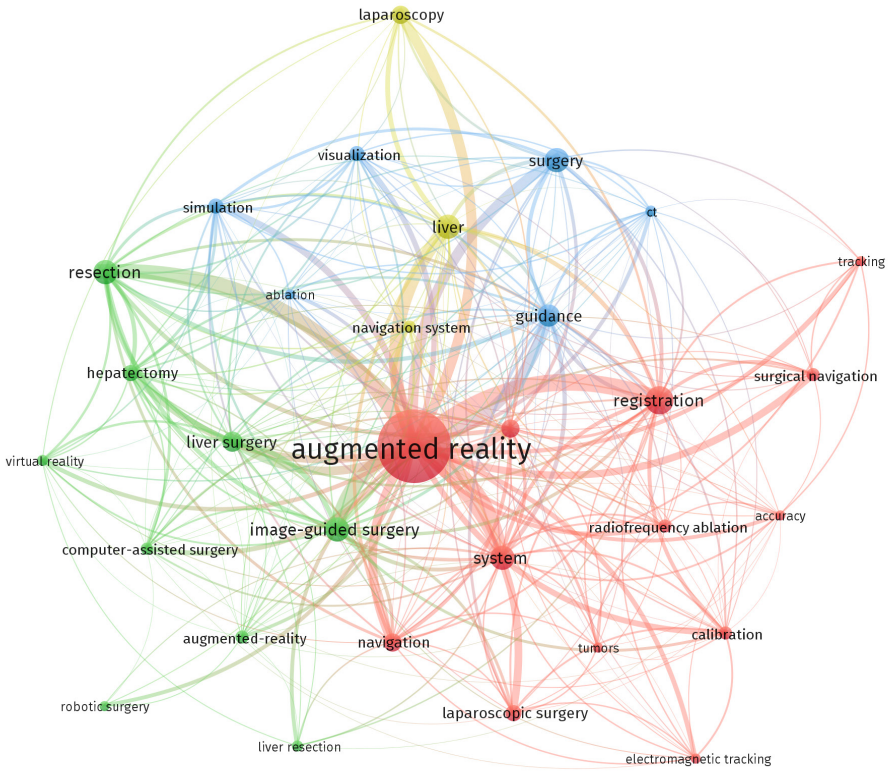

Figure 2. Net map of related literature keywords of AR application in HPB surgery.

HPB surgery has been increasing in the last 20 years. But unlike open surgery, the surgical skills of laparoscope were very hard to be practiced. Lack of sense of reality and detachment from clinical fact have always been two main learning obstacles troubling young HPB surgeons (*56*). After researchers had successfully applied of AR to anatomy education, more effort was made to develop an AR-based surgery simulation system. In 2011, Strickland developed an ex vivo simulated training model for laparoscopic liver resection (*57*). Then in 2015, Nomura *et al*. developed a VR-based training system and AR-based assessing system for laparoscopic cholecystectomy (*58*). The result showed medical students could improve their laparoscopic skills with such a form of simulator.

In addition, AR was also applied to learn complex anatomy of HPB. Viglialoro *et al*. used a well-design AR to teach the concept of Calot's triangle in laparoscopic cholecystectomy and the key points of isolation (*59*). According to the result, they believe that AR was an effective tool to learn organs with invisible vessel structure anatomy. Furthermore, Schott *et al*. appraised the effect of a VR/AR environment on multi-user liver anatomy education (*60*). The result showed that their prototype was usable, induced presence, and potentially supported the teaching of liver anatomy and surgery in the future. Interestingly, the objects who benefit from AR's education could also be the patients. Andolfi *et al*. made a 3D digital cancer model of the head of the pancreas for resident training and then 3D printed the model to carry on patient education for biliary obstruction (*61*). All these attempts expand the application field of AR and are good for cultivating new-generation HPB surgeons.

4. Conclusion and Outlook

AR technology applied in the clinical diagnosis and treatment of auxiliary has a unique advantage. The precision and safety of the AR have been testified widely. AR technology, especially for computer-combined navigation, has become a trend of medical development in the future (*62*). In 2022, the Food and Drug Administration (FDA) approved the first AI-driven AR guidance system called HOLO Portal for spinal surgery, marking the official health agency's recognition of AR-mediated surgery. Unfortunately, so far, AR products specifically suitable for HPB surgery have not yet come out, leaving a huge gap.

In the past 30 years, more than 3,000 pieces of literature about AR's application in surgery have been published, and the top three areas are neurosurgery, liver surgery, and orthopedic surgery accordingly. Obviously, HPB is undoubtedly to be the hottest spot in AR-assisted surgery research. In recent related 300 articles and reports, high-frequency keywords were analyzed (Figure 2), showing that HPB surgeons pay more attention to AR's registration and application in liver, image-guided, laparoscopic surgery.

After proving that AR was safe for HPB surgical use, many researchers began to compare its effect with traditional surgical patterns. In 2017, Diana *et al*. prospectively evaluated the identifying precision of the bile duct using AR-VR navigation and X-ray-based intraoperative cholangiography during robotic cholecystectomy for 58 patients. Ultimately, AR-VR enabled the identification of 12 anatomical variants in 8 patients, of which only 7 could be correctly reported by the radiologists (*63*). This showed that AR is a powerful complement to the surgeon's visual observation. In 2018, Cheung *et al*. compared the surgical effect between laparoscopic hepatectomy guided by AR and Indocyanine Green (ICG) fluorescence imaging and open hepatectomy for HCC (*64*). In 2023, Zhu *et al*. also made a similar effect comparison of hepatectomy for centrally located HCC guided by AR or not (*65*). Although all these studies' results were positive, it's still limited so far. High-level multicenter prospective control experiments are necessary to further evaluate the effectiveness and safety of AR in HPB surgery.

Combined with its own development trends, the hybridization of AR and other disciplines or technologies is also worthy of expecting:

i) Artificial intelligence is the star field of computer science in recent years. A typical example is the Chat Generative Pre-trained Transformer (ChatGPT) dialogue program developed by Open-AI Ltd. In addition to language models, machine learning can also help train AR systems. For the technical barriers that limit the wide application of AR in the field of surgery: action prediction and simulated elastic deformation of organs, machine learning will greatly reduce the threshold for using AR.

ii) 5th Generation Mobile Communication Technology (5G) makes short-term high-throughput information transmission possible. In fact, telemedicine that relies on 5G networks is promoting the sinking of top medical resources. The 5G network can greatly reduce the signal delay caused by huge calculations. In fact, many remote robotic surgeries have been used in medical practice. AR can better allow surgeons who are thousands of miles away to understand the details of the surgery. If it can be combined with more difficult tactile simulation, AR surgery will completely break the geographical restrictions of doctors and benefit more patients.

iii) 3D holographic projection also gives AR more possibilities. Scopis Ltd. uses a head-mounted display called HoloLens to project AR holographic images directly onto patients for surgical navigation. After testing, this 3D technology is also well-compatible with the AR system.

Despite AR has achieved series of successes in surgery, there are still some problems that need to be solved:

i) During the HPB surgery, the organ's shape and inter structure would be changing accordingly as surgeons adjust the organ's position frequently to expose a better view field. Realizing real-time reconstruction and registration of the new model is a huge challenge. It will be very low effective that surgeons have to wait for new scanning and calculating after each operation step.

ii) Visual occlusion is another problem that needs to be addressed. Especially for head-mounted sets, how to deal with optimizing their display layers? Could AR, in some extreme conditions, be a distraction for surgeons to focus on the most important part of the surgical field of vision? In other words, how to make sure the information AR presented is really constructive not a publicity stunt.

iii) Whether the over-involvement of AR will lead to the abuse of technology and make surgeons over-rely on AR slashing their creativity and could not accomplish a complex surgery when AR is not accessible? In extreme cases, the calculation and reconstruction of the model will increase the time required for preoperative preparation. For some patients who need to receive emergency surgery, is it worth spending extra time at the risk of fatal danger doing visualization instead of sending patients into operation theatre as soon as possible?

In general, from aseptic techniques, anesthesia techniques, laparoscopic techniques, and robotic techniques to today's AR, each prosperity in the surgery area have been accompanied by the application of a kind of revolutionary technology (*66*). So, it's reasonable to expect that AR will open a new chapter in the evolution of surgery. Meantime, it also needs to be recognized that, as a technology with great potential, AR is still in its infancy and requires further innovations, improvement, and grinding.

Acknowledgements

We thank Xuyin Di and Lei Zhao for their contribution to the content polishing and thank Matthieu Poyade and Daisy Abbott for their contribution to the expertise in medical visualization.

Funding: None.

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

References

1. Venkatesan M, Mohan H, Ryan JR, Schürch CM, Nolan GP, Frakes DH, Coskun AF. Virtual and augmented reality for biomedical applications. Cell Rep Med. 2021; 2:100348.
2. Shaikh TA, Dar TR, Sofi S. A data-centric artificial intelligent and extended reality technology in smart healthcare systems. Soc Netw Anal Min. 2022; 12:122.
3. Zhang J, Lu V, Khanduja V. The impact of extended

reality on surgery: a scoping review. Int Orthop. 2023; 47:611-621.
4. Lwin TM, Hoffman RM, Bouvet M. Fluorescence-guided hepatobiliary surgery with long and short wavelength fluorophores. Hepatobiliary Surg Nutr. 2020; 9:615-639.
5. Scherer R, Giebler R, Erhard J, Lange R, Günnicker M, Schmutzler M, Paar D, Kox WJ. A new method of veno-venous bypass during human orthotopic liver transplantation. Anaesthesia. 1994; 49:398-402.
6. de Jong KP, Terpstra OT, Blankensteijn JD, Laméris JS. Intraoperative ultrasonography and ultrasonic dissection in liver surgery. Am J Gastroenterol. 1989; 84:933-936.
7. Broering DC, Sterneck M, Rogiers X. Living donor liver transplantation. J Hepatol. 2003; 38 Suppl 1:S119-S135.
8. van Hilst J, de Rooij T, Bosscha K, Brinkman DJ, van Dieren S, Dijkgraaf MG, Gerhards MF, de Hingh IH, Karsten TM, Lips DJ, Luyer MD, Busch OR, Festen S, Besselink MG; Dutch Pancreatic Cancer Group. Laparoscopic versus open pancreatoduodenectomy for pancreatic or periampullary tumours (LEOPARD-2): a multicentre, patient-blinded, randomised controlled phase 2/3 trial. Lancet Gastroenterol Hepatol. 2019; 4:199-207.
9. Alderson D. The future of surgery. Br J Surg. 2019; 106:9-10.
10. Botden SMBI, Jakimowicz JJ. What is going on in augmented reality simulation in laparoscopic surgery? Surg Endosc. 2009; 23:1693-1700.
11. Cipresso P, Giglioli IAC, Raya MA, Riva G. The Past, Present, and Future of Virtual and Augmented Reality Research: A Network and Cluster Analysis of the Literature. Front Psychol. 2018; 9:2086.
12. Werrlich S, Nitsche K, Notni G, Assoc Comp M. Demand Analysis for an Augmented Reality based Assembly Training. In: Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA 17). 2017; pp. 416-422.
13. Ma L, Huang T, Wang J, Liao H. Visualization, registration and tracking techniques for augmented reality guided surgery: a review. Phys Med Biol. 2023; 68. doi: 10.1088/1361-6560/acaf23.
14. Xiong J, Hsiang E-L, He Z, Zhan T, Wu S-T. Augmented reality and virtual reality displays: emerging technologies and future perspectives. Light Sci Appl. 2021; 10:216.
15. Heinrich F, Schwenderling L, Joeres F, Lawonn K, Hansen C. Comparison of Augmented Reality Display Techniques to Support Medical Needle Insertion. IEEE Trans Vis Comput Graph. 2020; 26:3568-3575.
16. O'Reilly MK, Heagerty PJ, Gold LS, Kallmes DF, Jarvik JG. Augmented Reality. AJNR Am J Neuroradiol. 2020; 41:E67-E68.
17. Mojica CMM, Velazco-Garcia JD, Zhao HR, Seimenis I, Leiss EL, Shah D, Webb A, Becker AT, Tsiamyrtzis P, Tsekos NV, Ieee. Interactive and Immersive Image-guided Control of Interventional Manipulators with a Prototype Holographic Interface. In: 19th Annual IEEE International Conference on Bioinformatics and Bioengineering (BIBE) (Athens, GREECE, 2019; pp. 1002-1005.
18. Spijkerboer KGP, Fitski M, Siepel FJ, van de Ven CP, van der Steeg AFW. Augmented reality-guided localization of a chest wall tumor in a pediatric patient. Eur J Cancer. 2022; 170:103-105.
19. Reipschlager P, Flemisch T, Dachselt R. Personal Augmented Reality for Information Visualization on Large Interactive Displays. IEEE Transactions on Visualization and Computer Graphics. 2021; 27:1182-1192.
20. Hasan MS, Yu HN. Innovative Developments in HCI and Future Trends. International Journal of Automation and Computing. 2017; 14:10-20.
21. Innocente C, Ulrich L, Moos S, Vezzetti E. Augmented Reality: Mapping Methods and Tools for Enhancing the Human Role in Healthcare HMI. AAppl. Sci. 2022; 12: 4295; https://doi.org/10.3390/app12094295
22. Lang H, Huber T. Virtual and Augmented Reality in Liver Surgery. Ann Surg. 2020; 271:e8.
23. Cleary K, Peters TM. Image-guided interventions: technology review and clinical applications. Annu Rev Biomed Eng. 2010; 12:119-142.
24. San Martin-Rodriguez L, Soto-Ruiz MN, Echeverria-Ganuza G, Escalada-Hernandez P. Augmented reality for training operating room scrub nurses. Med Educ. 2019; 53:514-515.
25. McKendrick M, Yang S, McLeod GA. The use of artificial intelligence and robotics in regional anaesthesia. Anaesthesia. 2021; 76 Suppl 1:171-181.
26. Bruno RR, Wolff G, Wernly B, Masyuk M, Piayda K, Leaver S, Erkens R, Oehler D, Afzal S, Heidari H, Kelm M, Jung C. Virtual and augmented reality in critical care medicine: the patient's, clinician's, and researcher's perspective. Crit Care. 2022; 26:326.
27. Yeung AWK, Tosevska A, Klager E, Eibensteiner F, Laxar D, Stoyanov J, Glisic M, Zeiner S, Kulnik ST, Crutzen R, Kimberger O, Kletecka-Pulker M, Atanasov AG, Willschke H. Virtual and Augmented Reality Applications in Medicine: Analysis of the Scientific Literature. J Med Internet Res. 2021; 23:e25499.
28. Ma C, Chen G, Zhang X, Ning G, Liao H. Moving-Tolerant Augmented Reality Surgical Navigation System Using Autostereoscopic Three-Dimensional Image Overlay. IEEE J Biomed Health Inform. 2019; 23:2483-2493.
29. Marescaux J, Clément JM, Tassetti V, Koehl C, Cotin S, Russier Y, Mutter D, Delingette H, Ayache N. Virtual reality applied to hepatic surgery simulation: the next revolution. Ann Surg. 1998; 228:627-634.
30. Krummel TM. Surgical simulation and virtual reality: the coming revolution. Ann Surg. 1998; 228:635-637.
31. Liao H, Hata N, Nakajima S, Iwahara M, Sakuma I, Dohi T. Surgical navigation by autostereoscopic image overlay of integral videography. IEEE Trans Inf Technol Biomed. 2004; 8:114-121.
32. Studeli T, Kalkofen D, Risholm P, Ali W, Freudenthal A, Samset E. Visualization tool for improved accuracy in needle placement during percutaneous radio-frequency ablation of liver tumors. in Proceedings of Medical Imaging 2008: Visualization, Image-Guided Procedures, and Modeling. PTS 1 and 2, 2008.
33. Sugimoto M, Yasuda H, Koda K, Suzuki M, Yamazaki M, Tezuka T, Kosugi C, Higuchi R, Watayo Y, Yagawa Y, Uemura S, Tsuchiya H, Azuma T. Image overlay navigation by markerless surface registration in gastrointestinal, hepatobiliary and pancreatic surgery. J Hepatobiliary Pancreat Sci. 2010; 17:629-636.
34. Hansen C, Ritter F, Wieferich J, Hahn H, Peitgen HO. Illustration of Vascular Structures for Augmented Reality in Liver Surgery. In: World Congress on Medical

Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany: Vol. 25/4 Image Processing, Biosignal Processing, Modelling and Simulation, Biomechanics. 2010; pp. 2113-2116.
35. Shekhar R, Dandekar O, Bhat V, Philip M, Lei P, Godinez C, Sutton E, George I, Kavic S, Mezrich R, Park A. Live augmented reality: a new visualization method for laparoscopic surgery using continuous volumetric computed tomography. Surg Endosc. 2010; 24:1976-1985.
36. Konishi K, Nakamoto M, Kakeji Y, Tanoue K, Kawanaka H, Yamaguchi S, Ieiri S, Sato Y, Maehara Y, Tamura S, Hashizume M. A real-time navigation system for laparoscopic surgery based on three-dimensional ultrasound using magneto-optic hybrid tracking configuration. Int J Comput Assist Radiol Surg. 2007; 2:1-10.
37. Gavaghan KA, Peterhans M, Oliveira-Santos T, Weber S. A portable image overlay projection device for computer-aided open liver surgery. IEEE Trans Biomed Eng. 2011; 58:1855-1864.
38. Freudenthal A, Studeli T, Lamata P, Samset E. Collaborative co-design of emerging multi-technologies for surgery. J Biomed Inform. 2011; 44:198-215.
39. Okamoto T, Onda S, Matsumoto M, Gocho T, Futagawa Y, Fujioka S, Yanaga K, Suzuki N, Hattori A. Utility of augmented reality system in hepatobiliary surgery. J Hepatobiliary Pancreat Sci. 2013; 20:249-253.
40. Marzano E, Piardi T, Soler L, Diana M, Mutter D, Marescaux J, Pessaux P. Augmented Reality-Guided Artery-First Pancreatico-Duodenectomy. J Gastrointest Surg. 2013; 17:1980-1983.
41. Kenngott HG, Wagner M, Gondan M, Nickel F, Nolden M, Fetzer A, Weitz J, Fischer L, Speidel S, Meinzer HP, Böckler D, Büchler MW, Müller-Stich BP. Real-time image guidance in laparoscopic liver surgery: first clinical experience with a guidance system based on intraoperative CT imaging. Surg Endosc. 2014; 28:933-940.
42. Katic D, Wekerle AL, Gortler J, Spengler P, Bodenstedt S, Rohl S, Suwelack S, Kenngott HG, Wagner M, Muller-Stich BP, Dillmann R, Speidel S. Context-aware Augmented Reality in laparoscopic surgery. Comput Med Imaging Graph. 2013; 37:174-182.
43. Haouchine N, Dequidt J, Peterlik I, Kerrien E, Berger MO, Cotin S, Ieee. Towards an Accurate Tracking of Liver Tumors for Augmented Reality in Robotic Assisted Surgery. In: 2014 IEEE International Conference on Robotics and Automation (ICRA). 2014; pp. 4121-4126.
44. Giannone F, Felli E, Cherkaoui Z, Mascagni P, Pessaux P. Augmented Reality and Image-Guided Robotic Liver Surgery. Cancers (Basel). 2021; 13:6268.
45. Okamoto T, Onda S, Yasuda J, Yanaga K, Suzuki N, Hattori A. Navigation Surgery Using an Augmented Reality for Pancreatectomy. Dig Surg. 2015; 32:117-123.
46. Tang R, Ma LF, Xiang CH, Wang XD, Li A, Liao HE, Dong JH. Augmented reality navigation in open surgery for hilar cholangiocarcinoma resection with hemihepatectomy using video-based in situ three-dimensional anatomical modeling A case report. Medicine (Baltimore). 2017; 96:e8083.
47. Lin JY, Tao HS, Wang ZX, Chen R, Chen YL, Lin WJ, Li BH, Fang CH, Yang J. Augmented reality navigation facilitates laparoscopic removal of foreign body in the pancreas that cause chronic complications. Surg Endosc. 2022; 36:6326-6330.
48. Balci D, Kirimker EO, Raptis DA, Gao YJ, Kow AWC. Uses of a dedicated 3D reconstruction software with augmented and mixed reality in planning and performing advanced liver surgery and living donor liver transplantation. Hepatobiliary Pancreat Dis Int. 2022; 21:455-461.
49. Huber T, Huettl F, Hanke LI, Vradelis L, Heinrich S, Hansen C, Boedecker C, Lang H. Liver Surgery 4.0-Planning, Volumetry, Navigation and Virtual Reality. Zentralbl Chir. 2022; 147:361-368. (in German)
50. Ayache N, Cotin S, Delingette H, Clemente JM, Russier Y, Marescaux J. Simulation of endoscopic surgery. Minimally Invasive Therapy and Allied Technologies. 1998; 7:71-77.
51. Bornik A, Beichel R, Reitinger B, Gotschuli G, Sorantin E, Leberl F, Sonka M. Computer-aided liver surgery planning: an augmented reality approach. In: Medical Imaging 2003: Visualization, Image-Guided Procedures, and Display. 2003; pp. 395-406.
52. Reitinger B, Bornik A, Beichel R, Werkgartner G, Sorantin E. Tools for augmented reality based liver resection planning. In: Medical Imaging 2004: Visualization, Image-Guided Procedures, and Display. 2004; pp. 88-99.
53. Deng K, Wei B, Chen M, Huang ZY, Wu H. Realization of real-time X-ray stereoscopic vision during interventional procedures. Sci Rep. 2018; 8:15852.
54. De Paolis LT. Augmented Visualization as Surgical Support in the Treatment of Tumors. In: Bioinformatics and Biomedical Engineering , IWBBIO 2017, PT I, 2017; pp. 432-443.
55. Tang R, Ma LF, Rong ZX, Li MD, Zeng JP, Wang XD, Liao HE, Dong JH. Augmented reality technology for preoperative planning and intraoperative navigation during hepatobiliary surgery: A review of current methods. Hepatobiliary Pancreat Dis Int. 2018; 17:101-112.
56. Beyer-Berjot L, Palter V, Grantcharov T, Aggarwal R. Advanced training in laparoscopic abdominal surgery: A systematic review. Surgery. 2014; 156:676-688.
57. Strickland A, Fairhurst K, Lauder C, Hewett P, Maddern G. Development of an ex vivo simulated training model for laparoscopic liver resection. Surg Endosc. 2011; 25:1677-1682.
58. Nomura T, Mamada Y, Nakamura Y, Matsutani T, Hagiwara N, Fujita I, Mizuguchi Y, Fujikura T, Miyashita M, Uchida E. Laparoscopic skill improvement after virtual reality simulator training in medical students as assessed by augmented reality simulator. Asian J Endosc Surg. 2015; 8:408-412.
59. Viglialoro RM, Esposito N, Condino S, Cutolo F, Guadagni S, Gesi M, Ferrari M, Ferrari V. Augmented Reality to Improve Surgical Simulation: Lessons Learned Towards the Design of a Hybrid Laparoscopic Simulator for Cholecystectomy. IEEE Trans Biomed Eng. 2018; doi: 10.1109/TBME.2018.2883816
60. Schott D, Saalfeld P, Schmidt G, Joeres F, Boedecker C, Huettl F, Lang HK, Huber T, Preim B, Hansen C, Society IC. A VR/AR Environment for Multi-User Liver Anatomy Education. In: 2021 IEEE VIRTUAL REALITY AND 3D USER INTERFACES (VR) (2021; pp. 296-305.
61. Andolfi C, Plana A, Kania P, Banerjee PP, Small S. Usefulness of Three-Dimensional Modeling in Surgical

Planning, Resident Training, and Patient Education. J Laparoendosc Adv Surg Tech A.. 2017; 27:512-515.
62. Ghaednia H, Fourman MS, Lans A, Detels K, Dijkstra H, Lloyd S, Sweeney A, Oosterhoff JHF, Schwab JH. Augmented and virtual reality in spine surgery, current applications and future potentials. Spine J. 2021; 21:1617-1625.
63. Diana M, Soler L, Agnus V, D'Urso A, Vix M, Dallemagne B, Faucher V, Roy C, Mutter D, Marescaux J, Pessaux P. Prospective Evaluation of Precision Multimodal Gallbladder Surgery Navigation Virtual Reality, Near-infrared Fluorescence, and X-ray-based Intraoperative Cholangiography. Ann Surg. 2017; 266:890-897.
64. Cheung TT, Ma KW, She WH, Dai WC, Tsang SHY, Chan ACY, Chok KSH, Lo CM. Pure laparoscopic hepatectomy with augmented reality-assisted indocyanine green fluorescence versus open hepatectomy for hepatocellular carcinoma with liver cirrhosis: A propensity analysis at a single center. Asian J Endosc Surg. 2018; 11:104-111.
65. Zhu W, Zeng XJ, Hu HY, Xiang N, Zeng N, Wen S, Tian J, Yang J, Fang CH. Perioperative and Disease-Free Survival Outcomes after Hepatectomy for Centrally Located Hepatocellular Carcinoma Guided by Augmented Reality and Indocyanine Green Fluorescence Imaging: A Single-Center Experience. J Am Coll Surg. 2023; 236:328-337.
66. Hargest R. Five thousand years of minimal access surgery: 1990-present: organisational issues and the rise of the robots. J R Soc Med. 2021; 114:69-76.

Received April 18, 2023; Revised June 21, 2023; Accepted June 24, 2023.

Address correspondence to:
Chuan Li, Division of Liver Surgery, Department of General Surgery, West China Hospital, Sichuan University, 37 Guoxue Lane, Wuhou District, Chengdu City, Sichuan Province, China.
E-mail: lichuan@scu.edu.cn

Released online in J-STAGE as advance publication June 26, 2023.

Mini-Review

DOI: 10.5582/bst.2023.01076

Donor-recipient matching in adult liver transplantation: Current status and advances

Caterina Accardo[1], Ivan Vella[1], Duilio Pagano[1], Fabrizio di Francesco[1], Sergio Li Petri[1], Sergio Calamia[1], Pasquale Bonsignore[1], Alessandro Tropea[1], Salvatore Gruttadauria[1,2,*]

[1] Department for the Treatment and Study of Abdominal Diseases and Abdominal Transplantation, Istituto di Ricovero e Cura a Carattere Scientifico-Istituto Mediterraneo per i Trapianti e Terapie ad alta specializzazione (IRCCS-ISMETT), University of Pittsburgh Medical Center (UPMC), Palermo, Italy;
[2] Department of Surgery and Medical and Surgical Specialties, University of Catania, Catania, Italy.

SUMMARY The match between donor and recipient (D-R match) in the field of liver transplantation (LT) is one of the most widely debated topics today. Within the cohort of patients waiting for a transplant, better matching of the donor organ to the recipient will improve transplant outcomes, and benefit the waiting list by minimizing graft failure and the need for re-transplantation. In an era of suboptimal matches due to the sparse organ pool and the increase in extended criteria donors (ECD), ensuring adequate outcomes becomes the primary goal for clinicians in the field. The objective of this mini-review is to analyze the main variables in the evaluation of the D-R match to ensure better outcomes, the existence of scores that can help in the realization of this match, and the latest advances made thanks to the technology and development of artificial intelligence (AI).

Keywords liver, transplantation, donor-recipient, matching, outcomes, allocation

1. Introduction

With the increasing number of liver transplant (LT) candidates and the sparse pool of available organs, the rationale for allocating a graft to a potential recipient on the list is now a lively topic in the field of liver transplantation. In addition, at a time in which marginal organs are increasingly available, to ensure a good post-transplant outcome it is necessary to find the most appropriate way to allocate these grafts to the most suitable recipients. Allograft allocation should and can be more precise and personalized, but D-R matching is really a problem of classification, in which some donor variables are combined with variables of the listed recipients, surgical considerations, and logistical factors; in short, it is a quite complex process (*1,2*).

In 2002, liver allograft allocation changed with the implementation of the Model for End-Stage Liver Disease (MELD) scoring system. Given its effectiveness in predicting short-term mortality, priority in this score is based on the "sicker first" principle (*3*). However, some of its limitations have been underscored over time. One of these is that some pathologies do not have an adequate priority (*e.g.*, hepatocellular carcinoma) because their prognosis is not directly related to the underlying liver function but to the risk of disease progression.

The MELD score system is not useful for predicting survival after LT in an era in which recipient and donor combinations can be suboptimal matches.

Another aspect that needs to be analyzed is the impact of the advent of extended criteria donors (ECDs). The use of ECDs has increased the donor pool but, on the other hand, it has also worsened its quality (*4*). Today, we have various organ preservation techniques that have allowed us to increase the pool of transplantable organs among ECDs, increasing and improving recipient outcomes (*5,6*). Despite this important goal, a good percentage of grafts remain non-transplantable (8.4% in a U.S. series) (*7*).

If, with an appropriate matching between the donor and the recipient, even organs defined as marginal can produce a good outcome in select patients (*8*) , an ideal prioritization system should be valid for all patients and diseases, and should be able to assign the organ to the patient with the highest risk of mortality, and at the same time with the best predictable post-transplant survival (*3*).

Various scores have therefore been developed over the years with the aim of guaranteeing an adequate donor-recipient match, and an improvement in post-transplant survival rates. Moreover, in the last decade,

investments have been made in AI for the definition of the best D-R match, though we are still awaiting satisfactory results on its real clinical applicability.

The aim of this mini-review is to analyze the most widely used scores and the most studied variables in D-R matching, providing a snapshot of the current state of the literature, with an eye to new frontiers.

2. The most widely analyzed variables in the literature

The principal variables analyzed in donor-recipient matching are summarized in Table 1.

2.1. Donor-to-recipient age match

Considering the increase in marginal grafts, several studies have focused on the role of recipient and donor characteristics such as age match, seeking to identify predictive factors that reduce the risk of graft failure and patient death (*9*).

A single-center retrospective analysis of 849 deceased donor LTs by Gilbo *et al*. (*10*) shows that matching older donor livers with older recipients does not affect long-term outcomes because there is no exponential increase in age-related risks, provided that other risk factors are absent or minimized. Chapman *et al*. (*11*) also did not observe any difference in patient and graft survival in LTs matched or mismatched per age. Recently, Nakamura *et al*. found that elderly liver grafts showed slower recovery trajectories in the acute phase, but finally achieved acceptable outcomes (*12*). These results are in contrast with results from large

Table 1. The main variables analyzed in the literature on donor-recipient matching

Variable	Authors	Place and year of publication	Conclusions
Age	Vitale A. *et al*. (*9*)	Italy, 2011	**Donor age >70 y** among the criteria for defining a **suboptimal liver.**
	Pagano D. *et al*. (*14*)	Italy, 2013	**Age mismatch** is an independent **risk factor for patient death**.
	Chapman WC *et al*. (*11*)	U.S., 2015	**Comparable outcomes** in graft and patient survivals **using older donors** (> 60 y) without increased rate of complications.
	Gilbo N. *et al*. (*10*)	Belgium, 2019	**Older livers can be safely used in older recipients** if other risk factors are minimized.
	Nakamura T. *et al*. (*12*)	Japan, 2022	**Elderly liver grafts** exhibit slower recovery trajectories in the acute phase but **finally achieve acceptable outcomes**.
	Caso maestro O. *et al*. (*15*)	Spain, 2022	The results of **LT with nonagenarian liver grafts** are not significantly different from those obtained with octogenarian donors, with **satisfactory outcomes**.
Size	Levesque E. *et al*. (*21*)	France, 2013	Using **large grafts for recipient size did not impair liver function and did not modify graft and patient outcomes at one year**.
	Croome KP *et al*. (*22*)	U.S., 2015	Donors with a calculated **sTLV size ratio ≥ 1.25 have an increased risk of EAD**.
	KW Ma *et al*. (*20*)	China, 2019	**SFSG is associated with inferior medium-term but not long-term graft survival.**
	Reyes J. *et al*. (*19*)	U.S., 2019	In deceased donor LT, **the D/R body surface area ratio is a significant predictor of graft survival**.
	Kubal CA *et al*. (*24*)	U.S., 2021	**Significant donor-recipient body size mismatch did not have a negative impact on early and long-term outcomes**.
	Addeo P. *et al*. (*25*)	France, 2022	**Combination of anthropometrics of the donors with imaging of the recipients** can be **helpful in improving** the process of **donor-recipient matching** and avoiding complications.
	Kostakis ID *et al*. (*23*)	U.K., 2023	**Donor-recipient size mismatch affects the rates of portal vein thrombosis within the first 3 months** and overall graft survival.
Gender	Rustgi VK *et al*. (*26*)	U.S., 2022	**Gender-mismatched patients have a 6.9% increase in likelihood of graft failure**.
	Lehner F. *et al*. (*28*)	Germany, 2009	**Gender-incompatible LT is not a confounder in patient survival**.
	Schoening WN *et al*. (*27*)	Germany, 2016	The impressive **long-term graft survival benefit of gender mismatch** versus matched groups in LT may be caused by significant differences in donor quality and recipient characteristics, and **may not be related to gender itself**.
	Lai Q. *et al*. (*29*)	Italy, 2018	**Gender mismatch is a risk factor for poor graft survival after LT** (female-to-male mismatch represents the worst combination).
	Germani G. *et al*. (*30*)	Italy, 2020	**Donor/recipient gender mismatch in male recipients**, and the use of obese donors in female recipients are associated with **reduced survival after LT**.

registry studies from the past (*13*).

Our previous analysis (*14*) indicated that both recipient and donor ages were predictors of transplant outcome; patients of the same age were more likely to show better graft survival and longer lifespan.

Finally, there is currently no consensus on the donor age limit for liver transplantation, due mainly to recent improvements in outcomes with elderly donors (*15*).

2.2. Donor-to-recipient dimensional match

The impact of donor and recipient size mismatch in deceased whole liver transplantation has not been well studied. The consequences of size mismatch using whole grafts have been shown to increase the risk of developing "small for size syndrome", in which the transplanted liver is unable to ensure the functional and metabolic needs of the recipient.

On the other hand, a "large for size donor" can cause graft damage caused by vascular thrombosis or graft necrosis secondary to poor blood supply (*16*).
Body surface area (BSA) has proven to be an excellent indicator of metabolic mass because it is less influenced by fat mass, and therefore allows prediction with a good approximation the liver volume (*17*). The donor to recipient body surface area ratio (DR_BSAR = BSAdonor/BSArecipient) has been studied to determine its influence on graft survival (*18*).

Reyes *et al*. conducted a retrospective analysis of 79,704 liver transplants performed in the U.S., and found that in whole liver transplantation from deceased donors, DR_BSAR is a significant predictor of graft survival (*19*), thus demonstrating the importance of the correct dimensional match between donor and recipient.

Most studies have concluded that a liver graft needs to have at least 0.8% of the recipient's weight or 35% of his/her ideal liver volume (*20*), but not more than 2.5% of the recipient's weight or 125% of his/her ideal liver volume (*21,22*). In a series by Kostakis *et al*. (*23*), in which 11,245 transplants were considered (the liver grafts were from donors after brain death (DBD) in 9,504 (84.5%) transplants and from donors after circulatory death (DCD) in 1,741 (15.5%) transplants, three distinct categories were identified: donor size of 85% of recipient size or less, donor size more than 85%, but up to 140% of recipient size, and donor size more than 140% of recipient size. There were statistically significant differences for overall graft survival among these 3 categories ($P < 0.001$): the first category had shorter overall graft survival (75^{th} percentile: 4,514 days) than the second category (75th percentile: 5,721 days), while the third had much shorter overall graft survival than both of the others. The data was confirmed by multivariable Cox regression analysis. The findings of this study demonstrated that donor-recipient size mismatch affected the risk of developing portal vein thrombosis within the first 3 months after transplantation, and reduced the overall graft survival. These data were confirmed by other studies on the same topic (*19,24*).

Another useful dimensional parameter is that of the Strasbourg group: the maximum volume that a graft to be implanted in patients with cirrhosis can have is the sum of the volume of the recipient's liver and the dimensions of the right upper abdominal cavity. This overall volume correlates with the severity of portal hypertension (ascites, large direct portosystemic shunt) and right anteroposterior cavity diameter (RAP; measured above the hepatic dome). The RAP reflects the compliance of the right hypochondria and correlates linearly with the overall volume. This combination of donor anthropometry with recipient imaging can be helpful in improving the donor-recipient size match, according to Addeo (*25*).

2.3. Donor-to-recipient gender match

Gender match seems to be one of the aspects that influences outcomes after LT, though this association is controversial. In the past, some monocentric studies have underlined a correlation between donor gender and graft loss (*26*), especially in male recipients of female donors (*27*), in contrast to other studies (*28*). In recent times, a number of scores have been proposed with the aim of predicting the post-transplant outcome, though none has identified the donor gender as a risk factor for poor graft survival.

A more recent meta-analysis conducted in 2018 suggests a detrimental role of the female-male (F-M) mismatch in terms of graft survival (*29*). These results are absolutely in line with several experiences worldwide (*27*). Nevertheless, Lai's group argues that there are several confounding factors to consider in these analyses.

Also Germani *et al*. pointed out recently that donor/recipient gender mismatch in male recipients and use of obese donors in female recipients are associated with decreased survival after LT, highlighting the importance of associating an anthropometric evaluation with the gender mismatch in the allocation process to have better long-term outcomes (*30*).

In conclusion, the impact of gender mismatch on post-transplant outcomes is still much debated in the literature. Further large, well-calibrated studies are needed, with the aim of definitively clarifying the potential harmful role of gender mismatch in the liver transplantation setting.

3. Scores in the literature

Liver transplantation is today the most appropriate treatment for end-stage liver disease, and a myriad of factors, as we have seen, relating to the donor, the recipient, the anesthetic-surgical procedure, and the

Table 2. Main scores proposed for donor-recipient matching in the graft allocation process

Score	Authors	Place and year of publication	Variables
DRI (Donor Risk Index)	Feng S. *et al*. (*32*)	U.S., 2006	**Donor age, race, height**, death from cerebrovascular accident (**CVA**), donation after cardiac death (**DCD**), **cause of death classified as "other"** (excluding trauma, CVA, or anoxia), **split or partial graft, cold ischemia time and location of organs** based on donor service area.
P-SOFT (the Preallocation score to predict Survival Out-comes Following Liver Transplant Score) and **SOFT Score**	Rana A. *et al*. (*41*)	U.S., 2008	**Age, BMI, previous transplant or abdominal surgery, albumin < 2.0 g/dL, dialysis before transplantation, ICU pretransplant, MELD score, life support pretransplant, encephalopathy, portal vein thrombosis, ascites pretransplant.** **SOFT: P-SOFT + points awarded from donor criteria,** 1 recipient condition **(portal bleed 48-h pre-transplant)** and two logistical factors **(CIT and national allocation)** at the time of graft allocation.
D-MELD	Halldorson JB *et al*. (*42*)	U.S., 2009	The product of **donor age** and **preoperative MELD**, calculated from laboratory values.
BAR (Balance of Risk)	Dutkowski P. *et al*. (*33*)	Switzerland, 2011	**Donor age** (years), **recipient age** (years), **cold ischemia time** (hours), **retransplantation** (yes/no), **life support** (yes/no), and the **MELD score** at the time of liver transplant (true value without exception points).
ET-DRI (Eurotransplant-Donor-Risk-Index)	Braat AE *et al*. (*13*)	Eurotransplant region, 2012	**Donor age, cause of death, donation after cardiac death, split liver graft, organ location** (regional or national), **cold ischemia time, rescue allocation**, and **gamma-glutamyltransferase levels.**
ISO (Italian Score for Organ allocation)	Cillo U. *et al*. (*39*)	Italy, 2015	Based on principles of urgency, utility, and transplant benefit, the score considers in addition to pure **MELD**, **exceptions**, and **hepatocellular carcinoma**

management of the intensive care unit are involved in the onset of complications, survival, and related costs. So with the aim of being able to analyze all these variables in the shortest possible time in the organ allocation process, various indexes have been developed (*31*), summarized in Table 2. The objective is to predict post-transplant outcomes.

3.1. Donor risk index (DRI)

Feng *et al*. (*32*) discuss the concept of the DRI, which objectively evaluates donor variables involved in transplant outcomes: donor age, DCD, and split/partial grafts are strongly associated with graft failure; African-American race, short stature, and cerebrovascular accident as the cause of death are modestly associated with graft failure. The DRI assessment offers an evidence-based D-R match. However, when DRI is assessed at the time of offering, D-R matching is not easily programmable, and the distribution of risks from donor and recipient is more dependent on the allocation scheme. By itself, DRI is a suboptimal tool for D-R matching.

3.2. Balance of risk (BAR) score

In 2011, Dutkowski *et al*. designed the BAR score, based on a combination of the principles of prognosis and justice. The main advantage is that it is based on objective factors available at the time of organ allocation, with the exception of cold ischemia time. The main disadvantage is that it does not consider other determinants, such as graft steatosis (*33*). Researchers, such as Schlegel, have proven that the BAR score is useful for finding good donor-recipient matches, but in other studies it showed a suboptimal ability (*34,35*). In the 336 patient sample of Lopez *et al*., the BAR score was found to be inaccurate in predicting liver transplant survival (*36*), and unable to identify which of several D–R pairs will get the best result.

3.2. Eurotransplant Donor Risk Index (ET-DRI)

Braat *et al*. advanced the idea that the ET-DRI, a scoring system tailored for the Eurotransplant region, may be a useful tool for liver allocation in the future (*13*). In a retrospective single center study by Schoening (*37*) *et al*., when combining donor (ET-DRI) and recipient factors (indication and/or lab-MELD), an estimation of long-term graft survival seems possible. This score was therefore considered to have a limited impact on the prediction of early outcome following LT in other series (38).

3.3. Italian Score for Organ allocation (ISO) system

The Italian Board of Experts in the Field of Liver Transplantation has developed the ISO system, incorporating a priority criterion for MELD exception conditions (*39*). In our previous monocentric case study (*40*), there was clearly a significant reduction in deaths while waiting for a liver, and an increase in the percentage of LT recipients with the application of the ISO score, though this would have to be validated

prospectively to confirm its superiority compared to the MELD score.

3.4. Pre-allocation Survival Out-comes Following liver Transplant (P-SOFT) score and SOFT score

Another ambitious scoring system that predicts recipient survival after LT was proposed by Rana *et al*. (*41*): they identified 4 donors, 13 recipients, and 1 operative variable as significant predictors of 3-month mortality following LT. Consequently, two complementary scoring systems were designed: the P-SOFT score and the SOFT score, which are the result of adding P-SOFT points to points awarded from donor criteria, 1 recipient condition (portal bleed 48 hours pre-transplant), and two logistical factors (cold ischemia time [CIT] and national allocation) at the time of graft allocation. It seems that the SOFT and P-SOFT scores are adequate in predicting 90-day mortality. However, the inclusion of multiple variables, some of which are partially subjective and only semi-quantitative (*e.g.*, encephalopathy, ascites) and a complex underlying statistical model, limits its clinical applicability in pre-transplant decision-making (34,38).

3.5. Donor age × recipient Modified for End-stage Liver Disease [MELD] score (D-MELD score)

In 2009, Halldorson *et al*. (*42*) proposed a simple score, D-MELD, combining the sickest-first principle (lab-MELD) and DRI (donor age). The product of these continuous variables produces an increased risk of mortality and complications, calculated as length of hospitalization. A D-MELD cut-off score of 1600 defines a subset of D-R matches with worse outcomes. The positive aspects of this system are simplicity, objectivity, and transparency, but the prognostic ability of the D-MELD is lacking in LT centers using a more complex D-R matching policy (*9*).

The difficulty in identifying an effective score is due to the myriad of variables to be considered in the match between donor and recipient in each transplant.

4. Artificial intelligence (AI)

The ideal D-R matching system remains a chimera. Unfortunately, to date, the scores available are not statistically robust enough. Arguably, the human mind may not be precise enough to put so many interacting variables in order.

In this context, new technologies that exploit AI have been developing recently.

AI is a branch of computational science that studies computational models capable of performing activities similar to human ones based on two characteristics: behavior and reasoning. Machine learning is defined as a branch of AI that focuses on using data and algorithms to mimic how humans learn, and gradually improve the accuracy of the algorithms (*43*).

Today, AI is revolutionizing the field of hepatology and liver surgery, and its application is becoming frequent in the clinical setting (*44*).

In the D-R match, different variables (donor, recipient, and logistics) are combined to obtain two possible outcomes: graft survival or graft loss at different endpoints (3 and 12 months are the most commonly used). However, no current allocation system is capable of achieving an ideal match. This means that these systems are unable to identify the candidate on the waiting list with the highest probability of death, and identify, among all available grafts, the one with the highest probability of post-transplant success for this candidate (*43*).

How does AI fit into the complex match between donor and recipient?

Clinical decisions have both an objective and a subjective component (*45*): scientific data, memory, and previous experiences serve as the basis of mature clinical reasoning, while other considerations, such as intuition or emotions, constitute the subjective component.

Therefore, clinical decisions in D-R matching have an inherent emotional bias; furthermore, a single D-R match may include about 100 parameters between donor and recipient characteristics and logistics to take into consideration.

The principal AI model used in D-R matching is artificial neural networks (ANN).

Deep learning classifiers use several previous experiences based on objective data (database) to be able to make the best decision for which they have been programmed. The subjective sphere of the decision is removed, and large amounts of data are managed in a short time, which is why AI and, in particular, deep learning classifiers represent an interesting alternative to traditional scores today (*46*).

Briceño *et al*. (*47*) were the first to apply a neural network combined with a system of rules to create a donor-recipient allocation model (M.A.D.R.E model). This multicenter study included a total of 1,003 liver transplants performed between 2007 and 2008, using 64 donor and recipient variables. The probability of graft failure at 3 months was the endpoint variable. The authors demonstrated the superiority of ANNs in donor allocation over biostatistics-based prioritization scores (MELD, D-MELD, SOFT, P-SOFT, DRI, and BAR).

A second study was performed with a dataset of 858 D-R pairs from liver transplants at King's College Hospital, in London. The authors found that the model obtained with this database achieved excellent results at 3 months and 12 months, and when compared with other scores such as MELD and BAR, had 15% more favorable results (*48*).

Indeed, in clinical scenarios, neural networks are

very useful in processing complex patterns because they can generate near-perfect predictions by analyzing multiple data rapidly, which is crucial in the allocation process (*49*).

The implementation of AI in the field of liver diseases has grown exponentially, but the number of clinical studies addressing D–R matching is small. Most of the studies mentioned are observational, and very few cases of clinical validation of the model are available (*48*). We are also starting to have studies that do not see the superiority of these complex deep learning systems when applied to larger databases, such as that of the United Network for Organ Sharing (UNOS) (Organ Procurement and Transplantation Network (OPTN). United Network for Organ Sharing (UNOS); 2020. Available from: *https://www.unos.org/*.) (*50*)

Therefore, they are not useful with large databases due to the extreme number of decision trees they would generate, making them impractical.

5. Conclusions

The match between donor and recipient in the delicate process of allocating grafts is a highly debated topic in the field of liver transplantation. The objective of the scientific community is to find a system that facilitates the match process in terms of speed, costs, and global outcomes.

In clinical practice, we do not yet have scores that are robust and effective in large databases.

The progress made in the application of deep learning to the field of liver transplantation bodes well for a future in which marginal grafts will increase, and the maintenance of adequate outcomes will increasingly depend on our ability to guarantee an adequate match between the donor and the recipient.

Funding: Funded by the Italian Ministry of Health - RC.

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

References

1. Marino IR. Are we ready to match donor and recipient in liver transplantation? Liver Transpl. 2006;12:1574-1576.
2. Widmer J, Eden J, Carvalho MF, Dutkowski P, Schlegel A. Machine Perfusion for Extended Criteria Donor Livers: What Challenges Remain? J Clin Med. 2022; 11:5218.
3. Jurado-García J, Muñoz García-Borruel M, Rodríguez-Perálvarez ML, Ruíz-Cuesta P, Poyato-González A, Barrera-Baena P, Fraga-Rivas E, Costán-Rodero G, Briceño-Delgado J, Montero-Álvarez JL, de la Mata-García M. Impact of MELD Allocation System on Waiting List and Early Post-Liver Transplant Mortality. PLoS One. 2016; 11:e0155822.
4. Akkina SK, Asrani SK, Peng Y, Stock P, Kim WR, Israni AK. Development of organ-specific donor risk indices. Liver Transpl. 2012; 18:395-404.
5. Sousa Da Silva RX, Weber A, Dutkowski P, Clavien PA. Machine perfusion in liver transplantation. Hepatology. 2022; 76:1531-1549.
6. Scalera I, De Carlis R, Patrono D, Gringeri E, Olivieri T, Pagano D, Lai Q, Rossi M, Gruttadauria S, Di Benedetto F, Cillo U, Romagnoli R, Lupo LG, De Carlis L. How useful is the machine perfusion in liver transplantation? An answer from a national survey. Front Surg. 2022; 9:975150.
7. Israni AK, Zaun D, Hadley N, Rosendale JD, Schaffhausen C, McKinney W, Snyder JJ, Kasiske BL. OPTN/SRTR 2018 Annual Data Report: Deceased Organ Donation. Am J Transplant. 2020; 20 Suppl s1:509-541.
8. Jackson KR, Motter JD, Haugen CE, Long JJ, King B, Philosophe B, Massie AB, Cameron AM, Garonzik-Wang J, Segev DL. Minimizing Risks of Liver Transplantation With Steatotic Donor Livers by Preferred Recipient Matching. Transplantation. 2020; 104:1604-1611.
9. Vitale A, Ramirez Morales R, dalla Bona E, Scopelliti M, Zanus G, Neri D, d'Amico F, Gringeri E, Russo F, Burra P, Angeli P, Cillo U. Donor-Model for End-Stage Liver Disease and donor-recipient matching in liver transplantation. Transplant Proc. 2011; 43:974-976
10. Gilbo N, Jochmans I, Sainz-Barriga M, Nevens F, van der Merwe S, Laleman W, Verslype C, Cassiman D, Verbeke L, van Malenstein H, Roskams T, Pirenne J, Monbaliu D. Age Matching of Elderly Liver Grafts With Elderly Recipients Does Not Have a Synergistic Effect on Long-term Outcomes When Both Are Carefully Selected. Transplant Direct. 2019; 5:e342.
11. Chapman WC, Vachharajani N, Collins KM, Garonzik-Wang J, Park Y, Wellen JR, Lin Y, Shenoy S, Lowell JA, Doyle MB. Donor Age-Based Analysis of Liver Transplantation Outcomes: Short- and Long-Term Outcomes Are Similar Regardless of Donor Age. J Am Coll Surg. 2015; 221:59-69.
12. Nakamura T, Nobori S, Harada S, Sugimoto R, Yoshikawa M, Ushigome H, Yoshimura N. Impact of Donor and Recipient Age on Outcomes After Living Donor Liver Transplant. Transplant Proc. 2022; 54:438-442.
13. Braat AE, Blok JJ, Putter H, Adam R, Burroughs AK, Rahmel AO, Porte RJ, Rogiers X, Ringers J; European Liver and Intestine Transplant Association (ELITA) and Eurotransplant Liver Intestine Advisory Committee (ELIAC). The Eurotransplant donor risk index in liver transplantation: ET-DRI. Am J Transplant. 2012; 12:2789-2796.
14. Pagano D, Grosso G, Vizzini G, Spada M, Cintorino D, Malaguarnera M, Donati M, Mistretta A, Gridelli B, Gruttadauria S. Recipient-donor age matching in liver transplantation: a single-center experience. Transplant Proc. 2013; 45:2700-2706.
15. Caso Maestro O, Justo Alonso I, Marcacuzco Quinto A, Manrique Municio A, Calvo Pulido J, García-Sesma A, Jiménez-Romero C. Expanding donor age in liver transplantation using liver grafts from nonagenarian donors. Clin Transplant. 2022; 36:e14684.
16. Kiuchi T, Kasahara M, Uryuhara K, Inomata Y, Uemoto S, Asonuma K, Egawa H, Fujita S, Hayashi M, Tanaka K. Impact of graft size mismatching on graft prognosis in liver transplantation from living donors. Transplantation. 1999; 67:321-327.
17. DuBois EF. Basal Metabolism in Health and Disease.Lea

& Febiger, Philadelphia, Pa, 1936.
18. Mosteller RD. Simplified calculation of body-surface area. N Engl J Med. 1987; 317:1098.
19. Reyes J, Perkins J, Kling C, Montenovo M. Size mismatch in deceased donor liver transplantation and its impact on graft survival. Clin Transplant. 2019; 33:e13662.
20. Ma KW, Wong KHC, Chan ACY, Cheung TT, Dai WC, Fung JYY, She WH, Lo CM, Chok KSH. Impact of small-for-size liver grafts on medium-term and long-term graft survival in living donor liver transplantation: A meta-analysis. World J Gastroenterol. 2019; 25:5559-5568.
21. Levesque E, Duclos J, Ciacio O, Adam R, Castaing D, Vibert E. Influence of larger graft weight to recipient weight on the post-liver transplantation course. Clin Transplant. 2013; 27:239-247
22. Croome KP, Lee DD, Saucedo-Crespo H, Burns JM, Nguyen JH, Perry DK, Taner CB. A novel objective method for deceased donor and recipient size matching in liver transplantation. Liver Transpl. 2015; 21:1471-7.
23. Kostakis ID, Raptis DA, Davidson BR, Iype S, Nasralla D, Imber C, Sharma D, Pissanou T, Pollok JM. Donor-Recipient Body Surface Area Mismatch and the Outcome of Liver Transplantation in the UK. Prog Transplant. 2023; 33:61-68.
24. Kubal CA, Mihaylov P, Fridell J, Cabrales A, Nikumbh T, Timsina L, Soma D, Ekser B, Mangus R. Donor-recipient body size mismatch has no impact on outcomes after deceased donor whole liver transplantation: Role of donor liver size measurement. Clin Transplant. 2021; 35:e14299.
25. Addeo P, Bachellier P, Noblet V. Combination of Donor Anthropometrics With Recipient Imaging to Improve Matching in Liver Transplantation. Liver Transpl. 2022; 28:512-513.
26. Rustgi VK, Marino G, Halpern MT, Johnson LB, Umana WO, Tolleris C. Role of gender and race mismatch and graft failure in patients undergoing liver transplantation. Liver Transpl. 2002; 8:514-518.
27. Schoening WN, Helbig M, Buescher N, Andreou A, Bahra M, Schmitz V, Pascher A, Pratschke J, Seehofer D. Gender Matches in Liver Transplant Allocation: Matched and Mismatched Male-Female Donor-Recipient Combinations; Long-term Follow-up of More Than 2000 Patients at a Single Center. Exp Clin Transplant. 2016; 14:184-90.
28. Lehner F, Becker T, Klempnauer J, Borlak J. Gender-incompatible liver transplantation is not a risk factor for patient survival. Liver Int. 2009; 29:196-202.
29. Lai Q, Giovanardi F, Melandro F, Larghi Laureiro Z, Merli M, Lattanzi B, Hassan R, Rossi M, Mennini G. Donor-to-recipient gender match in liver transplantation: A systematic review and meta-analysis. World J Gastroenterol. 2018; 24:2203-2210.
30. Germani G, Zeni N, Zanetto A, *et al*. Influence of donor and recipient gender on liver transplantation outcomes in Europe. Liver Int. 2020; 40:19611971.
31. Silveira F, Silveira FP, Freitas ACT, Coelho JCU, Ramos EJB, Macri MM, Tefilli N, Bredt LC. Liver transplantation: survival and indexes of donor-recipient matching. Rev Assoc Med Bras (1992). 2021; 67:690-695.
32. Feng S, Goodrich NP, Bragg-Gresham JL, Dykstra DM, Punch JD, DebRoy MA, Greenstein SM, Merion RM. Characteristics associated with liver graft failure: the concept of a donor risk index. Am J Transplant. 2006; 6:783-790.
33. Dutkowski P, Oberkofler CE, Slankamenac K, Puhan MA, Schadde E, Müllhaupt B, Geier A, Clavien PA. Are there better guidelines for allocation in liver transplantation? A novel score targeting justice and utility in the model for end-stage liver disease era. Ann Surg. 2011; 254:745-53; discussion 753.
34. Schlegel A, Linecker M, Kron P, Györi G, De Oliveira ML, Müllhaupt B, Clavien PA, Dutkowski P. Risk Assessment in High- and Low-MELD Liver Transplantation. Am J Transplant. 2017; 17:1050-1063.
35. de Campos Junior ID, Stucchi RS, Udo EY, Boin Ide F. Application of the BAR score as a predictor of short- and long-term survival in liver transplantation patients. Hepatol Int. 2015; 9:113-119.
36. López IP, Alonso AP, Carroll NZ, Villalba JS, Gómez EB, Jiménez AP, Herrera TV, Moral JMVD. Improving Matching Between Recipient and Donor in Liver Transplantation: Are Scores Able to Predict Long-term Survival? Transplant Proc. 2022; 54:41-44.
37. Schoening W, Helbig M, Buescher N, Andreou A, Schmitz V, Bahra M, Puhl G, Pascher A, Pratschke J, Seehofer D. Eurotransplant donor-risk-index and recipient factors: influence on long-term outcome after liver transplantation - A large single-center experience. Clin Transplant. 2016; 30:508-517.
38. Boecker J, Czigany Z, Bednarsch J, Amygdalos I, Meister F, Santana DAM, Liu WJ, Strnad P, Neumann UP, Lurje G. Potential value and limitations of different clinical scoring systems in the assessment of short- and long-term outcome following orthotopic liver transplantation. PLoS One. 2019; 14:e0214221.
39. Cillo U, Burra P, Mazzaferro V, Belli L, Pinna AD, Spada M, Nanni Costa A, Toniutto P; I-BELT (Italian Board of Experts in the Field of Liver Transplantation). A Multistep, Consensus-Based Approach to Organ Allocation in Liver Transplantation: Toward a "Blended Principle Model". Am J Transplant. 2015; 15:2552-61.
40. Khouzam S, Pagano D, Barbàra M, Cintorino D, Li Petri S, di Francesco F, Ricotta C, Bonsignore P, Seidita A, Calamia S, Canzonieri M, Tropea A, Gruttadauria S. Impact of Italian Score for Organ Allocation System on Deceased Donor Liver Transplantation: A Monocentric Competing Risk Time-to-Event Analysis. Transplant Proc. 2019; 51:2860-2864.
41. Rana A, Hardy MA, Halazun KJ, Woodland DC, Ratner LE, Samstein B, Guarrera JV, Brown RS Jr, Emond JC. Survival outcomes following liver transplantation (SOFT) score: a novel method to predict patient survival following liver transplantation. Am J Transplant. 2008; 8:2537-2546.
42. Halldorson JB, Bakthavatsalam R, Fix O, Reyes JD, Perkins JD. D-MELD, a simple predictor of post liver transplant mortality for optimization of donor/recipient matching. Am J Transplant. 2009; 9:318-326.
43. Calleja Lozano R, Hervás Martínez C, Briceño Delgado FJ. Crossroads in Liver Transplantation: Is Artificial Intelligence the Key to Donor-Recipient Matching? Medicina (Kaunas). 2022; 58:1743.
44. Veerankutty FH, Jayan G, Yadav MK, Manoj KS, Yadav A, Nair SRS, Shabeerali TU, Yeldho V, Sasidharan M, Rather SA. Artificial Intelligence in hepatology, liver surgery and transplantation: Emerging applications and frontiers of research. World J Hepatol. 2021; 13:1977-1990.
45. Briceño J, Calleja R, Hervás C. Artificial intelligence and liver transplantation: Looking for the best donor-recipient pairing. Hepatobiliary Pancreat Dis Int. 2022; 21:347-353.
46. Börner N, Schoenberg MB, Pöschke P, Heiliger C, Jacob

S, Koch D, Pöllmann B, Drefs M, Koliogiannis D, Böhm C, Karcz KW, Werner J, Guba M. A Novel Deep Learning Model as a Donor-Recipient Matching Tool to Predict Survival after Liver Transplantation. J Clin Med. 2022; 11:6422.
47. Briceño J, Cruz-Ramírez M, Prieto M, *et al*. Use of artificial intelligence as an innovative donor-recipient matching model for liver transplantation: results from a multicenter Spanish study. J Hepatol. 2014; 61:1020-1028.
48. Ayllón MD, Ciria R, Cruz-Ramírez M, Pérez-Ortiz M, Gómez I, Valente R, O'Grady J, de la Mata M, Hervás-Martínez C, Heaton ND, Briceño J. Validation of artificial neural networks as a methodology for donor-recipient matching for liver transplantation. Liver Transpl. 2018; 24:192-203.
49. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019; 17:195.
50. Guijo-Rubio D, Briceño J, Gutiérrez PA, Ayllón MD, Ciria R, Hervás-Martínez C. Statistical methods versus machine learning techniques for donor-recipient matching in liver transplantation. PLoS One. 2021; 16:e0252068.

Received April 3, 2023; Revised June 6, 2023; Accepted June 13, 2023.

**Address correspondence to:*
Salvatore Gruttadauria, IRCCS-ISMETT, University of Pittsburgh Medical Center (UPMC), 90127 Palermo, Italy.
E-mail: sgruttadauria@ismett.edu

Released online in J-STAGE as advance publication June 22, 2023.

Original Article

DOI: 10.5582/bst.2023.01128

The ability of Segmenting Anything Model (SAM) to segment ultrasound images

Fang Chen[1,2,*], Lingyu Chen[1,2], Haojie Han[1,2], Sainan Zhang[1,2], Daoqiang Zhang[1,2], Hongen Liao[3]

[1] Key Laboratory of Brain-Machine Intelligence Technology, Ministry of Education, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China;
[2] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China;
[3] Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China.

SUMMARY Accurate ultrasound (US) image segmentation is important for disease screening, diagnosis, and prognosis assessment. However, US images typically have shadow artifacts and ambiguous boundaries that affect US segmentation. Recently, Segmenting Anything Model (SAM) from Meta AI has demonstrated remarkable potential in a wide range of applications. The purpose of this paper was to conduct an initial evaluation of the ability for SAM to segment US images, particularly in the event of shadow artifacts and ambiguous boundaries. We evaluated SAM's performance on three US datasets of different tissues, including multi-structure cardiac tissue, thyroid nodules, and the fetal head. Results indicated that SAM generally performs well with US images with clear tissue structures, but it has limited performance in the event of shadow artifacts and ambiguous boundaries. Thus, creating an improved SAM that considers the characteristics of US images is significant for automatic and accurate US segmentation.

Keywords Segment Anything Model, ultrasound images, shadow artifacts

1. Introduction

Automatic segmentation of ultrasound (US) images can help disease screening, diagnosis, and assessment of prognosis. However, accurate US segmentation is a challenge due to the following difficulties. First, US images often suffer from a low signal-to-noise ratio (SNR) (*1*) and inhomogeneous intensity distribution (*2*). Second, shadows are a common occurrence due to inadequate contact between the US probe and the body surface or the presence of anatomical structures that interfere with the scanned tissue interfaces (*3*). These shadow regions, with their low intensity or dark pixels, are often integral to anatomical areas and lesions (*4*). As shown in Figure 1, shadow artifacts and ambiguous lesion boundaries are often observed in US images, posing significant challenges to accurate US segmentation.

Recently, the Segment Anything Model (SAM) (*5*) from Meta AI has been proposed as a promotable foundational model for natural image segmentation with minimal human intervention. SAM is a deep learning model (transformer-based) that has been trained on a huge number of images and masks - more than 1 billion masks in 11 million images. SAM is driven by various segmentation prompts (*e.g*., points, boxes, masks) to achieve zero-shot image segmentation. Due to its promising performance in several computer vision benchmarks, SAM has garnered a great deal of attention for use in medical image segmentation tasks (*6-9*). Specifically, Deng *et al*. (*6*) conducted experiments with SAM for tumor, non-tumor tissue, and cell nuclei segmentation, and empirical results indicated that SAM is amenable to the tasks of segmenting large connected objects. He *et al*. (*7*) evaluated more than 12 medical image segmentation datasets that used 5 imaging modalities (2D X-ray, histology, endoscopy, *etc*.) and that include different organs such as the brain, chest, lungs, and skin (*8*) in an attempt to validate the out-of-the-box zero-shot capabilities of SAM with an abdominal CT organ segmentation dataset, and they examined multiple scenarios, such as marking multiple points or boxes as prompts to obtain segmentation accuracy. Hu *et al*. (*9*) concluded that the more prompts were made, the more precise segmentation results were obtained by analyzing liver tumor segmentation for contrast-enhanced computed tomography volumes.

Although the aforementioned studies investigated

SAM's performance on medical images, they lack the comprehensive and in-depth assessment of SAM's performance on US images with shadow artifacts and inhomogeneous intensity distribution. The current study evaluated SAM's performance on three US datasets of different tissues in order to perform a comprehensive analysis of SAM's performance on US images. The hope is that this study can provide the community with some insights into the future development of an improved SAM for US image segmentation.

2. Materials and Methods

2.1. Overview

Figure 2 depicts the testing pipeline for SAM as applied to various US images in this study. SAM has three main components: an image encoder, a prompt encoder, and a mask decoder. The image encoder uses the Vision Transformer (ViT) (*10*) as its backbone and is pre-trained using the masked strategy from the masked autoencoder (MAE) (*11*). Its role is to provide the embedding of the input tensor so that it can be combined with the embedding of manual prompts in subsequent steps. The prompt encoder handles various types of sparse (multiple points, boxes, or texts) and dense (masks) prompts *via* distinct branches comprising a basic convolutional neural network. Ultimately, the mask decoder uses all embedding to determine the segmentation labels.

During the testing phase, SAM was comprehensively compared to related deep segmentation models using three different US datasets pertaining to various tissues, including multi-structure cardiac tissue, thyroid nodules, and the fetal head. Moreover, the testing involving US images was divided into two sets, the first consisting of images with shadow artifacts and the second consisting of clean images without any obvious shadow artifacts. This division enabled evaluation of SAM's effectiveness on US images with shadow artifacts. Moreover, four different methods of prompt selection were attempted and a regular grid of foreground points was used as prompts to generate US image segmentation results.

Positive and negative prompt points exist, indicating foreground or background points, respectively. To ensure experimental and model reproducibility, randomness, and accuracy, prompt points were chosen using the following three methods: (*i*) SAM-MPP: foreground points from the GT mask were randomly selected to serve as positive prompt points, with a range of 1-10 points; (*ii*) SAM-MPN: a background point was randomly marked as a negative point and multiple positive points were marked; (*iii*) SAM-CP: the central point of the image was identified and whether it is a positive or negative prompt

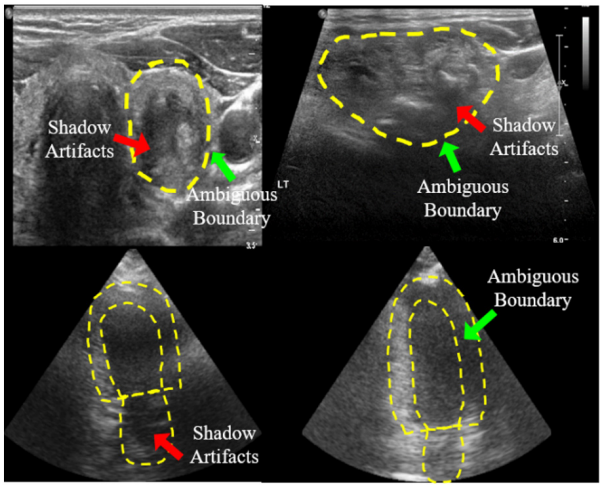

Figure 1. Difficultly of segmenting US images that contain shadow artifacts or ambiguous boundaries.

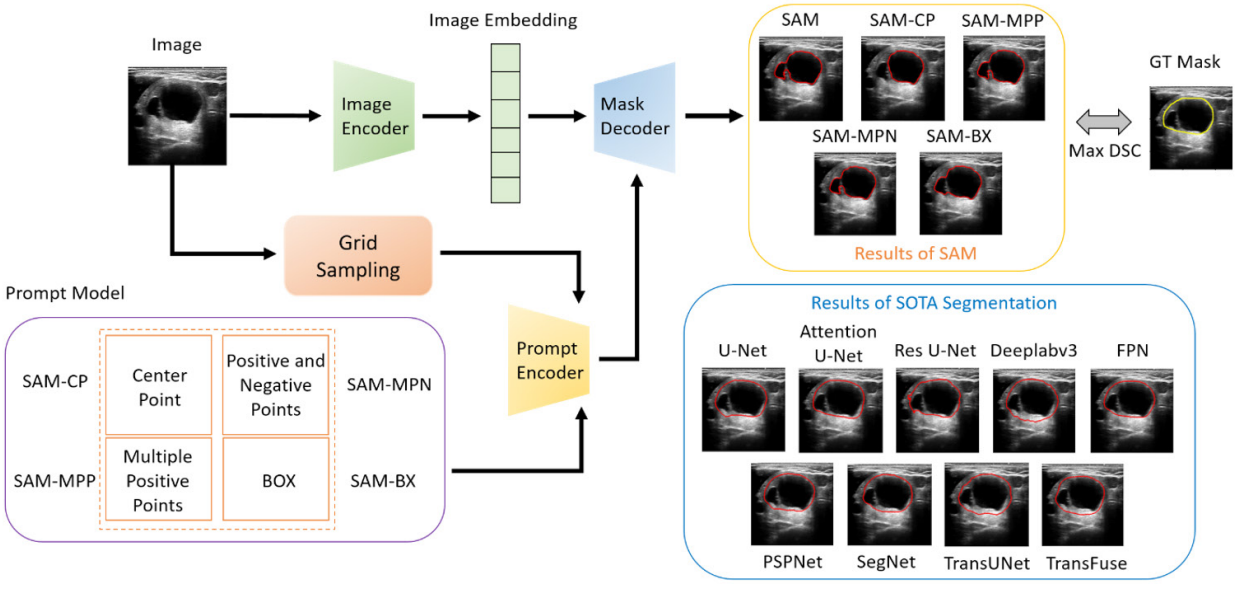

Figure 2. Testing pipeline for SAM in this study.

was determined using the GT mask. Additionally, (*iv*) SAM-BX: the bounding box of the GT mask was directly acquired without further steps. To evaluate SAM's segmentation performance, regular grid sampling was used to predict multiple masks per image and the highest quality mask was selected as the final segmentation result after comparison.

2.2. US datasets

SAM's performance in the US segmentation task was evaluated using three classic publicly available datasets. These datasets contain different scenarios of cardiac ultrasound, thyroid, and fetal head. The first dataset, CAMUS (*12*), is a large, fully annotated 2D echocardiographic assessment dataset collected from 500 patients and it includes manual annotations of the left ventricular endocardium (LV(Endo)), myocardium (LV(Epi)), and left atrium (LA) by experts. The second dataset, TN-SCUI (*13*), from the MICCAI 2020 Challenge, presents a challenging US segmentation task due to the various shapes of thyroid nodules, missing areas of lesions, ambiguous boundaries, and artifacts due to the way US is imaged. This dataset contains 3644 thyroid nodules from 3,644 patients that were manually annotated by experienced radiologists. The third dataset, HC18 (*14*), is a fetal US dataset consisting of 1,334 images used to measure the fetal head circumference.

2.3. Evaluation metrics

This study used four evaluation metrics to assess the performance of the different segmentation methods: the Dice Similarity Coefficient (DSC), Jaccard index (JI), Hausdorff Distance (HD), and Average Surface Distance (ASD).

Dice Similarity Coefficient (DSC, %) (*15*): This measures the similarity between the prediction and ground-truth sets, with a value range of [0,1]. A higher value indicates better model performance and it is often used to calculate the similarity of closed regions.

Jaccard index (JI, %) (*16*): This measures the ratio between the intersection and union of a category prediction and the ground-truth using fuzzy set theory.

Average Surface Distance (ASD, pixel) (*17*): This measures the average surface distance from all points of the prediction to the ground-truth, which assesses the surface variation between the segmentation and the GT.

Hausdorff Distance (HD, pixel) (*18*): This calculates the distance between the two sets of the prediction and ground-truth, with smaller values indicating higher similarity between the two sets. It is more sensitive to boundaries than DSC.

2.4. Methods of comparison

This study compared segmentation by SAM to several classic segmentation methods including U-Net, Attention U-Net, ResU-Net, DeepLabV3, PSPNet, SegNet, FPN, TransUnet, and TransFuse.

U-Net (*19*): U-Net is a U-shaped structure that uses skip connections to capture contextual information.

Attention U-Net (*20*): To address the problem that many redundant underlying features are extracted due to U-Net skip connections, Attention U-Net adds an attention module in skip connections to effectively suppress activations in irrelevant regions, thereby reducing the number of redundant features.

ResU-Net (*21*): ResUNet is a deep learning model based on residual connectivity for image segmentation tasks. It combines the advantages of ResNet and U-Net to better solve the problems of gradient disappearance and missing semantic information.

DeepLabV3 (*22*): DeeplabV3 provides the ability to arbitrarily control the resolution of features extracted by the encoder, with the encoder section having a large number of hole convolutions to balance accuracy and time consumption without loss of information.

PSPNet (*23*): PSPNet is based on FCN with a global mean pooling operation and feature fusion to obtain more contextual information. The features have a pyramid structure, so it is also called pyramid pooling.

SegNet (*24*): The decoder structure of SegNet performs non-linear upsampling using the pooling index computed in the maximum pooling step of the corresponding encoder. This reduces the number of parameters and operations compared to deconvolution and eliminates the need to learn upsampling.

FPN (*25*): The FPN algorithm uses both the high resolution of the lower-layer features and the semantic information of the higher-layer features to achieve prediction by fusing these different feature layers. In addition, the prediction is performed on each fused feature layer separately.

TransUnet (*26*): This is the first time that transformer was used as a promising alternative for medical image segmentation, and it has the merits of both transformers and U-Net.

TransFuse (*27*): TransFuse fuses a transformer and CNN to achieve long-range dependency modeling and reduce computational redundancy.

2.5. Implementation details

This study divided three publicly available datasets into training and testing sets in a 4:1 ratio. The training and testing datasets for each dataset are shown in Table 1.

Table 1. Number of training and testing images for TN-SCUI, CAMUS, and HC18

Dataset	TN-SCUI	CAMUS	HC18
Training Dataset Size	2,916	1,440	800
Testing Dataset Size	728	360	199

Due to the availability of training weights for SAM, this study is consistent with related studies on SAM (*6,28-29*) to evaluate SAM's performance on ultrasound images using the testing sets. Inspired by studies that compared SAM to related segmentation methods (*30-31*), segmentation models for comparison were first trained on the training set, and then corresponding segmentation metrics were obtained using the testing set.

3. Results

3.1. Segmentation results for US images overall

Quantitative comparison: Tables 2, 3, and 4 show the quantitative results of different segmentation methods in terms of the four evaluation metrics using different US datasets. In this experiment, the US images for testing were selected based on the same data partitioning of the public dataset. Results indicated that:

(1) Interestingly, SAM and SAM-CP performed poorly on the TN-SCUI dataset in terms of segmentation performance, possibly due to the unique image characteristics of the US dataset since it contains shadow artifacts and missing or unclear boundaries, hampering the model's ability to differentiate between foreground and background. However, SAM-MPP and SAM-MPN had substantially improved segmentation performance by incorporating manually labeled point prompts, achieving comparable or even better results than ResU-Net. Moreover, SAM-BX performed well on all four evaluation metrics, surpassing popular fully supervised segmentation models such as Deeplabv3, FPN, and SegNet, indicating that SAM is more effective at segmenting large connected areas. SAM-BX also benefited from the effective prompt of the bounding box, allowing the model to focus only on the box region and achieve a higher accuracy. However, providing box prompts may be time-consuming in real clinical scenarios, so the focus of this study is primarily on the segmentation results of the baseline SAM.

(2) Experiments on three different anatomical structures were conducted using CAMUS and Table 3 shows a comparative analysis of the results obtained. Results indicate that the best DSC obtained by SAM was 0.617 for the LV (Endo), 0.380 for the LV (Epi),

Table 2. Comparison to seven state-of-the-art fully-supervised methods using the TN-SCUI dataset

Items	DSC	JI	ASD	HD
U-Net	0.846	0.764	16.177	4.423
Attention U-Net	0.827	0.745	15.433	4.644
ResU-Net	0.696	0.572	44.865	10.661
Deeplabv3	0.861	0.778	13.256	4.028
FPN	0.869	0.792	12.049	4.029
PSPNet	0.858	0.780	23.516	6.183
SegNet	0.867	0.791	11.186	3.483
TransUNet	0.787	0.683	14.746	4.623
TransFuse	0.802	0.703	19.237	5.725
SAM	0.195	0.118	106.075	31.253
SAM-CP	0.345	0.257	76.281	28.056
SAM-MPP	0.716	0.603	28.419	9.222
SAM-MPN	0.721	0.605	28.536	9.029
SAM-BX	0.889	0.805	9.155	2.873

Table 4. Comparison to seven state-of-the-art fully-supervised methods using the HC18 dataset

Items	DSC	JI	ASD	HD
U-Net	0.979	0.958	3.984	1.411
Attention U-Net	0.978	0.957	4.310	1.504
ResU-Net	0.964	0.933	21.059	3.645
Deeplabv3	0.980	0.960	3.517	1.338
FPN	0.979	0.959	3.320	1.327
PSPNet	0.979	0.959	3.230	1.326
SegNet	0.980	0.960	3.108	1.304
TransUNet	0.966	0.936	6.059	2.105
TransFuse	0.974	0.950	3.654	1.532
SAM	0.539	0.380	58.622	17.771
SAM-CP	0.820	0.709	23.516	8.473
SAM-MPP	0.856	0.760	19.742	6.751
SAM-MPN	0.860	0.764	21.130	7.197
SAM-BX	0.951	0.908	8.330	2.601

Table 3. Comparison to seven state-of-the-art fully-supervised methods using the CAMUS dataset

Items	LV (Endo)				LV (Epi)				LA			
	DSC	JI	ASD	HD	DSC	JI	ASD	HD	DSC	JI	ASD	HD
U-Net	0.929	0.875	63.511	21.203	0.866	0.769	64.539	11.791	0.888	0.818	135.457	28.424
Attention U-Net	0.930	0.877	63.795	21.246	0.863	0.766	65.042	11.851	0.884	0.814	135.666	28.557
ResU-Net	0.920	0.856	68.042	22.057	0.834	0.720	67.437	13.167	0.858	0.771	135.818	29.149
Deeplabv3	0.936	0.883	64.414	21.291	0.873	0.778	65.703	11.841	0.901	0.828	135.437	28.454
FPN	0.934	0.882	65.458	21.506	0.873	0.780	66.679	11.937	0.905	0.835	136.691	28.515
PSPNet	0.937	0.886	64.289	21.165	0.875	0.783	65.515	11.709	0.899	0.932	135.567	28.321
SegNet	0.934	0.880	64.319	21.476	0.868	0.771	65.552	11.923	0.891	0.822	136.105	28.553
TransUNet	0.914	0.851	63.276	21.450	0.831	0.720	64.811	11.835	0.879	0.797	135.278	28.670
TransFuse	0.915	0.850	65.394	21.991	0.832	0.720	66.766	12.255	0.876	0.799	136.011	28.856
SAM	0.241	0.140	89.938	35.827	0.233	0.133	90.724	34.634	0.140	0.076	152.234	61.710
SAM-CP	0.555	0.413	49.529	12.554	0.256	0.148	53.860	17.010	0.161	0.089	128.767	37.364
SAM-MPP	0.595	0.452	49.035	11.639	0.280	0.165	77.102	28.331	0.428	0.319	75.077	22.800
SAM-MPN	0.617	0.470	46.164	11.182	0.290	0.172	72.311	26.111	0.477	0.356	57.552	17.158
SAM-BX	0.600	0.440	26.224	8.093	0.380	0.238	31.484	9.699	0.867	0.770	8.813	2.907

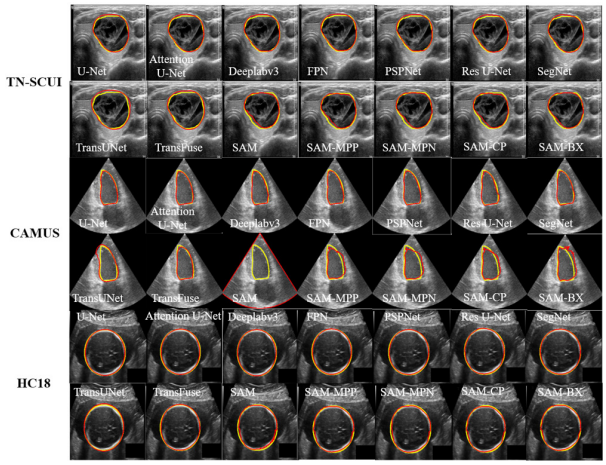

Figure 3. Good examples of segmentation on three US datasets.

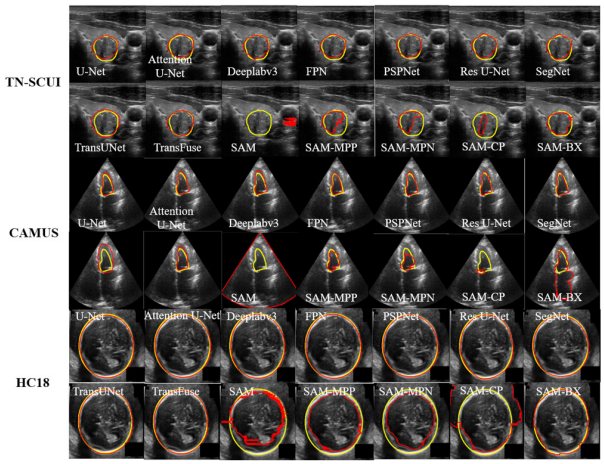

Figure 4. Bad examples of segmentation on three US datasets.

and 0.867 for the LA, which are significantly lower than the best DSC results obtained using representative deep networks (0.937, 0.875, and 0.905). This difference can be attributed to several reasons. First, the heart has complex and multiple structures that differ greatly from the shapes of thyroid nodules and the fetal head. Second, CAMUS images contain a large number of shadow artifacts and ambiguous boundaries, hampering SAM's ability to accurately distinguish between foreground and background regions. Lastly, the segmentation performance of the LV (Epi) was substantially lower compared to the LV (Endo) and LA, which may be due to interference from the LV (Endo) structure surrounding the LV (Epi).

(3) The results for HC18 are shown in Table 4, where all four types of SAM-based models, except for SAM, achieved DSC scores exceeding 0.8, and the accuracy of SAM-BX reached 0.95. This excellent performance can be attributed to the fact that the target anatomical regions in the HC18 dataset have large connected regions, which allows the additional prompts provided by the models to help achieve accurate segmentation. However, there is a noticeable gap between the ASD and HD obtained by SAM and classic fully supervised segmentation models, indicating that SAM has a weakness in fine segmentation tasks.

To summarize, the segmentation performance of basic SAM needs to be improved according to all three US datasets. This could be the result of the unique image characteristics of US, such as shadow artifacts and ambiguous or missing boundaries, posing significant challenges for SAM in identifying foreground and background regions accurately. In addition, various prompt methods were used, and results indicated that using the bounding box of the GT mask is the most effective solution. However, this approach is also very stringent and restricts the analysis and clinical application of medical images. A series of SAM models displayed better segmentation performance in areas with large connectivity and regularity but struggled with complex anatomical structures.

Qualitative comparison: Figure 3 shows some good examples of segmentation achieved by SAM, while Figure 4 displays examples where SAM failed to segment the target areas accurately. Figure 3 indicates that SAM performs as well as or better than other methods in instances where the tissue or anatomical structures are relatively distinct. However, SAM struggles with segmenting complex structures such as thyroid nodules or cardiac structures, which may be obscured by shadow artifacts or have ambiguous or missing boundaries. As shown in Figure 4, this results in low accuracy for SAM-based models, with a significant performance gap compared to popular deep models.

3.2. Segmentation results for US images with and without shadow artifacts

Tables 2-4 indicate that SAM consistently produces the worst segmentation results compared to other popular segmentation models using the TN-SCUI and CAMUS datasets. To investigate whether the presence of shadow artifacts in US images of the TN-SCUI dataset affected SAM's performance, the dataset was divided into two groups: US samples with shadow artifacts and clear US images without obvious shadow artifacts. With the guidance of a sonographer with five years of experience, segmentation results for US images with and without shadow artifacts were compared using different methods.

Quantitative comparison: As shown in Table 5, SAM methods differ most dramatically between the TN-SCUI with and without shadows. Especially with SAM, SAM-MPP, SAM-MPN, and SAM-CP, the DSC difference was 22.1% (SAM-MPN), along with maximum differences in the ASD of 18.882 and the HD of 4.919 (SAM-MPN). In addition, SAM-BX had an 8% DSC difference. The analysis of complete datasets identified the bounding box as the most

Table 5. Comparison to seven state-of-the-art fully-supervised methods using the SHADOW TN-SCUI dataset

Items	w/ Shadow Artifacts				w/o Shadow Artifacts			
	DSC	JI	ASD	HD	DSC	JI	ASD	HD
U-Net	0.828	0.734	17.663	4.920	0.865	0.796	14.605	3.897
Attention U-Net	0.805	0.713	16.736	5.305	0.851	0.780	14.086	3.962
ResU-Net	0.676	0.545	45.193	10.896	0.717	0.600	44.520	10.415
Deeplabv3	0.830	0.735	15.322	4.899	0.894	0.825	11.084	3.113
FPN	0.840	0.751	14.181	5.061	0.901	0.835	9.805	2.944
PSPNet	0.830	0.740	27.350	7.978	0.889	0.823	19.499	4.301
SegNet	0.841	0.754	12.304	4.047	0.895	0.829	10.015	2.892
TransUNet	0.773	0.663	16.580	5.587	0.832	0.745	14.101	4.182
TransFuse	0.787	0.679	19.408	5.818	0.817	0.728	19.059	5.627
SAM	0.178	0.108	107.762	33.683	0.216	0.129	104.455	28.844
SAM-CP	0.281	0.196	83.101	30.453	0.412	0.322	69.102	25.534
SAM-MPP	0.633	0.502	36.132	12.103	0.820	0.720	20.527	6.285
SAM-MPN	0.616	0.482	36.832	12.195	0.837	0.740	17.950	5.254
SAM-BX	0.811	0.705	15.923	8.714	0.889	0.828	10.243	5.989

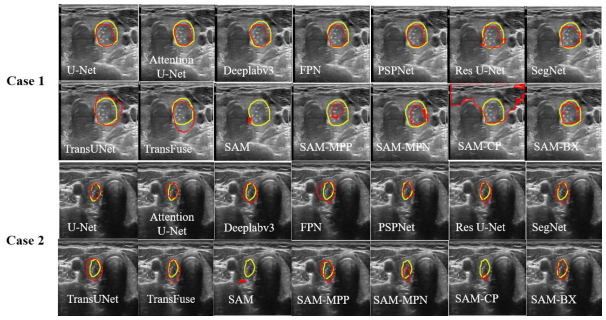

Figure 5. Failure to segment US images of the thyroid with shadow artifacts.

powerful prompt for improving SAM's segmentation performance. However, the significant difference in results between images with and without shadows persisted despite using a given GT mask's bounding box, indicating that the US dataset contains a significant number of shadow artifacts that can have a considerable impact on SAM's segmentation results.

Further analysis of other representative deep segmentation models revealed that their DSC difference did not exceed 6% (FPN) when tested on TN-SCUI with or without shadows. In addition, the differences in ASD and HD did not exceed 7.851 and 3.677, respectively (PSPNet). These findings suggest that the segmentation performance of deep models was relatively stable. Moreover, deep segmentation models had ASD and HD metrics that were generally better than those of SAM, indicating that additional refinement is necessary to enable SAM to better complete segmentation tasks for shadow artifacts and ambiguous boundaries present in US images.

Qualitative comparison: The segmentation results of different models using US images of the thyroid containing shadow artifacts are shown in Figure 5. Areas with large shadow artifacts are evident, so the segmentation results of SAM are not ideal. In areas with missing or ambiguous boundaries, segmentation results with SAMs were rough, and the corresponding ASD and HD were not as good.

4. Discussion and Conclusion

As a foundational model for image segmentation, SAM has shown great potential for natural images. This study completed an initial evaluation of SAM's ability to perform medical US image segmentation using three US image datasets of different organs, including multi-structure cardiac tissue, thyroid nodules, breast nodules, and the fetal head. This study particularly examined shadow artifacts in US images as a factor affecting SAM's accuracy. While we acknowledge the advancement of large foundational models for CV, the current experiments demonstrate that there is still room for improvement in SAM's performance on this specific task of medical US image segmentation.

Results of medical US image segmentation are compared in Tables 2-4, which indicate that everything mode is not suitable for most medical US datasets. In other words, SAM is not as accurate as dataset specific deep-learning algorithms (*19-27*) for medical US segmentation tasks. Therefore, applying the trained model from SAM directly to a medical US segmentation task will not result in satisfactory performance. In the future, more medical US images need to be used to fine-tune SAM to create a highly accurate benchmark mode.

Given that US imaging is a special imaging modality that commonly contains shadow artifacts, SAM's performance was compared on US images with and without shadow artifacts. Thus, a strength of this study is that it fully explored shadow artifacts as a factor that affect SAM's accuracy in US images. As shown by Table 5 and Figure 2, SAM had significant performance degradation with shadow artifacts. Hence, future research should investigate how to improve

SAM's reliability for US image segmentation with shadow artifacts. A possible solution (not explicitly addressed in the current work) is adding a shadow learning mechanism to SAM. Previous studies have proven that generating and injecting simulated shadows into US images and teaching them is helpful for US segmentation tasks (*32,33*). Another possible solution is using US images with shadow artifacts to finetune SAM. In summary, the current study has shown that additional work is needed to improve SAM's performance on this specific US segmentation task. The hope is that this study can provide the community with some insights into the future development of a improved SAM for US image segmentation.

Funding: This work was supported by the National Nature Science Foundation of China (62271246, U20A20389, 82027807, U22A2051), the National Key Research and Development Program of China (2022YFC2405200), the Beijing Hospital Authority's Special Funding Support for Development of Clinical Medicine (XMLX202141), and the China Postdoctoral Science Foundation (2020M671484).

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

References

1. Honarvar F, Sheikhzadeh H, Moles M, Sinclair AN. Improving the time-resolution and signal-to-noise ratio of ultrasonic NDE signals. Ultrasonics. 2004; 41:755-763.
2. Xiao G, Brady M, Noble JA, Zhang Y. Segmentation of ultrasound B-mode images with intensity inhomogeneity correction. IEEE Trans Med Imaging. 2002; 21:48-57.
3. Singla R, Ringstrom C, Hu R, Lessoway V, Reid J, Nguan C, Rohling R. Speckle and shadows: ultrasound-specific physics-based data augmentation applied to kidney segmentation. In: Medical Imaging with Deep Learning. 2022;1-10.
4. Noll M, Puhl J, Wesarg S. Achieving fluid detection by exploiting shadow detection methods. In: Imaging for Patient-Customized Simulations and Systems for Point-of-Care Ultrasound: International Workshops, BIVPCS 2017 and POCUS 2017; 121-128.
5. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo W-Y. Segment anything. arXiv:230402643. 2023;1-30. https://doi.org/10.48550/arXiv.2304.02643
6. Deng R, Cui C, Liu Q, Yao T, Remedios LW, Bao S, Landman BA, Wheless LE, Coburn LA, Wilson KT. Segment Anything Model (SAM) for digital pathology: Assess zero-shot segmentation on whole slide imaging. arXiv:230404155. 2023; 1-6. https://arxiv.org/pdf/2304.04155.pdf
7. He S, Bao R, Li J, Grant PE, Ou Y. Accuracy of Segment-Anything Model (SAM) in medical image segmentation tasks. arXiv:230409324. 2023;1-8. https://arxiv.org/pdf/2304.09324.pdf
8. Roy S, Wald T, Koehler G, Rokuss MR, Disch N, Holzschuh J, Zimmerer D, Maier-Hein KH. SAM. MD: Zero-shot medical image segmentation capabilities of the Segment Anything Model. arXiv:230405396. 2023;1-4. arXiv:230405396
9. Hu C, Li X. When SAM meets medical images: An investigation of Segment Anything Model (SAM) on multi-phase liver tumor segmentation. arXiv:230408506. 2023;1-5. https://doi.org/10.48550/arXiv.2304.08506
10. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:201011929. 2020;1-22. https://doi.org/10.48550/arXiv.2010.11929
11. He K, Chen X, Xie S, Li Y, Dollár P, Girshick R. Masked autoencoders are scalable vision learners. In: Proceedings IEEE/CVF Conference Comp Vision Pattern Recog. 2022; 16000-16009.
12. Leclerc S, Smistad E, Pedrosa J, Østvik A, Cervenansky F, Espinosa F, Espeland T, Berg EAR, Jodoin P-M, Grenier T. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE Trans Med Imaging. 2019; 38:2198-2210.
13. Gireesha H, Nanda S. Thyroid nodule segmentation and classification in ultrasound images. Internatl J Engineer Res Tech. 2014; 21-31.
14. van den Heuvel TL, de Bruijn D, de Korte CL, Ginneken Bv. Automated measurement of fetal head circumference using 2D ultrasound images. PloS One. 2018; 13:e0200412.
15. Bilic P, Christ P, Li HB, Vorontsov E, Ben-Cohen A, Kaissis G, Szeskin A, Jacobs C, Mamani GEH, Chartrand G. The liver tumor segmentation benchmark (LiTS). Med Image Anal. 2023; 84:102680;1-24.
16. Crum WR, Camara O, Hill DL. Generalized overlap measures for evaluation and validation in medical image analysis. IEEE Trans Med Imaging. 2006; 25:1451-1461.
17. Dubuisson M-P, Jain AK. A modified Hausdorff distance for object matching. In: Proceedings 12th Internatl Conference Pattern Recog. IEEE, 1994; 566-568.
18. Zhou D, Fang J, Song X, Guan C, Yin J, Dai Y, Yang R. IOU loss for 2d/3d object detection. In: 2019 International Conference on 3D Vision (3DV). IEEE, 2019; 85-94.
19. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, 2015; 234-241.
20. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B. Attention u-net: Learning where to look for the pancreas. arXiv:180403999. 2018;1-10. https://doi.org/10.48550/arXiv.1804.03999
21. Zhang Z, Liu Q, Wang Y. Road extraction by deep residual u-net. IEEE GeosciI Remote S. 2018; 15:749-753.
22. Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. arXiv:170605587. 2017;1-14. https://doi.org/10.48550/arXiv.1706.05587
23. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. In: Proceedings IEEE Conference Comp Vision Pattern Recog. 2017; 2881-2890.
24. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell. 2017; 39:2481-2495.

25. Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings IEEE Conference Comp Vision Pattern Recog. 2017; 2117-2125.
26. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, Lu L, Yuille AL, Zhou Y. Transunet: Transformers make strong encoders for medical image segmentation. arXiv:210204306. 2021;1-13. https://arxiv.org/pdf/2102.04306.pdf
27. Zhang Y, Liu H, Hu Q. Transfuse: Fusing transformers and CNNs for medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. Springer, 2021; 14-24.
28. Jie L, Zhang H. When SAM meets shadow detection. arXiv:230511513. 2023;1-6. https://doi.org/10.48550/arXiv.2305.11513
29. Zhou T, Zhang Y, Zhou Y, Wu Y, Gong C. Can sam segment polyps? arXiv:230407583. 2023;1-5. https://doi.org/10.48550/arXiv.2304.07583
30. Wang A, Islam M, Xu M, Zhang Y, Ren H. SAM meets robotic surgery: An empirical study in robustness perspective. arXiv:230414674. 2023;1-5. https://doi.org/10.48550/arXiv.2304.14674
31. Williams D, MacFarlane F, Britten A. Leaf Only SAM: A Segment Anything pipeline for zero-shot automated leaf segmentation. arXiv preprint arXiv:230509418. 2023;1-9.
32. Meng Q, Sinclair M, Zimmer V, Hou B, Rajchl M, Toussaint N, Oktay O, Schlemper J, Gomez A, Housden J. Weakly supervised estimation of shadow confidence maps in fetal ultrasound imaging. IEEE Trans Med Imaging. 2019; 38:2755-2767.
33 Yasutomi S, Arakaki T, Matsuoka R, Sakai A, Komatsu R, Shozu K, Dozen A, Machino H, Asada K, Kaneko S. Shadow estimation for ultrasound images using auto-encoding structures and synthetic shadows. Applied Sciences. 2021; 11:1127-1147.

Received June 2, 2023; Revised June 15, 2023; Accepted June 20, 2023.

*Address correspondence to:
Fang Chen, Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing211016, Jiangsu, China.
E-mail: chenfang@nuaa.edu.cn

Released online in J-STAGE as advance publication June 22, 2023.

Original Article

DOI: 10.5582/bst.2023.01116

EMG-FRNet: A feature reconstruction network for EMG irrelevant gesture recognition

Wenli Zhang[1,*], Yufei Wang[1], Jianyi Zhang[2], Gongpeng Pang[1]

[1] Faculty of Imformation Technology, Beijing University of Technology, Beijing, China;
[2] College of Art and Design, Beijing University of Technology, Beijing, China.

SUMMARY With the development of deep learning technology, gesture recognition based on surface electromyography (EMG) signals has shown broad application prospects in various human-computer interaction fields. Most current gesture recognition technologies can achieve high recognition accuracy on a wide range of gesture actions. However, in practical applications, gesture recognition based on surface EMG signals is susceptible to interference from irrelevant gesture movements, which affects the accuracy and security of the system. Therefore, it is crucial to design an irrelevant gesture recognition method. This paper introduces the GANomaly network from the field of image anomaly detection into surface EMG-based irrelevant gesture recognition. The network has a small feature reconstruction error for target samples and a large feature reconstruction error for irrelevant samples. By comparing the relationship between the feature reconstruction error and the predefined threshold, we can determine whether the input samples are from the target category or the irrelevant category. In order to improve the performance of EMG irrelevant gesture recognition, this paper proposes a feature reconstruction network named EMG-FRNet for EMG irrelevant gesture recognition. This network is based on GANomaly and incorporates structures such as channel cropping (CC), cross-layer encoding-decoding feature fusion (CLEDFF), and SE channel attention (SE). In this paper, Ninapro DB1, Ninapro DB5 and self-collected datasets were used to verify the performance of the proposed model. The Area Under the receiver operating characteristic Curve (AUC) values of EMG-FRNet on the above three datasets were 0.940, 0.926 and 0.962, respectively. Experimental results demonstrate that the proposed model achieves the highest accuracy among related research.

Keywords surface EMG signals, human-computer interaction, irrelevant gesture recognition, reconstruction error

1. Introduction

With the rapid development of artificial intelligence and deep learning technology, human-computer interaction through gesture recognition technology has gradually become a hot research topic nowadays. Gesture recognition based on surface electromyography (EMG) signals has advantages over vision-based gesture recognition, such as being less affected by changes in the external environmental background, requiring less computational effort, and offering higher real-time performance (*1-2*). The surface EMG signals are bioelectrical information obtained from the skin surface, which has the advantages of non-invasive, non-traumatic, and simple operation. The surface EMG signals directly reflect the state of muscle contraction that causes limb movements and contains rich motor information, which can realize the prediction of hand movements intention. In recent years, pattern recognition technology based on surface myoelectricity has shown promising applications in the field of human-computer interaction, such as intelligent prostheses (*3*), rehabilitation exoskeletons (*4*), sign language interpretation (*5*), *etc*.

In practical applications of EMG interaction, the recognition of the target gesture category is crucial, along with the need to mitigate various interferences, including irrelevant gestures, electrode displacement, muscle fatigue, and user variations (*6-7*). Among these interferences, the presence of irrelevant gestures is a common and significant concern. The target gestures refer to the user-defined hand gestures that are used to train the classifier and facilitate human-computer interaction. The irrelevant gestures refer to the unintentional hand gestures made by users during system usage that do not belong to the predefined target categories. In such cases, the classifier is forced

to select one of the trained motions, resulting in erroneous recognition results and compromising the safety of both the device and the user (*8-9*). In the field of target gesture recognition, numerous studies have achieved high accuracy rates across various hand gesture actions (*10-11*). In our recent research at our laboratory, Zhang *et al*. (*12*) proposed the LST-EMG-Net model, which further improves the accuracy of EMG-based gesture recognition. Therefore, the focus of this study lies in the recognition of irrelevant gestures. Existing methods in the field of irrelevant gesture recognition can be mainly categorized into probability-based approaches (*13-17*) and one-vs-all classification rule-based approaches (*18-21*).

Probability-based methods: The core idea of this type of method is to effectively differentiate target and irrelevant samples by comparing the classifier's predicted probability values for the test samples with a preset probability threshold. Specifically, the classifier calculates a predicted probability value for the test sample, and if the predicted probability value is higher than the preset threshold, the sample is classified as a target sample; otherwise, it is classified as an irrelevant sample. Scheme *et al*. (*13*) proposed a method based on linear discriminant analysis (LDA) that generates a confidence score for each decision, providing the ability to reject decisions with scores below a threshold. Robertson *et al*. (*14*) found limitations in the confidence features of LDA and used support vector machine (SVM) confidence scores to make correct decisions. Tomczynski *et al*. (*15*) used the entropy function output by an artificial neural network classifier as a loss function and as a criterion for accepting or rejecting gestures. Bao *et al*. (*16*) generated confidence scores based on the posterior probability of the CNN, estimating the probability of each output of the classifier is correct. Zhou *et al*. (*17*) proposed a two-layer classifier that combines Gaussian mixture model (GMM) and k-nearest neighbor (KNN) models. The classifier determines that a gesture is irrelevant when the output probabilities of both layers are below a predefined threshold. The probability-based method has the advantages of simple principle and low implementation cost. However, the classification probabilities of many target samples may be low, while those of irrelevant samples may be high. This ultimately affects the accuracy of irrelevant actions discrimination.

Methods based on one-vs-all classification rules: The core idea of this type of method is to train a one-class classifier for each target class, achieving effective discrimination between target and irrelevant samples. Specifically, the test sample is input into all classifiers to obtain binary classification results for each classifier, which are used to determine whether the test sample belongs to the corresponding target class of that classifier. If the test sample does not belong to any known target class, it is classified as irrelevant, otherwise, it is classified as a target sample. Ding *et al*. (*18*) used a set of classifiers composed of one-class Gaussian classifiers (GC) to determine whether the input sample belongs to the irrelevant class. The purpose of the Gaussian classifier is to fit a Gaussian distribution to samples belonging to the same target class. Ding *et al*. (*19*) used a set of classifiers composed of one-class support vector data description (SVDD) to exclude irrelevant motion interference. The purpose of SVDD is to find a minimum volume hyper-sphere to enclose samples belonging to the same target class. Wu *et al*. (*20-21*) used a set of classifiers composed of one-class autoencoder (AE) to address irrelevant motion interference. The purpose of AE is to reconstruct the input and judge whether the sample belongs to the target class based on the relationship between the reconstruction error and the threshold. Among the methods based on one-vs-all classification rules, some simple machine learning methods such as GC and SVDD are used to distinguish irrelevant gestures. However, these methods assume that the target gestures and irrelevant gestures are significantly different in the feature space, while the gestures in practical applications are indeed unpredictable. In contrast, the AE method can more fully exploit the small differences between the target and irrelevant gestures by calculating the reconstruction error, and improve the discriminative performance and stability of the model. However, this method is currently mostly used in high-density myoelectric systems, and the AE reconstruction process is easily affected by noise, leading to the limited reconstruction performance of the model.

The detection of irrelevant actions and anomaly detection solve very similar problems. Pang *et al*. (*22*) pointed out that anomaly detection, also known as outlier detection or novelty detection, refers to the process of detecting instances that deviate significantly from the majority of data. This has been a persistent and active research area for decades. The AE is a well-established method in the field of anomaly detection (*23-25*). It is trained on normal data and during testing, the reconstruction error of abnormal data is typically much larger than that of normal data. By measuring the magnitude of the reconstruction error, it becomes possible to identify anomalous samples. However, methods based on AE and AE variants are usually susceptible to data noise presented in the training data. With the development of GAN networks, GAN-based anomaly detection has rapidly become a popular deep anomaly detection method. Schlegl *et al*. (*26*) proposed the AnoGAN for anomaly detection in clinically-assisted retinal diseases. AnoGAN was the first paper to use GAN for anomaly detection. Zenati *et al*. (*27*) subsequently introduced EGBAD for anomaly detection on handwritten digit image datasets such as MNIST and network intrusion datasets like KDD99. Akcay *et al*. (*28*) proposed the GANomaly method for detecting dangerous items such as guns and knives in X-ray luggage or package datasets. GANomaly achieved state-

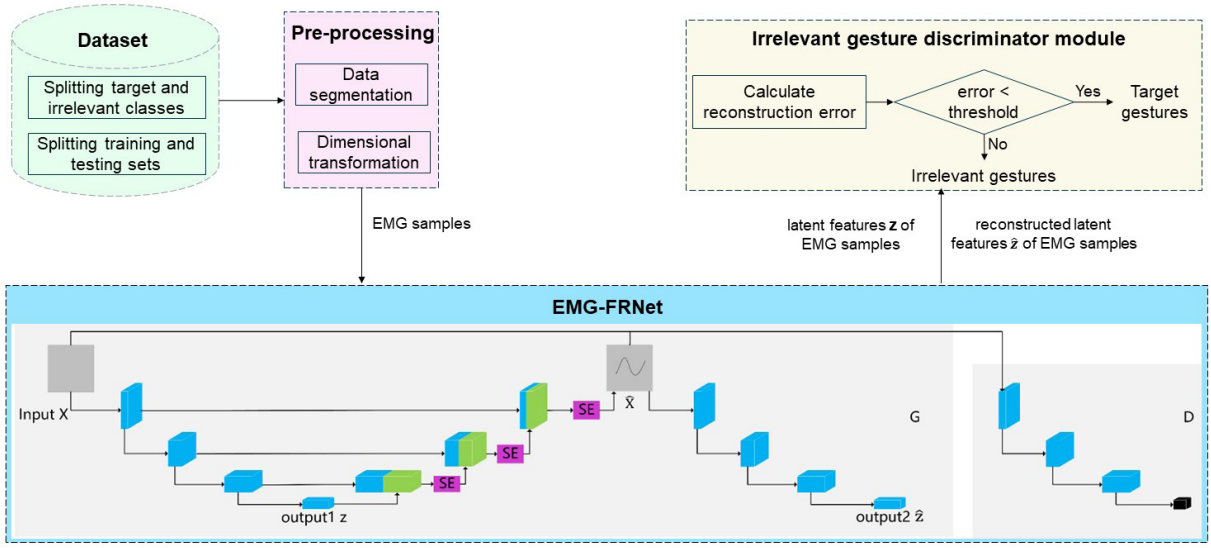

Figure 1. Overall flow chart of irrelevant gesture recognition method. EMG samples are obtained by pre-processing module; then latent features z and reconstructed latent features ẑ of EMG samples are obtained by EMG-FRNet module; finally, the reconstructed error between z and ẑ is calculated by irrelevant gesture discrimination module and compared with the threshold value to determine the class of EMG samples.

of-the-art performance in statistics and computation. Li *et al*. (*29-30*) implemented user authentication and improved system and device security based on the GANomaly anomaly detection method and multi-channel surface EMG signals of hand gestures. However, the detection performance of GANomaly in the field of EMG irrelevant gesture recognition still needs further exploration and improvement.

Based on the aforementioned issues, this paper proposes a feature reconstruction network named EMG-FRNet for EMG irrelevant gesture recognition. For the first time, we introduce GANomaly into EMG irrelevant gesture recognition. Building upon this, we incorporate additional structures such as channel cropping (CC), cross-layer encoding-decoding feature fusion (CLEDFF), and SE channel attention (SE) mechanisms. These enhancements contribute to the improved performance of irrelevant gesture recognition. The main innovative points are as follows:

1) To address the problem that the input of the original GANomaly is a three-channel RGB image, while this paper is a single-channel myoelectric input, the number of feature channels in the network layer is redundant. In this paper, we propose a CC method to optimize the number of channels in the network feature layer, which can significantly reduce the number of network parameters while improving the accuracy of the network.

2) To address the issue of spatial information loss during the downsampling process in the encoder of the original GANomaly generator, which affects the decoding performance, this paper proposes the use of CLEDFF. This method connects features of different scales from the encoding stage to the decoding stage, compensating for the lost information in the downsampling process of the encoder. As a result, the feature maps restored by the decoder contain more low-level semantic information, leading to better reconstruction quality.

3) To address the problem that the CLEDFF method may propagate a large amount of useless information and noise from the encoding process to the decoding layer, although it can achieve parallel transmission of feature information and improve the degree of information reuse, this paper proposes to use SE method. After the CLEDFF concatenates the feature channels of the encoding and decoding, the decoding stage introduces channel attention mechanism. This mechanism assigns different weights to each feature channel, capturing important feature information while suppressing unimportant channels. This approach helps improve the reconstruction accuracy.

2. Materials and Methods

The overall process of the proposed method for irrelevant gesture recognition is shown in Figure 1, which consists of three modules: data pre-processing module, feature reconstruction network EMG-FRNet module, and irrelevant gesture discrimination module. They are respectively introduced in sections 2.2, 2.3, and 2.4. First, the data pre-processing module is used to obtain EMG samples by processing the datasets with data segmentation and dimensional transformation, which provides the database for training and testing of the network model. Then, the EMG-FRNet module is used to extract the latent features of the input EMG samples and reconstruct the latent features of the EMG samples, and finally output the latent features z of the EMG samples and the reconstructed latent features ẑ. The module trains the network using datasets from the target category and

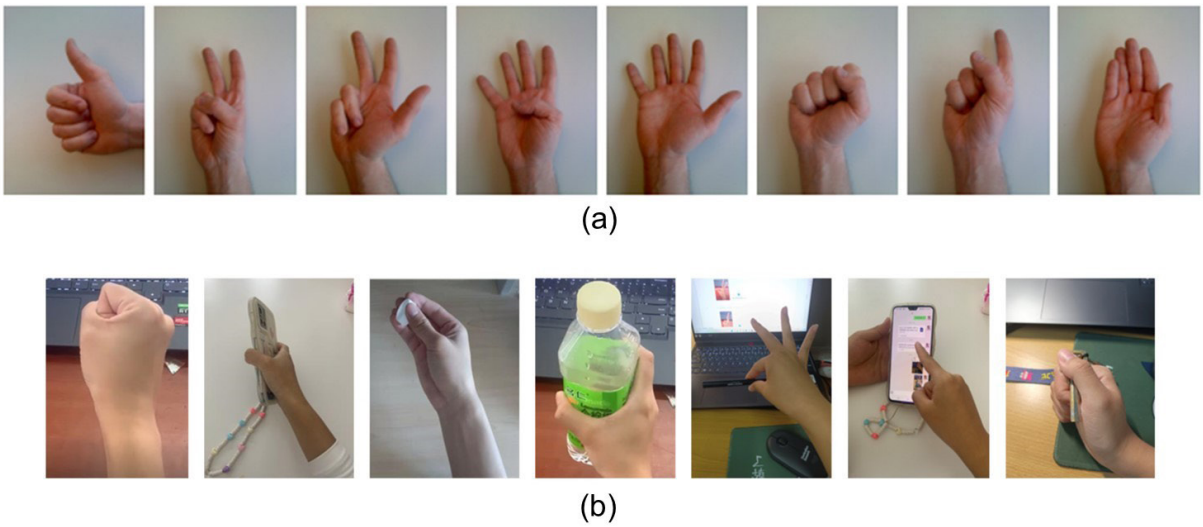

Figure 2. Types of gestures in the dataset used in this paper. (a) 8 gestures in the DB1/DB5 Exercise B dataset; **(b)** 7 gestures in the self-collected dataset.

tests the network using datasets from all categories. The network is tested with a small feature reconstruction error for the target category samples and a large feature reconstruction error for the irrelevant category samples. The feature reconstruction error is the difference between the latent feature z and the reconstructed latent feature $\hat{z}$. Finally, the irrelevant gesture discrimination module calculates the reconstruction error for z and $\hat{z}$ and compares it with a predefined threshold to determine whether the EMG samples belong to the target gestures or the irrelevant gestures.

2.1. Dataset

The datasets used in this paper for EMG-FRNet are the public datasets NinaproDB1, NinaproDB5, and the self-collected dataset. The self-collected dataset was obtained from stroke patients at Beijing Rehabilitation Hospital, Capital Medical University, and informed consent was obtained from all subjects in accordance with the Declaration of Helsinki, and ethics number 2022bkky-048 was obtained.

1) Public dataset DB1 (*31*): Eight basic hand postures of DB1 dataset Exercise B (as in Figure 2(a)) were used, and each gesture was acquired 10 times for a total of 10 healthy subjects. Its acquisition device was a 10-channel OttoBock 13E200 with a sampling frequency of 100 Hz. The equipment is manufactured by Ottobock, Germany. 2) Public dataset DB5 (*32*): Eight basic hand postures of DB5 dataset Exercise B (as in Figure 2(a)) were used, and each gesture was acquired six times for a total of 10 healthy subjects. Their acquisition devices were two Myo EMG bracelets, where each Myo bracelet had 8 channels and a sampling frequency of 200 Hz. The device is manufactured by Canadian company Thalmic Labs. 3) Self-collected dataset: 7 hand movements commonly used in life (as in Figure 2(b)) were used, and each gesture was acquired 6 times for a total of 6 subjects. The acquisition device was a Myo EMG bracelet with 8 channels and a sampling frequency of 200 Hz, manufactured by Thalmic Labs, Canada.

Target category and irrelevant category data division: experiments were set up with 1 gesture as the target category action and the remaining gestures as the irrelevant category actions, traversing all possible situations. Specifically, a total of 8 experiments were conducted per subject in DB1 and DB5, and 7 experiments were conducted per subject in the self-collected dataset.

Training set and test set data division: the training set has only target category data, and the test set has both target category and irrelevant category data. Specifically, in the DB1 dataset experiments, the 1st, 3rd, 4th, 6th, 8th, 9th, and 10th gesture repetitions of the target category are used to build the training set, and the 2nd, 5th, and 7th gesture repetitions of the target category and all gesture repetitions of the irrelevant category are used to build the test set. In DB5 and self-collected dataset experiments, the 1st, 3rd, 4th, and 6th gesture repetitions of the target category are used to construct the training set, and the 2nd and 5th repetitions of the target category and all gesture repetitions of the irrelevant category are used to construct the test set.

2.2. Data pre-processing

First, for all datasets in this paper, only eight channels of EMG data were used. The specific reasons are as follows: the most important muscle group for EMG gesture recognition is concentrated around the forearm brachioradialis muscle below the elbow. Commercially available eight-channel EMG bracelets are able to cover this part of the muscle. Additionally, the configuration of such bracelets is portable and has wide practical application prospects. Therefore, this data acquisition

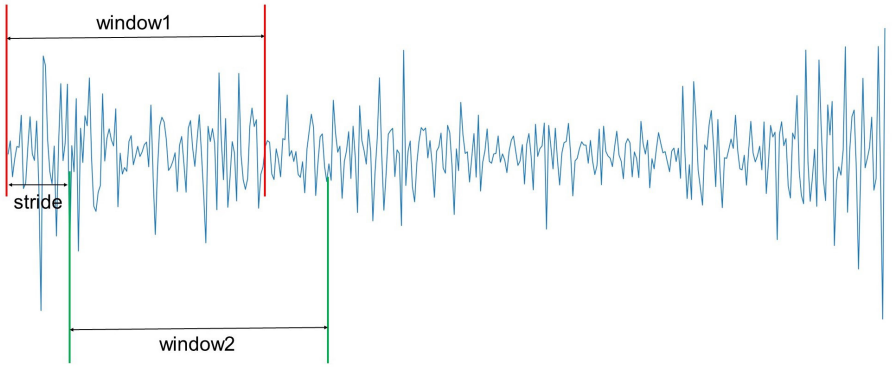

Figure 3. Schematic diagram of data segmentation. Windows1 and Windows2 represent myoelectric windows with a length of 128 sampling points, and stride represents a step size of 16 sampling points.

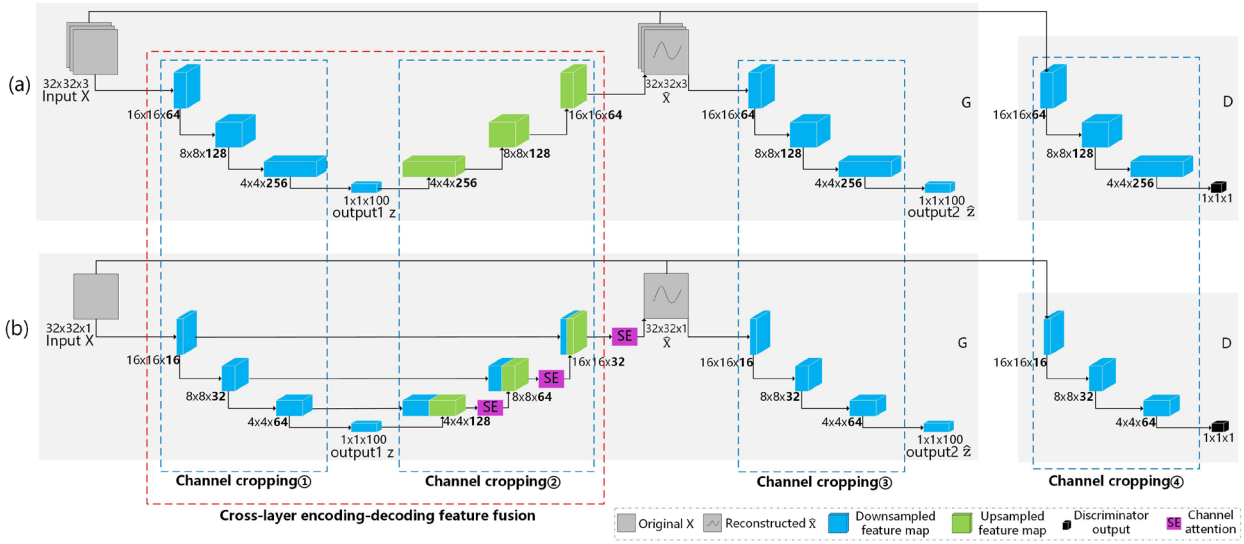

Figure 4. Model structure diagram. (a) Structure of the original GANomaly model; **(b)** Structure of the EMG-FRNet model.

scheme has become the first choice for many EMG gesture recognition studies.

Data segmentation: This study utilizes the sliding window method to segment the multi-channel EMG signals and obtain muscle activity samples of hand gestures. Specifically, as shown in Figure 3, we select windows of length 128 from the time series data and slide them with a fixed step size of 16. This means that multiple windows are extracted, with each window containing 128 consecutive time values. In this case, the data size is 1 × 128. Considering the presence of 8 channels, the data size of the EMG window samples is 8x128. This method can effectively segment the EMG signals to extract useful time-domain features and provide a database for subsequent model training and testing.

Data dimensional transformation: To enhance the feature learning in GAN networks, this study conducts a data dimensional transformation. Specifically, the segmented multichannel EMG samples are transformed from a 2D matrix of size 8 × 128 to a 2D matrix of size 32 × 32, which better matches the input format of the network.

2.3. Feature reconfiguration network EMG-FRNet model

The original GANomaly network architecture, as shown in Figure 4(a), consists of two main components: the generator G and the discriminator D. The generator network G is composed of an encoder G_{E1}, a decoder G_D, and another encoder G_{E2}, forming an "encoder-decoder-encoder" structure. Firstly, G_{E1} learns the latent features z of the input data X through downsampling. Then, G_D upsamples z to generate $\hat{X}$. $\hat{X}$ is further downsampled by G_{E2} to learn the feature representation $\hat{z}$ of $\hat{X}$. $\hat{z}$ is the reconstructed latent feature of the input data. G_{E2} employs the same network structure as G_{E1}. The main purpose of the discriminator network D is to distinguish between the real input data X and the reconstructed input data $\hat{X}$. During training, the parameters of the generator network G and the discriminator network D are alternately updated. During testing, the discriminator D is discarded. Ultimately, in the testing phase, the network obtains the latent features z and reconstructed latent

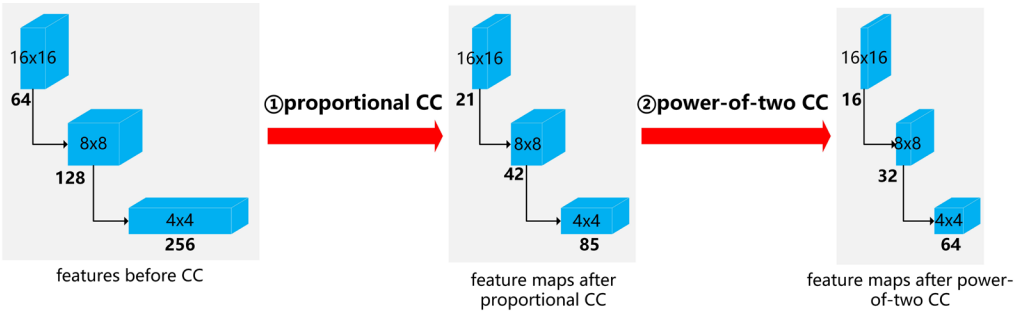

Figure 5. CC methods. Including proportional CC and power-of-two CC. Bold numbers in the figure represent the number of channels in the feature maps.

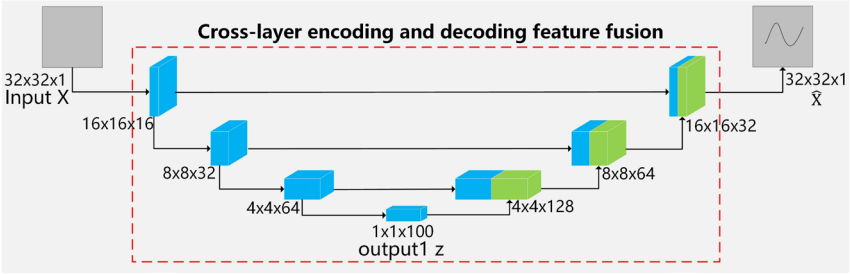

Figure 6. Schematic diagram of CLEDFF. Downsampled feature maps and upsampled feature maps for feature fusion.

features $\hat{z}$ of the input data.

The original GANomaly network has shown good performance in image anomaly detection, but its performance still needs to be improved when applied to EMG gesture recognition. Therefore, this paper proposes a feature reconstruction network, EMG-FRNet, for EMG gesture recognition. The model structure of EMG-FRNet is shown in Figure 4(b). It is an improvement on the original GANomaly network by adding CC, CLEDFF, and SE.

(1) CC: As shown in Figure 4(a), the original GANomaly network takes in three-channel RGB images, while this paper's input is single-channel EMG samples. The network contains a large number of redundant feature channels, which increases the number of parameters in the network, makes training more difficult, and affects the performance of the network.

In order to reduce the redundant feature channels of the original GANomaly network, we propose a CC method. This method consists of two parts: proportional CC and power-of-two CC. Proportional CC reduces the number of feature channels in the network according to the channel proportion of input samples. Power-of-two CC ensures that the number of channels after cropping can still be divisible by 2, to maximize the computer's processing capability. The CC method proposed in this paper is illustrated in Figure 5. In the proportional CC part, the feature maps of the original GANomaly network are cropped according to a 3:1 ratio, which is due to the channel ratio of three-channel RGB images and single-channel EMG samples. In the power-of-two CC part, the number of channels after cropping is required to be an integer power of two. Figure 5 shows the CC process of the downsampling feature maps. The number of channels in the downsampling feature layer is modified from (64, 128, 256) to (16, 32, 64). Similar modifications are made to the upsampling feature layer, with the number of channels changed from (256, 128, 64) to (64, 32, 16). Furthermore, the number of channels for the latent features z and $\hat{z}$ output by the network remains at 100, as shown in Figure 4. This configuration allows for more information to be contained in the latent features z and $\hat{z}$, which leads to more accurate reconstruction error calculation in the irrelevant gesture recognition module in Figure 1.

(2) CLEDFF: During the downsampling encoding process G_{E1} of the original GANomaly network from the input data X to the latent feature z, the original data is compressed continuously through the use of convolution and pooling operations. This compression can result in information loss, which limits the available feature information during the upsampling decoding process G_D from the latent feature z to the reconstructed data $\hat{X}$. As a result, the reconstruction performance of the generator is restricted and the recovery effect of the original data is affected.

To compensate for the information loss during the downsampling process, this paper proposes a CLEDFF method. As shown in Figure 6, the feature blocks of the encoder G_{E1} are visualized as blue, and the feature blocks of the decoder G_D are visualized as green. This method adds a series of skip connections between the encoder G_{E1} and decoder G_D of the generator G. These skip connections can directly transmit the original

high-resolution feature information to the decoder, allowing effective fusion of feature information during the encoding and decoding process, thus avoiding information loss caused by downsampling. The implementation steps of this method are as follows: in the decoding and recovery process from latent feature z to $\hat{X}$, first, z is upsampled by 2× to obtain the 4 × 4 × 64 feature block of decoder G_D. Secondly, the 4 × 4 × 64 feature blocks of encoder G_{E1} and decoder G_D are concatenated into a 4 × 4 × 128 feature block, which is then upsampled by 2× to obtain the 8 × 8 × 32 feature block of decoder G_D. Then, the 8 × 8 × 32 feature blocks of encoder G_{E1} and decoder G_D are concatenated into an 8 × 8 × 64 feature block, which is further upsampled by 2× to obtain the 16 × 16 × 16 feature block of decoder G_D. Finally, the 16 × 16 × 16 feature blocks of encoder G_{E1} and decoder G_D are concatenated into a 16 × 16 × 32 feature block, which is upsampled by 2× to obtain $\hat{X}$.

(3) SE: Since CLEDFF directly connects the same-scale encoding and decoding features, this connection mechanism enables cross-layer information transmission and can compensate for information loss. However, same-scale features often contain similar but not exactly the same information, so the feature information transmitted through CLEDFF may contain redundant information. The existence of this redundant feature information can reduce the network's generalization ability and may lead to overfitting problems.

To avoid the problem of redundant information in CLEDFF, this paper proposes the use of SE mechanism. As shown in Figure 7, after obtaining the concatenated feature maps M through CLEDFF, this method uses the SE mechanism to obtain the feature maps M-weight containing weight information. The SE mechanism mainly consists of three steps: Squeeze, Excitation, and Scale (*33*).

Squeeze: Firstly, the feature concatenation map M of the encoder and decoder is reduced in dimensionality through global average pooling, resulting in a numerical representation for each feature channel, and yielding a feature representation z. This is shown in Equation (1), where z represents the feature representation, M represents the feature concatenation map, and H, W, and C represent the height, width, and number of channels of the feature concatenation map.

$$z = Squeeze(M) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} M(i,j) \quad (1)$$

Excitation: Next, the feature representation z is non-linearly transformed and mapped into a weight vector s. This process is accomplished through two fully connected layers, where different numerical values in s represent the weight information of different channels. As shown in Equation (2), where s represents the weight vector, W_1 represents the parameters of the first fully connected layer, Relu is the activation function of the first fully connected layer, W_2 represents the parameters of the second fully connected layer, and Sigmoid is the activation function of the second fully connected layer.

$$s = Excitation(z) = sigmoid(W_2 \, \mathrm{Re}lu(W_1 z)) \quad (2)$$

Scale: Finally, the weight vector s is applied to the original feature maps M to obtain the weighted feature maps M-weight. Specifically, the feature concatenation map is weighted by multiplying it with the weight vector s generated in the third step, resulting in a feature concatenation map containing weight information. As shown in Equation (3), where M-weight represents the weighted feature maps.

$$M - weight = Scale(M, s) = M \times s \quad (3)$$

2.4. Irrelevant gesture discriminator module

The process of this module is shown in Figure 1. Firstly, the reconstruction error between the feature vectors z and $\hat{z}$ is calculated using L2 distance, as shown in equation (4). Then, the reconstruction error is compared with a pre-defined threshold value. If the reconstruction error is greater than the threshold, it is classified as an irrelevant gesture; otherwise, it is classified as the target gesture. The equation for the classification is shown in equation (5), where 0 represents the target gesture and 1 represents the irrelevant gesture.

$$\text{error} = \left\| z - \hat{z} \right\|_2 \quad (4)$$

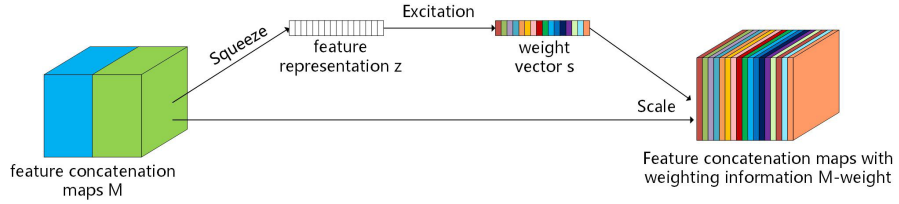

Figure 7. SE Diagram. The Squeeze operation obtains the feature representation z for each channel of the feature concatenation map M; the Excitation operation obtains the weight s for each channel; the Scale operation obtains the feature concatenation map M-weight containing weight information.

$$\text{label} = \begin{cases} 0, & error <= threshold \\ 1, & error > threshold \end{cases} \quad (5)$$

The principles for threshold selection are as follows: Firstly, the reconstruction errors of the feature vectors from all samples in the test set are sorted. Secondly, these values are iterated in ascending order to serve as the current threshold. Subsequently, the samples are predicted as either belonging to the target category or the irrelevant category based on the threshold. Next, by combining the predicted labels with the true labels, the false positive rate (FPR) and true positive rate (TPR) are calculated for each threshold. Finally, the optimal threshold is determined by selecting the value at which the difference between the TPR and FPR is minimized, ensuring the best trade-off between TPR and FPR.

3. Results

3.1. Experimental environment and parameter settings

The computer configuration used in the experiments of this study is as follows: an Intel Core i5-8250U CPU processor (with 8GB of memory), an NVIDIA GeForce 940MX graphics card (with 2GB of memory), and the Windows 10 operating system. The network model was trained and tested using the Python 3.7 programming language in the PyTorch 1.2.0 deep learning framework.

The training parameter settings of the EMG-FRNet: Adam optimizer was used with a smoothing constant ($\beta1$, $\beta2$) of (0.5, 0.999), the initial learning rate lr was set to 2e−3, the batch size was set to 64, and the number of training epochs was set to 200.

3.2. Evaluation Indicators

In this paper, the area under the Receiver Operating Characteristic (ROC) curve (AUC) is used to evaluate the performance of the model. AUC values range from 0 to 1, and the larger the AUC value, the better the performance of the model.

The confusion matrix is the basis for drawing the ROC curve. In the confusion matrix, TP represents true positive, indicating that the sample's true class is positive and the model recognition result is also positive. Similarly, FN represents false negative, FP represents false positive, and TN represents true negative (*34*).

Based on the formulas (6) and (7), the FPR and TPR can be obtained. FPR represents the ratio of negative samples that are incorrectly classified as positive. TPR represents the ratio of positive samples that are correctly classified as positive.

$$FPR = \frac{FP}{FP + TN} \quad (6)$$

$$TPR = \frac{TP}{TP + FN} \quad (7)$$

For binary classification tasks, a fixed threshold can be set to obtain a (FPR, TPR) pair. By plotting the (FPR, TPR) pairs corresponding to different thresholds on a coordinate system, the ROC curve can be obtained. The ROC curve represents the recognition performance of the model under different thresholds.

Furthermore, this paper uses the AUC to quantitatively evaluate the performance of the model in recognizing irrelevant gestures.

3.3. Comparison experiment

In this paper, the proposed EMG-FRNet for EMG-based irrelevant gesture recognition is compared with the existing methods based on SVDD (*19*) and AE (*20*). These comparative algorithms are all state-of-the-art methods in the field of irrelevant gesture recognition and have demonstrated good performance in this area, thus we chose them as the comparison algorithms in this study.

The experimental settings are as described in Section 2.1, where each subject undergoes multiple experiments, with a different target gesture category set for each experiment. DB1 and DB5 each have 10 subjects, and each subject performs 8 different target gesture experiments. The self-collected dataset consisted of 6 participants, with each participant performing 7 different target gesture experiments. Firstly, the AUC values are recorded for each subject when setting different target gestures. Then, the AUC values corresponding to different target gestures are averaged to obtain the subject's AUC value. Finally, the AUC values of all subjects in the dataset are averaged to obtain the dataset's AUC value. Figure 8 shows the AUC line graph for different subjects in each dataset. Table 1 shows the AUC values for each dataset.

From Figure 8, it can be seen that for different subjects, compared to the SVDD and AE comparison algorithms, the AUC value of the EMG-FRNet model can always maintain a high and relatively stable level. Specifically, the AUC values on different datasets are described in Table 1.

In the aforementioned comparative experiments, SVDD performed the worst. The reason for this may be that the traditional one-class support vector machine method is used, based on spherical hyperplanes. This method leads to poor classification performance in the case of complex data distributions such as EMG samples. The AE achieves improvement based on the SVDD. This may be because AE uses an autoencoder method for feature learning and extraction, which can better explore the intrinsic features of data. The EMG-FRNet achieves improvement based on the AE. The reason for this may be that the method utilizes the advantages of Generative Adversarial Networks (GANs) in anomaly detection. It also combines strategies such as CC, CLEDFF, and SE to further improve the model's performance in EMG

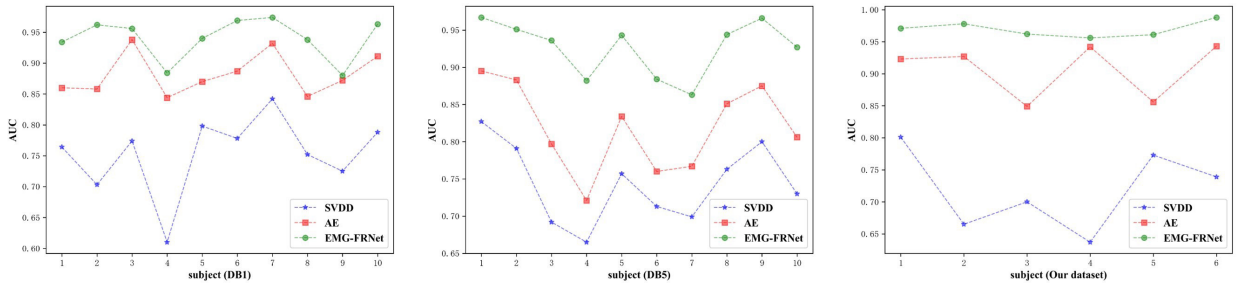

Figure 8. AUC line chart for different subjects in each dataset. (a)-(c) represent the AUC line charts of SVDD, AE, and EMG-FRNet on different subjects in DB1, DB5, and the self-collected dataset, respectively.

Table 1. AUC values for each dataset (Comparison experiment)

Model Name	DB1	DB5	Our dataset
SVDD (*19*)	0.744	0.753	0.719
AE (*20*)	0.882	0.819	0.907
EMG-FRNet	0.940	0.926	0.969

irrelevant gesture recognition tasks.

In conclusion, the proposed EMG-FRNet achieves state-of-the-art (SOTA) performance in the task of EMG irrelevant gesture recognition.

3.4. Ablation experiments

To validate the effectiveness of the proposed method, the following ablation experiments were conducted: based on the original GANomaly network, CC, CLEDFF, and SE were successively added. In the legend below, GANomaly is represented as Baseline, GANoamly+CC is represented as Model 1, GANomaly+CC+CLEDFF is represented as Model 2, and GANomaly+CC+CLEDFF+SE (EMG-FRNet) is represented as Model 3.

The experimental setup is described in Section 2.1. Figure 9 demonstrates the AUC fold plots for different subjects for each dataset. Table 2 demonstrates the AUC values for each dataset.

From Figure 9, it can be seen that the performance of Baseline, Model 1, Model 2, and Model 3 increases in turn. Specifically, the AUC values on different datasets are described in Table 2.

In the ablation experiments, Model 1 achieves improvement based on the Baseline. The reason for this improvement is that the CC reduces the impact of redundant features in the original GANomaly network on model performance. Model 2 further improves upon Model 1. The reason for this improvement is that the CLEDFF can compensate for information loss during downsampling. Model 3 has been improved from model 2. The reason for this improvement is that the SE can avoid the problem of redundant information brought by feature fusion, making the network focus on more important features and improving the reconstruction performance of the model.

In summary, the proposed model EMG-FRNet achieves the best performance in the task of recognizing irrelevant gestures. CC, CLEDFF, and SE all improve the model's performance to varying degrees.

4. Discussion

Currently, the recognition performance of most studies on EMG irrelevant gesture recognition is unstable. This paper establishes a connection between the EMG irrelevant gesture recognition and anomaly detection fields, and for the first time applies GANomaly to EMG irrelevant gesture recognition. Based on this, a feature reconstruction network, EMG-FRNet, is proposed for EMG irrelevant gesture recognition. The network exhibits a small feature reconstruction error for target class samples and a large feature reconstruction error for irrelevant class samples, which improves the ability to distinguish between target and irrelevant samples. We verify the feasibility of the proposed method through experiments, and the results show that our method can maintain a high level of performance in all EMG datasets.

In this paper, we have achieved high reliability in distinguishing target gestures from multiple irrelevant gestures. However, in practical myoelectric interaction applications, there are often multiple types of target gestures, which not only require us to distinguish between target and irrelevant gestures but also to distinguish between different categories of target gestures. In addition, there are other types of interference in practical myoelectric interaction applications, such as electrode displacement, muscle fatigue, user differences, *etc*., which can lead to poor interaction effects. Therefore, for future research, we will further explore the following aspects: (*i*) Based on the EMG-FRNet method proposed in this paper, we aim to achieve the distinction between multiple target gestures and multiple irrelevant gestures, as well as the distinction between different categories of target gestures. (*ii*) Additionally, we will seek corresponding solutions for other types of interference, with the goal of improving the interaction effects in

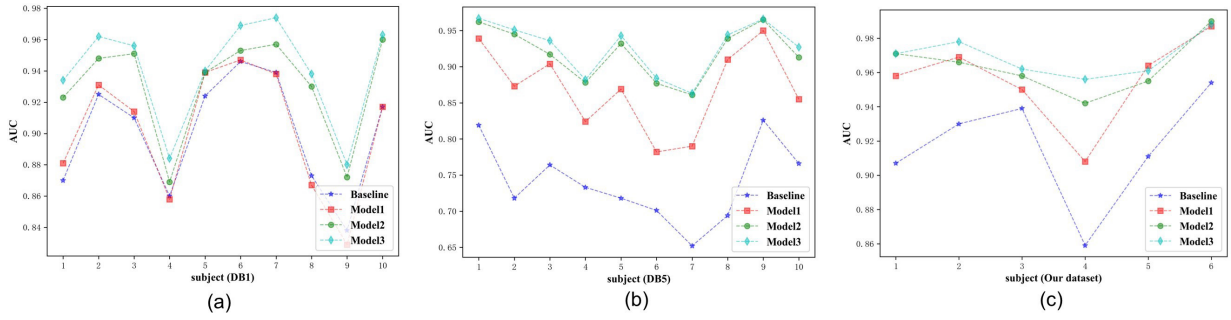

Figure 9. AUC line chart for different subjects in each dataset. (a)-(c) represent the AUC line charts of GANomaly, GANomaly+CC, GANomaly+CC+ CLEDFF, and EMG-FRNet on different subjects in DB1, DB5, and the self-collected dataset, respectively.

Table 2. AUC values for each dataset (Ablation experiments)

Items	CC	CLEDF	SE	DB1	DB5	Our dataset
Baseline				0.900	0.739	0.915
Model 1	√			0.902	0.870	0.956
Model 2	√	√		0.930	0.919	0.963
Model 3	√	√	√	0.940	0.926	0.969

practical myoelectric applications.

Funding: This study was partially supported by Beijing Municipal Science & Technology Comission No.Z221100007422068

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

Author statemen: Zhang W and Wang Y conceived the ideas and designed the methodology. Zhang W and Wang Y implemented the technical pipeline, conducted the experiments, and analyzed the results. Zhang J and Pang G provided the dataset for the experiments; All authors discussed, wrote the manuscript, and gave final approval for publication.

References

1. Lv Z, Xiao F, Wu Z, Liu Z, Wang Y. Hand gestures recognition from surface electromyogram signal based on self-organizing mapping and radial basis function network. Biomed Signal Process Control. 2021, 68:102629.
2. Zhang Z, He C, Yang K. A Novel Surface Electromyographic Signal-Based Hand Gesture Prediction Using a Recurrent Neural Network. Sensors (Basel). 2020; 20:3994.
3. Xu Z, Tian Y, Li Y. sEMG Pattern Recognition of Muscle Force of Upper Arm for Intelligent Bionic Limb Control. J Bionic Eng. 2015; 12, 316–323.
4. Cases C, Baldovino R, Manguerra M, Dupo V. An EMG-based Gesture Recognition for Active-assistive Rehabilitation. 2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM). DOI: 10.1109/HNICEM51456.2020.9400132
5. Yang X, Chen X, Cao X, Wei S, Zhang X. Chinese Sign Language Recognition Based on an Optimized Tree-Structure Framework. IEEE J Biomed Health Inform. 2017; 21:994-1004.
6. Xu H, Xiong A. Advances and Disturbances in sEMG-Based Intentions and Movements Recognition: A Review. IEEE Sens J. 2021; 21:3019-13028. doi: 10.1109/JSEN.2021.3068521.
7. Li X, Chen S, Zhang H, Samuel OW, Wang H, Fang P, Zhang X, Li G. Towards reducing the impacts of unwanted movements on identification of motion intentions. J Electromyogr Kinesiol. 2016; 28:90-98.
8. Chang J, Phinyomark A, Bateman S, Scheme E. Wearable EMG-Based Gesture Recognition Systems During Activities of Daily Living: An Exploratory Study. Annu Int Conf IEEE Eng Med Biol Soc. 2020; 2020:3448-3451.
9. Robertson JW. Toward an enhanced understanding of rejection in pattern recognition-based myoelectric control. University of New Brunswick, 2019.
10. Chen H, Zhang Y, Zhou D, Liu H. Improving Gesture Recognition by Bidirectional Temporal Convolutional Netwoks. International Conference on Robotics and Rehabilitation Intelligence. Springer, Singapore, 2020; 413-424.
11. Rahimian E, Zabihi S, Asif A, Farina D, Atashzar SF, Mohammadi A. TEMGNet: Deep Transformer-based Decoding of Upperlimb sEMG for Hand Gestures Recognition. arXiv:2109.12379, 2021. https://doi.org/10.48550/arXiv.2109.12379
12. Zhang W, Zhao T, Zhang J, Wang Y. LST-EMG-Net: Long short-term transformer feature fusion network for sEMG gesture recognition. Front Neurorobot. 2023; 17:1127338.
13. Scheme E J, Hudgins B S, Englehart K B. Confidence-based rejection for improved pattern recognition myoelectric control. IEEE Trans Biomed Eng. 2013; 60:1563-1570
14. Robertson J W, Englehart K B, Scheme E J. Effects of confidence-based rejection on usability and error in pattern recognition-based myoelectric control. IEEE J Biomed Health Inform. 2019; 23:2002-2008.
15. Tomczyński J, Kaczmarek P, Mańkowski T. Hand gesture-based interface with multichannel sEMG band enabling unknown gesture discrimination. 2015 10th International Workshop on Robot Motion and Control (RoMoCo). DOI: 10.1109/RoMoCo.2015.7219713
16. Bao T, Zaidi S AR, Xie SQ, Yang PF, Zhang ZQ. CNN Confidence Estimation for Rejection-Based Hand Gesture Classification in Myoelectric Control. IEEE Trans Hum Mach Syst. 2021; 52:99-109.

17. Zhou S, Fei F, Yin K. Toward improving the reliability of discrete movement recognition of sEMG signals. Applied Sciences. 2022; 12: 3374.
18. Ding Q, Li Z, Zhao X, Xiao Y, Han J. Real-time myoelectric prosthetic-hand control to reject outlier motion interference using one-class classifier. 2017 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC). DOI: 10.1109/YAC.2017.7967385
19. Ding Q, Zhao X, Han J, Bu C, Wu C. Adaptive Hybrid Classifier for Myoelectric Pattern Recognition Against the Interferences of Outlier Motion, Muscle Fatigue, and Electrode Doffing. IEEE Trans Neural Syst Rehabil Eng. 2019; 27:1071-1080.
20. Wu L, Chen X, Chen X, Zhang X. Rejecting novel motions in high-density myoelectric pattern recognition using hybrid neural networks. Front Neurorobot. 2022; 16:862193.
21. Wu L, Zhang X, Zhang X, Chen X, Chen X. Metric learning for novel motion rejection in high-density myoelectric pattern recognition. Knowl Based Syst. 2021; 227:107165.
22. Pang G, Shen C, Cao L, Hengel A. Deep learning for anomaly detection: A review. ACM Comput Surv (CSUR), 2021; 54:1-38.
23. Bati E, Çalışkan A, Koz A, Alatan AA. Hyperspectral anomaly detection method based on auto-encoder. Proc. SPIE 9643, Image and Signal Processing for Remote Sensing XXI, 96430N. 2015; https://doi.org/10.1117/12.2195180.
24. Cowton J, Kyriazakis I, Plötz T, Bacardit J. A Combined Deep Learning GRU-Autoencoder for the Early Detection of Respiratory Disease in Pigs Using Multiple Environmental Sensors. Sensors (Basel). 2018; 18:2521.
25. An J, Cho S. Variational autoencoder based anomaly detection using reconstruction probability. Special lecture on IE. 2015; 2:1-18. *http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf*
26. Schlegl T, Seeböck P, Waldstein SM, Schmidt-Erfurth U, Langs G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings. Cham: Springer International Publishing, 2017: 146-157.
27. Zenati H, Foo CS, Lecouat B, Manek G, Chandrasekhar VR. Efficient gan-based anomaly detection. arXiv:1802.06222, 2018. https://doi.org/10.48550/arXiv.1802.06222
28. Akcay S, Atapour-Abarghouei A, Breckon TP. GANomaly: Semi-supervised anomaly detection *via* adversarial training. Computer Vision – ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2-6, 2018, Revised Selected Papers, Part III 14. Springer International Publishing, 2019; 622-637.
29. Li Q, Luo Z, Zheng J. Deep Learning-based User Authentication with Surface EMG Images of Hand Gestures. Annu Int Conf IEEE Eng Med Biol Soc. 2021; 2021:2038-2041.
30. Li Q, Luo Z, Zheng J. A new deep anomaly detection-based method for user authentication using multichannel surface EMG signals of hand gestures. IEEE Trans Instrum Meas. 2022; DOI: 10.1109/TIM.2022.3164162
31. Atzori M, Gijsberts A, Heynen S, Hager AGM, Deriaz O, Van Der Smagt P, Castellini C, Caputo B, Müller H. Building the Ninapro database: A resource for the biorobotics community.2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob). IEEE, 2012; DOI: 10.1109/BioRob.2012.6290287
32. Pizzolato S, Tagliapietra L, Cognolato M, Reggiani M, Müller H, Atzori M. Comparison of six electromyography acquisition setups on hand movement classification tasks. PLoS One. 2017; 12:e0186132.
33. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: DOI: 10.1109/CVPR.2018.00745
34. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Information processing & management. 2009; 45:427-437.

Received May 21, 2023; Revised June 21, 2023; Accepted June 29, 2023.

**Address correspondence to:*
Wenli Zhang, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China.
E-mail: zhangwenli@bjut.edu.cn

Released online in J-STAGE as advance publication June 30, 2023.

Correspondence

DOI: 10.5582/bst.2023.01119

The potential of 'Segment Anything' (SAM) for universal intelligent ultrasound image guidance

Guochen Ning[1,*], Hanyin Liang[1], Zhongliang Jiang[2], Hui Zhang[1], Hongen Liao[1,*]

[1] Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China;
[2] Computer Aided Medical Procedures, Technical University of Munich, Germany.

SUMMARY Ultrasound image guidance is a method often used to help provide care, and it relies on accurate perception of information, and particularly tissue recognition, to guide medical procedures. It is widely used in various scenarios that are often complex. Recent breakthroughs in large models, such as ChatGPT for natural language processing and Segment Anything Model (SAM) for image segmentation, have revolutionized interaction with information. These large models exhibit a revolutionized understanding of basic information, holding promise for medicine, including the potential for universal autonomous ultrasound image guidance. The current study evaluated the performance of SAM on commonly used ultrasound images and it discusses SAM's potential contribution to an intelligent image-guided framework, with a specific focus on autonomous and universal ultrasound image guidance. Results indicate that SAM performs well in ultrasound image segmentation and has the potential to enable universal intelligent ultrasound image guidance.

Keywords medical artificial intelligence, intelligent medical image guidance, large models, ultrasound image

Ultrasound image guidance is a method often used to help provide care, and it relies on the accurate perception of information from ultrasound images to guide medical procedures. It has become an indispensable tool in many clinical settings, ranging from diagnostics and interventions to minimally invasive therapies (*1*). Moreover, ultrasound images can be used as guidance to direct surgery and can also be acquired by a robot to further develop an autonomous image-guided process (*2*). One critical aspect of ultrasound image guidance is image segmentation, which involves identifying and delineating specific structures or regions of interest within ultrasound images. Image segmentation plays a crucial role in ultrasound-guided procedures as it enables physicians to accurately locate and target the tissues or structures of interest, monitor their changes in real-time, and guide the intervention or therapy accordingly (*3*). Moreover, the segmented structures can serve as the basis for planning and guiding robotic ultrasound systems, enabling autonomic positioning of an ultrasound probe for optimal imaging of the regions of interest (*4*). Therefore, accurate segmentation is essential for precise targeting and localization of structures, and it is currently achieved with AI models that have different structures. However, the reliance on specific AI models for tissue segmentation limits autonomous ultrasound image-guided procedures, like high-intensity focused ultrasound (HIFU) therapy, in clinical settings (*5*). It may require frequent updates and retraining of the AI models to adapt to different tissues, which is time-consuming and may not always be feasible in clinical settings (*6*). In addition, the availability of labeled ultrasound datasets for training may be limited, and especially those for rare diseases or specific procedures, further adding to the challenges of developing accurate and robust tissue segmentation models for intelligent ultrasound image guidance (*6*).

Large models, also called foundation models, have brought about a revolution in many fields, driving breakthroughs in diverse applications (*7*). One such model, ChatGPT, has garnered considerable recognition for its ability to generate human-like text responses and engage in interactive conversations (*8*). Numerous experts have described the potential for large models to be used in healthcare and their impact on traditional healthcare models (*9*). Inspired by the success of ChatGPT, the Segment Anything Model (SAM) is a state-of-the-art foundation model that enhances the generalization capability of image segmentation to new heights (*10*). Although several studies are attempting to improve segmentation generalizability by introducing new structures, large models are still only trained for a single anatomy. In contrast, SAM leverages advanced machine learning algorithms to automatically identify and delineate objects, structures, or regions of interest within

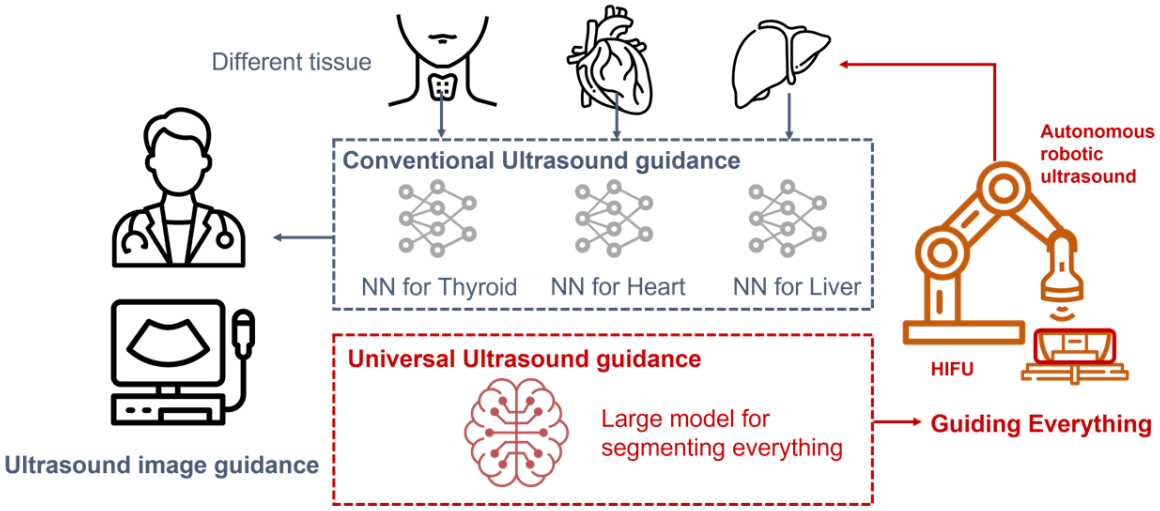

Figure 1. Can "Segment Anything" help with fully autonomous ultrasound image guidance?

images, regardless of their multimodality or diversity. With its unprecedented capabilities, SAM opens up new possibilities for visual analysis, understanding, and manipulation in different fields. The versatility of SAM in segmenting different organs enhances the potential of intelligent medical image guidance in various scenarios. Unlike traditional models that require the training of specific segmentation models for each organ, SAM's ability to segment different organs with a high level of accuracy allows for a more generalized approach (*11*). The current study has evaluated the performance of SAM on ultrasound image segmentation and assessed if the results can serve as a catalyst for a universal image-guided system, as shown in Figure 1.

The evaluation was performed in two parts, focusing on SAM's performance on ultrasound image data and the reliability of the segmentation results for guidance. To assess SAM's performance, three models of interaction were selected, namely Everything mode, Click mode, and Box mode, with each representing a level of interaction. In Everything mode, SAM automatically segments the image within its range of detection. In Click mode, SAM utilizes human clicks on the image as prompts for segmentation, with these clicks randomly generated within the region of the ultrasound image segmentation label. Lastly, in Box mode, SAM segments are based on boxes drawn by a human on the image as prompts, where the enclosing boxes are generated from the segmentation labels of the ultrasound image. In addition, Click mode and Box mode were combined in Click-Box mode to include more interactive information.

The quantitative evaluation metric used in this study is the DICE coefficient between the labels and the segmentation results obtained from SAM (*12*). DICE measures the similarity between two samples and is a reproducibility validation metric most frequently used in medical image segmentation. Three public ultrasound image datasets, denoted as Carotid Artery (*13*), Thyroid (*14*), and Heart (*15*), were selected for evaluation. These datasets consist of continuous ultrasound images, which are more representative of the real-time acquisition of ultrasound images during guidance. In the Click model, the accuracy of segmentation was 0.877 ± 0.115 for the Carotid artery, 0.427 ± 0.320 for the Thyroid, and 0.326 ± 0.153 for the Heart. In the Box model, the accuracy of segmentation was 0.908 ± 0.041 for the Carotid artery, 0.829 ± 0.126 for the Thyroid, and 0.841 ± 0.098 for the Heart. In the Click-Box model, the accuracy of segmentation was 0.909 ± 0.041 for the Carotid artery, 0.829 ± 0.121 for the Thyroid, and 0.867 ± 0.051 for the Heart. Figure 2 shows some ultrasound images obtained using different segmentation models. Results generated by the Everything model are difficult to evaluate quantitatively due to the inclusion of multiple objectives. The results of quantitative segmentation reveal that SAM can achieve a high level of accuracy in carotid artery segmentation when provided with correct prompts, such as Click mode and Box mode. Click-Box mode performs better as the prompts are more informative. In contrast, Everything mode segments all potential structures, including blood vessels, and may have difficulty depicting the major structure of interest. This suggests that ultrasound image segmentation requires more informative prompts for SAM compared to natural image segmentation in order to accurately reveal the target structure.

Another important aspect to consider in evaluating the usability of segmentation results for autonomic guidance is continuity, which determines the stability and flexibility of guidance in a continuous process (*16-18*). To assess this, the stability of the segmentation results was evaluated based on sequential images, in specific terms, the standard deviation of the segmentation results within an image sequence. For each dataset, five sets of images were selected for evaluation. Each contains 10 consecutive images. The standard deviation of the

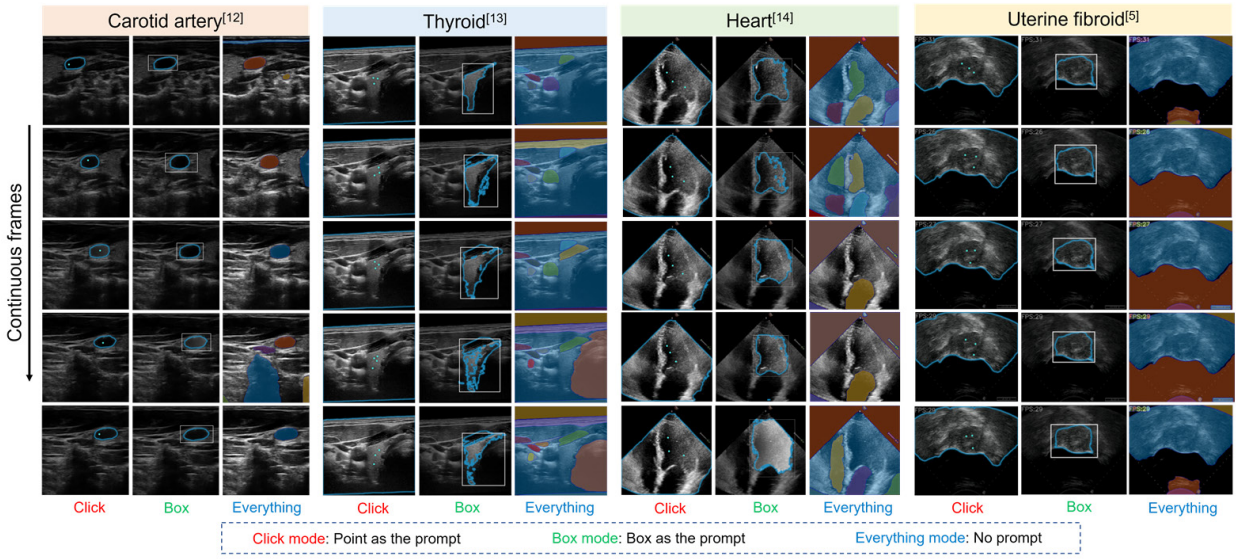

Figure 2. Qualitative results of SAM segmentation of ultrasound images. For simple and obvious structures, SAM yields accurate segmentation results and good sequence stability, even in less interactive Click mode. For less clear tissues such as the thyroid, however, SAM still fails to segment the tissue accurately in Box mode. And due to the lack of prompt input, Everything mode can automatically segment easily discerned tissues in ultrasound images but the results obtained are not those desired. Pictures were generated from: *https://segment-anything.com*

three SAM segmentation models on the Carotid artery sequence was 0.137, 0.066, and 0.065. The standard deviation of the three SAM segmentation models on the Thyroid sequence was 0.273, 0.153, and 0.150. The standard deviation of the three SAM segmentation models on the Heart sequence was 0.112, 0.117, and 0.074. Data on the Heart had a lower standard deviation than data on carotid artery segmentation, which is inconsistent with segmentation accuracy. These results indicate that while the overall accuracy of the segmentation results may be satisfactory in carotid artery images, there are still fluctuations in the stability of the continuous sequential images. These instabilities can lead to shifts and errors in autonomic guidance during imaging and can accumulate during clinical guidance. This is presumably because SAM does not currently take into account the continuity between input images, as it is designed to segment a single image without considering the contextual relationship between similar frames in video data.

The breakthrough achieved with SAM has brought possibilities for downstream tasks based on image segmentation, including fully autonomous ultrasound image-guided medical procedures. Preliminary results have revealed that SAM can achieve a good segmentation accuracy for ultrasound images with sufficient prompts, although it requires an additional interaction process compared to a specially trained model. However, the instability in continuous image segmentation, due to the lack of contextual information, introduces errors in long-term guidance. Overall, SAM has great potential as a large model for image segmentation in general-purpose tasks, but there are some areas that can be improved. First, the prompts need to be enhanced. Like ChatGPT, SAM relies on well-developed prompts to yield stable and accurate outputs, which requires a certain level of operator experience (*19-22*). This also means that fully autonomous ultrasound image guidance for different tissues still requires a lot of work. Second, the performance on medical data, such as ultrasound images, needs to be improved. Medical images have distinct imaging principles and characteristics compared to natural images, which may limit the applicability of SAM's segmentation model to ultrasound images. Therefore, developing a general segmentation model specifically for medical images is worth investigating (*23*). Moreover, continuous data, such as continuous ultrasound images or video, containing temporal information can provide effective cues for segmentation. Thus, exploring the use of continuous data in conjunction with SAM for image segmentation can be beneficial.

In conclusion, the emergence of SAM and other large models has opened up numerous possibilities for image-based downstream tasks, including an intelligent or autonomous ultrasound image guidance system. With further iteration of large models and refinement of data, the current authors believe that fully intelligent medical image-guiding systems and autonomous image-guided therapy will become a reality in the future.

Funding: This work has been supported by the National Key Research and Development Program of China (2022YFC2405200), the National Natural Science Foundation of China (82027807, U22A2051), the Beijing Municipal Natural Science Foundation (7212202), the Institute for Intelligent Healthcare, Tsinghua University (2022ZLB001), and the Tsinghua-Foshan Special Fund for Innovation (2021THFS0104).

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

References

1. Noble J A, and Djamal B. Ultrasound image segmentation: A survey. IEEE Trans Med Imaging. 2006; 25: 987-1010.
2. Salcudean SE, Moradi H, Black DG, Navab N. Robot-assisted medical imaging: A review. Proceedings of the IEEE. 2022; 110:951-967.
3. Ning G, Zhang X, and Liao H. Autonomic robotic ultrasound imaging system based on reinforcement learning. IEEE Trans Biomed Engineer 2021; 68:2787-2797.
4. Roshan MC, Pranata A, Isaksson M. Robotic ultrasonography for autonomous non-invasive diagnosis – A systematic literature review. IEEE Trans Med Robotics Bionics. 2022; 4:863-874.
5. Ning G, Zhang X, Zhang Q, Wang Z, Liao H. Real-time and multimodality image-guided intelligent HIFU therapy for uterine fibroid. Theranostics. 2020; 10:4676-4693.
6. Wang L, Guo D, Wang G, Zhang S. Annotation-efficient learning for medical image segmentation based on noisy pseudo labels and adversarial learning. IEEE Trans Med Imaging. 2021; 40:2795-2807.
7. Brown T, Mann B, Ryder N, *et al*. Language models are few-shot learners. Advances in neural information processing systems. 2020; 33:1877-1901.
8. Haupt CE, Marks M. AI-Generated Medical Advice-GPT and Beyond. JAMA. 2023; 329:1349-1350.
9. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ and Rajpurkar P. Foundation models for generalist medical artificial intelligence. Nature. 2023; 616:259-265.
10. Kirillov A, Mintun E, Ravi N, Mao HZ, Rolland C, Gustafson L, Xiao TT, Whitehead S, Berg AC, Lo WY, Dollár P, Girshick R. Segment anything. arXiv preprint arXiv:2304.02643, 2023. https://doi.org/10.48550/arXiv.2304.02643
11. Ji G, Fan D, Xu P, Cheng M, Zhou B, Gool LV, SAM struggles in concealed scenes –Empirical study on "Segment Anything". arXiv:2304.06022, 2023. https://doi.org/10.48550/arXiv.2304.06022
12. Wang G, Zhai S, Lasio G, Zhang B, Yi B, Chen S, Macvittie TJ, Metaxas D, Zhou J, Zhang S. Semi-supervised segmentation of radiation-induced pulmonary fibrosis from lung CT scans with multi-scale guided dense attention. IEEE Trans Med Imaging. 2022; 41:531-542.
13. Wunderling T, Golla B, Poudel P, Arens C, Friebe M and Hansen C, Comparison of thyroid segmentation techniques for 3D ultrasound. Proceedings of SPIE Medical Imaging, Orlando, USA, 2017.
14. Říha K, Mašek J, Burget R, Beneš R, Závodná E. Novel method for localization of common carotid artery transverse section in ultrasound images using modified Viola-Jones detector. Ultrasound Med Biol. 2013; 39:1887-902.
15. Leclerc S, Smistad E, Pedrosa J, Ostvik A, Cervenansky F, Espinosa F, Espeland T, Berg EAR, Jodoin PM, Grenier T, Lartizien C, Dhooge J, Lovstakken L, Bernard O. Deep Learning for Segmentation Using an Open Large-Scale Dataset in 2D Echocardiography. IEEE Trans Med Imaging. 2019; 38:2198-2210.
16. Ouyang D, He B, Ghorbani A, Lungren MP, Ashley EA, Liang DH, Zou JY. Echonet-dynamic: A large new cardiac motion video data resource for medical machine learning, NeurIPS ML4H Workshop: Vancouver, BC, Canada. 2019. *https://echonet.github.io/dynamic/NeuroIPS_2019_ML4H%20Workshop_Paper.pdf* (accessed May 20, 2023).
17. Ning G, Liang H, Zhang X, Liao H. Inverse-reinforcement-learning-based robotic ultrasound active compliance control in uncertain environments. IEEE Trans Indust Electron. 2023:1-10. doi:10.1109/tie.2023.3250767
18. Jiang Z, Li Z, Grimm M, Zhou MC, Esposito M, Wein W, Stechele W, Wendler T, Navab N. Autonomous robotic screening of tubular structures based only on real-time ultrasound imaging feedback. IEEE Trans Indust Electron. 2022; 69:7064-7075.
19. Shen Y, Tan X, Li D, Lu W, Zhuang Y. HuggingGPT: Solving AI tasks with ChatGPT and its friends in Hugging Face. arXiv:230317580. https://doi.org/10.48550/arXiv.2303.17580
20. Deng R, Cui C, Liu Q, *et al*. Segment anything model (SAM) for digital pathology: Assess zero-shot segmentation on whole slide imaging. arXiv: 2304.04155, 2023. https://doi.org/10.48550/arXiv.2304.04155
21. Yuan B, Jiang ZL, Clarenbach R, Ghotbi R, Karlas A, Navab N. MI-SegNet: Mutual information-based US segmentation for unseen domain generalization. arXiv:2303.12649, 2023. https://doi.org/10.48550/arXiv.2303.12649
22. Jiang Z, Gao Y, Xie L, Navab N. Towards autonomous atlas-based ultrasound acquisitions in presence of articulated motion. IEEE Robot Automation Letters 7.3 2022: 7423-7430.
23. Ma J, Wang B, Segment anything in medical images, arXiv:2304.12306, 2023. https://doi.org/10.48550/arXiv.2304.12306

Received May 23, 2023; Revised June 10, 2023; Accepted June 15, 2023.

**Address correspondence to:*
Guochen Ning, Hongen Liao Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China.
E-mail: liao@tsinghua.edu.cn

Released online in J-STAGE as advance publication June 22, 2023.

Correspondence

DOI: 10.5582/bst.2023.01087

Devising novel near-infrared aggregation-induced-emission luminogen labeling for point-of-care diagnosis of *Mycobacterium tuberculosis*

Guiqin Dai[1,2,3,§], Pengfei Zhao[1,3,§], Lijun Song[2,§], Zhuojun He[1,2,3,§], Deliang Liu[1,3], Xiangke Duan[1,3], Qianting Yang[3], Wenchang Zhao[2,*], Jiayin Shen[1,4], Tetsuya Asakawa[1], Mingbin Zheng[1,2,3,*], Hongzhou Lu[1,5,*]

[1] Institute of Neurology, National Clinical Research Center for Infectious Diseases, Third People's Hospital of Shenzhen, Shenzhen, Guangdong, China;
[2] Key Laboratory for Nanomedicine, Guangdong Medical University, Dongguan, Guangdong, China;
[3] Institute for Hepatology, National Clinical Research Center for Infectious Diseases, Third People's Hospital of Shenzhen, Shenzhen, Guangdong, China;
[4] Department of Science and Education, National Clinical Research Center for Infectious Diseases, Third People's Hospital of Shenzhen, Shenzhen, Guangdong, China;
[5] Department of Infectious Diseases, National Clinical Research Center for Infectious Diseases, Third People's Hospital of Shenzhen, Shenzhen, Guangdong, China.

SUMMARY Detecting and appropriately diagnosing a *Mycobacterium tuberculosis* infection remains technologically difficult because the pathogen commonly hides in macrophages in a dormant state. Described here is novel near-infrared aggregation-induced-emission luminogen (AIEgen) labeling developed by the current authors' laboratory for point-of-care (POC) diagnosis of an *M. tuberculosis* infection. The selectivity of AIEgen labeling, the labeling of intracellular *M. tuberculosis* by AIEgen, and the labeling of *M. tuberculosis* in sputum samples by AIEgen, along with its accuracy, sensitivity, and specificity, were preliminarily evaluated. Results indicated that this near-infrared AIEgen labeling had satisfactory selectivity and it labeled intracellular *M. tuberculosis* and *M. tuberculosis* in sputum samples. It had a satisfactory accuracy (95.7%), sensitivity (95.5%), and specificity (100%) for diagnosis of an *M. tuberculosis* infection in sputum samples. The current results indicated that near-infrared AIEgen labeling might be a promising novel diagnostic tool for POC diagnosis of *M. tuberculosis* infection, though further rigorous verification of these findings is required.

Keywords *Mycobacterium tuberculosis*, near-infrared aggregation-induced emission luminogen, point-of-care diagnosis, metabolic labeling, high-throughput screening

Tuberculosis (TB) is a common infectious disease caused by infection with *Mycobacterium tuberculosis* that may spread *via* aerosol and involve various organs (*1*). More than 10 million cases are confirmed per year. Other than COVID-19, TB is the second leading cause of death due to infectious disease (*2,3*). TB remains a major public health concern, particularly in developing countries, and indeed it cannot be ignored or underestimated. Earlier appropriate diagnosis and timely anti-TB therapy may result in better clinical outcomes. However, many conventional diagnostic methods, such as biomedical imaging, bacteriological diagnosis, molecular diagnosis, and immunologic testing, suffer from several methodological flaws, resulting in a low specificity, low sensitivity, and demanding conditions for pathogen detection (*4-6*). The main technological difficulty lies in the fact that *M. tuberculosis* might be shielded by host macrophages, so latent and dormant intracellular *M. tuberculosis* is extremely difficult to detect or eradicate (*7,8*). An ideal method of diagnosing TB in clinical practice should have several characteristics: *i*) satisfactory sensitivity and specificity; *ii*) rapid *M. tuberculosis* detection achieved with minimal sample processing; and *iii*) a high sensitivity so that it can be used to screen patients with TB from patients with a suspected *M. tuberculosis* infection (*9,10*). Since conventional methods are far from satisfactory, development of rapid, sensitive, and accurate tools for diagnosis of intracellular *M. tuberculosis* is an urgent task for global TB researchers.

Unlike other bacteria, *M. tuberculosis* has a special cellular structure. The cytoderm of *M. tuberculosis*

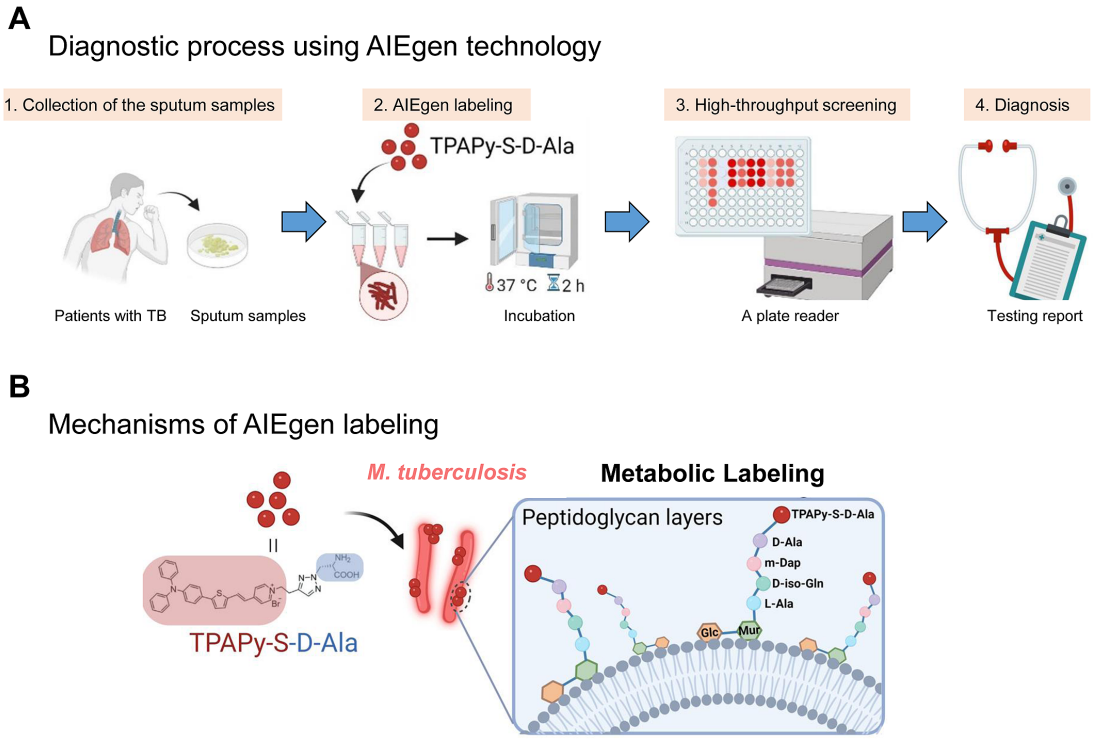

Figure 1. Introduction of a newly developed AIEgen for POC diagnosis of a *Mycobacterium tuberculosis* infection. (A) Process for diagnosis of an *M. tuberculosis* infection using TPAPy-S-D-Ala, an AIEgen with near-infrared fluorescence developed by the current authors' laboratory. Sputum samples were collected and labeled with TPAPy-S-D-Ala at 37°C for 2 h. Then, a plate reader was used to conduct high-throughput tuberculosis screening. **(B)** Underlying mechanisms of TPAPy-S-D-Ala in the metabolic labeling of *M. tuberculosis*. TPAPy-S-D-Ala molecules are integrated into the peptidoglycan layer of the cytoderm and "turn on" the fluorescence of *M. tuberculosis* after incubation for 2 h.

is reported to consist of a cytoplasmic lipid bilayer, peptidoglycan (PG) layer, arabinogalactan layer, mycomembrane, and capsule that contribute to building a solid barrier that facilitates the survival and virulence of *M. tuberculosis* (*11*). However, the complicated structure of the *M. tuberculosis* cytoderm enables the exploration of novel targets for detecting *M. tuberculosis*, such as targets in the PG layer. Specific pentapeptides (L-Ala-D-iso-Gln-m-DAP-D-Ala-D-Ala) are crossed-linked with N-acetylmuramic acid (MurNAc) in the PG layer *via* L, D-transpeptidases catalysis (*12*). Studies have reported that D-Ala-linked fluorescent probes have marked potential to label the PG layer of the *M. tuberculosis* cytoderm by serving as an external D-Ala substrate, enabling visualization of *M. tuberculosis* replication, thereby distinguishing living *M. tuberculosis* from dead cells (*13,14*). However, many of the commercially available fluorophores have a low emission efficiency and low sensitivity in an aggregate state due to the phenomenon of aggregation-caused quenching (ACQ) (*15*). Accordingly, aggregation-induced emission luminogens (AIEgens), which are molecules that are highly emissive in an aggregate state (limited intramolecular motion), had to be developed to address this problem (*16,17*). Thus, development of a metabolic-labeling AIEgen is a potential good solution for accurate point-of-care (POC) diagnosis of an *M. tuberculosis* infection. A previous study by the current authors described an *M. tuberculosis* cytoderm-labeling AIEgen for rapid detection and intracellular ablation of *M. tuberculosis* that had better detection sensitivity compared to conventional acid-fast staining (*17*). AIEgen with near-infrared fluorescence was urgently needed in order to better penetrate so as to facilitate *in vivo* labeling of *M. tuberculosis* in subjects (*18*). Thus, the current authors developed a near-infrared AIEgen (TPAPy-S-D-Ala) for metabolic labeling of PG to enable high-throughput diagnosis of *M. tuberculosis* (Figure 1). The AIE functional motif (TPApy-S) and the metabolic motif (D-Ala, with PG-layer-specific metabolic labeling) are linked using click chemistry to yield TPAPy-S-D-Ala. The entire diagnostic process involving TPAPy-S-D-Ala is shown in Figure 1A and it includes collection of the sputum sample, obtaining of TPAPy-S-D-Ala, incubation of samples with TPAPy-S-D-Ala, high-throughput screening, and finally readout of the diagnostic results (Figure 1A). The mechanisms of incubation are clear but extremely important (Figure 1B), namely, the external D-Ala molecules of the AIEgen are efficiently integrated into PG and are solidly aggregated on the cytoplasmic membrane, resulting "turning on" fluorescence so that the *M. tuberculosis* cells can be labeled. The fluorescence intensity of labeled *M. tuberculosis* can be quantified with a microplate reader to enable the diagnosis of TB (Figure 1).

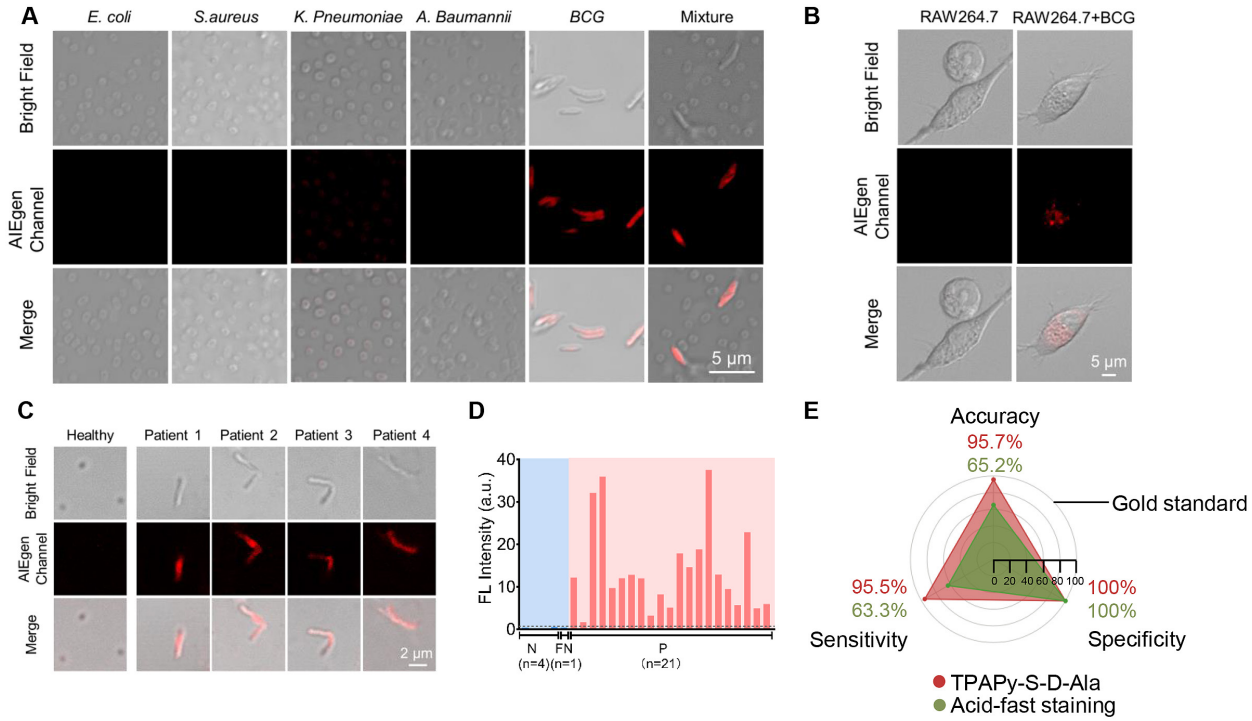

Figure 2. Verification of the labeling of *M. tuberculosis* by AIEgen. (A) Verification of the selective labeling of *M. tuberculosis* by TPAPy-S-D-Ala. Obviously, only samples infected with *M. tuberculosis* (BCG and a mixed infection) were labeled. Microbes were labeled in red by AIEgen. **(B)** Verification of the intracellular effects of TPAPy-S-D-Ala on *M. tuberculosis*. Only cells infected with *M. tuberculosis* were labeled. Microbes were labeled in red by AIEgen. **(C)** Verification of the labeling of *M. tuberculosis* in sputum samples by TPAPy-S-D-Ala. Only *M. tuberculosis* in sputum samples from infected patients was labeled. Microbes were labeled in red by AIEgen. **(D)** Outcomes of *M. tuberculosis* identification by TPAPy-S-D-Ala labeling. A total of 26 sputum samples (22 samples from patients infected with *M. tuberculosis* and 4 samples from healthy controls) were observed. N means negative, FN means false negative, and P means positive. **(E)** Evaluation of the accuracy, sensitivity, and specificity of TPAPy-S-D-Ala labeling in comparison to conventional acid-fast staining of sputum samples.

The metabolic labeling efficiency of TPAPy-S-D-Ala was preliminarily verified in terms of the following aspects: *i*) Selectivity of TPAPy-S-D-Ala The specificity with which TPAPy-S-D-Ala identified *M. tuberculosis* was investigated *via* clinical digestion and incubation with various bacteria. The species used included *Escherichia coli*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, and Bacillus Calmette-Guerin (BCG, a model for *M. tuberculosis*). As shown in Figure 2A, only the panels that included *M. tuberculosis* (BCG and a mixed infection) were found to be labeled by TPAPy-S-D-Ala, indicating that TPAPy-S-D-Ala was highly selective at labeling an *M. tuberculosis* infection. *ii*) Evaluation of the labeling efficiency for intracellular bacteria TPAPy-S-D-Ala was co-incubated with RAW264.7 cells infected with BCG and intact RAW264.7 cells (control) for 2 h. The results were observed using a laser scanning confocal microscope (Zeiss 700, Germany; excitation, 488 nm). Results indicated that the fluorescence intensity of intracellular BCG labeled by TPAPy-S-D-Ala was markedly greater than that of the control (Figure 2B). These results indicated that the AIEgen was effective at labeling intracellular *M. tuberculosis*, so it might be able to serve as a promising tool for diagnosis of latent TB (Figure 2A, B). *iii*) Evaluation of the labeling of *M. tuberculosis* in sputum samples by AIEgen Sputum samples from *M. tuberculosis*-positive patients and healthy controls were collected and exposed to TPAPy-S-D-Ala. As shown in Figure 2C, *M. tuberculosis*-positive sputum samples exhibited a clear "turning on" of fluorescence after TPAPy-S-D-Ala labeling, whereas no fluorescence was observed in *M. tuberculosis*-negative sputum samples from the healthy controls (Figure 2C). The fluorescence intensity of the *M. tuberculosis*-positive sputum samples was significantly greater than that of the sputum samples from the healthy controls (Figure 2D). *iv*) Evaluation of the accuracy, sensitivity, and specificity of the AIEgen A comprehensive gold standard was established to definitively diagnose TB by comprehensively considering clinical manifestations, radiological findings, and laboratory results from a given patient. Conventional acid-fast staining was performed in parallel. Based on the index of fluorescence intensity, TPAPy-S-D-Ala labeling had an accuracy of 95.7%, a sensitivity of 95.5%, and a specificity of 100% in diagnosing an *M. tuberculosis* infection in sputum samples (vs. an accuracy of 65.2%, a sensitivity of 63.3%, and a specificity of 100% for acid-fast staining) (Figure 2E). Hence, the selective labeling of intracellular *M. tuberculosis* and *M.*

tuberculosis in sputum samples by this near-infrared AIEgen (TPAPy-S-D-Ala), along with its accuracy, sensitivity, and specificity, were preliminarily verified. This near-infrared AIEgen was therefore considered to have great potential for use in the clinical diagnosis of TB due to its rapid, sensitive, and precise nature. However, well-designed randomized controlled trials (RCTs) with a large sample are needed to obtain robust evidence for clinical use of this AIEgen, and these RCTs are now being planned.

In conclusion, a near-infrared AIEgen (TPAPy-S-D-Ala) was developed for POC diagnosis of an *M. tuberculosis* infection. TPAPy-S-D-Ala can efficiently aggregate in the PG layer of the bacterial cytoderm by acting as an exogenous substrate of D-amino acid during PG biosynthesis, and this may facilitate the rapid and sensitive detection of living *M. tuberculosis* in macrophages. Preliminary data indicated that this near-infrared AIEgen selectively labels intracellular *M. tuberculosis* and *M. tuberculosis* in sputum samples, with satisfactory accuracy, sensitivity, and specificity in comparison to conventional acid-fast staining. Hence, TPAPy-S-D-Ala is a promising diagnostic tool for rapid and accurate detection of *M. tuberculosis* and it might be a good tool to mass screen for an *M. tuberculosis* infection, particularly in people in developing countries. The efficiency and safety of TPAPy-S-D-Ala should be further verified in large-scale RCTs before this labeling is widely used in clinical practice.

Funding: This work was supported by the National Natural Science Foundation of China (grant nos. 62005176, 82172286, 82070420), the Shenzhen Science and Technology Program (grant nos. JCYJ20210324115611032, JCYJ20220530163005012, and JCYJ20210324131603008), Projects in Key Areas for Universities in Guangdong Province (2022DZX2022). the Shenzhen Scientific and Technological Foundation (grant no. JSGG20220301090005007), and the Third People's Hospital of Shenzhen Foundation (grant nos. G2021027 and G2022062).

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

Ethical issues: This study was approved and supervised by the Ethics Committee of the Third People's Hospital of Shenzhen (approval number 2021-030). The study protocol was explained to all of the participants, who then provided written informed consent to participation in this study.

References

1. Furin J, Cox H, Pai *M. Tuberculosis*. Lancet. 2019; 393:1642-1656.
2. Comin J, Madacki J, Rabanaque I, Zuniga-Anton M, Ibarz D, Cebollada A, Vinuelas J, Torres L, Sahagun J, Klopp C, Gonzalo-Asensio J, Brosch R, Iglesias MJ, Samper S. The MtZ strain: Molecular characteristics and outbreak investigation of the most successful *Mycobacterium tuberculosis* strain in Aragon using whole-genome sequencing. Front Cell Infect Microbiol. 2022; 12:887134.
3. Narayan R, Kundu D, Ghatak A, Tripathi S, Datta S. Efficient elimination of airborne pathogens: A study on aerosolized *Mycobacterium tuberculosis* and SARS-CoV-2 using ZeBox technology. J Hosp Infect. 2022; 129:17-21.
4. Zhang M, Xie Y, Li S, Ye X, Jiang Y, Tang L, Wang J. Proteomics analysis of exosomes from patients with active tuberculosis reveals infection profiles and potential biomarkers. Front Microbiol. 2021; 12:800807.
5. Lu G, Jiang X, Wu A, Zhou J, Liu H, He F, Zhang Q, Zen K, Gu S, Wang J. Two small extracellular vesicle sRNAs derived from *Mycobacterium tuberculosis* serve as diagnostic biomarkers for active pulmonary tuberculosis. Front Microbiol. 2021; 12:642559.
6. Jenkins HE, Ayuk S, Puma D, Brooks MB, Millones AK, Jimenez J, Lecca L, Galea JT, Becerra M, Keshavjee S, Yuen CM. Geographic accessibility to health facilities predicts uptake of community-based tuberculosis screening in an urban setting. Int J Infect Dis. 2022; 120:125-131.
7. Thacker VV, Dhar N, Sharma K, Barrile R, Karalis K, McKinney JD. A lung-on-chip model of early *Mycobacterium tuberculosis* infection reveals an essential role for alveolar epithelial cells in controlling bacterial growth. Elife. 2020; 9:e59961.
8. Yang Q, Chen Q, Zhang M, Cai Y, Yang F, Zhang J, Deng G, Ye T, Deng Q, Li G, Zhang H, Yi Y, Huang RP, Chen X. Identification of eight-protein biosignature for diagnosis of tuberculosis. Thorax. 2020; 75:576-583.
9. Hong JM, Lee H, Menon NV, Lim CT, Lee LP, Ong CWM. Point-of-care diagnostic tests for tuberculosis disease. Sci Transl Med. 2022; 14:eabj4124.
10. Kidwai S, Bouzeyen R, Chakraborti S, et al. NU-6027 inhibits growth of *Mycobacterium tuberculosis* by targeting protein Kinase D and protein Kinase G. Antimicrob Agents Chemother. 2019; 63:e00996-19.
11. Wang X, Yang R, Liu S, Guan Y, Xiao C, Li C, Meng J, Pang Y, Liu Y. IMB-XMA0038, a new inhibitor targeting aspartate-semialdehyde dehydrogenase of *Mycobacterium tuberculosis*. Emerg Microbes Infect. 2021; 10:2291-2299.
12. Tolufashe GF, Sabe VT, Ibeji CU, Ntombela T, Govender T, Maguire GEM, Kruger HG, Lamichhane G, Honarparvar B. Structure and function of L,D- and D,D-transpeptidase family enzymes from *Mycobacterium tuberculosis*. Curr Med Chem. 2020; 27:3250-3267.
13. Lin L, Song J, Du Y, Wu Q, Gao J, Song Y, Yang C, Wang W. Quantification of bacterial metabolic activities in the gut by D-Amino acid-based *in vivo* labeling. Angew Chem Int Ed Engl. 2020; 59:11923-11926.
14. Chen X, Zhang Y, Yuan Q, Li M, Bian Y, Su D, Gao X. Bioorthogonal chemistry in metal clusters: A general strategy for the construction of multifunctional probes for bioimaging in living cells and *in vivo*. J Mater Chem B. 2021; 9:6614-6622.
15. Reza AM, Rakhi SF, Zhu X, Tang Y, Qin J. Visualising the emerging platform of using microalgae as a sustainable bio-factory for healthy lipid production through biocompatible AIE probes. Biosensors (Basel). 2022;

12:208.
16. Ma J, Gu Y, Ma D, Lu W, Qiu J. Insights into AIE materials: A focus on biomedical applications of fluorescence. Front Chem. 2022; 10:985578.
17. Dai G, Luo Y, Liao M, Zhang P, Pan H, Yin T, Yang Q, Zheng S, Liao J, Liu D. A cytoderm metabolic labeling AIEgen for rapid detection and intracellular ablation of *Mycobacterium tuberculosis*. Cell Reports Physical Science. 2023;101238.
18. Yan C, Dai J, Yao Y, Fu W, Tian H, Zhu WH, Guo Z. Preparation of near-infrared AIEgen-active fluorescent probes for mapping amyloid-beta plaques in brain tissues and living mice. Nat Protoc. 2023; 18:1316-1336.

Received April 18, 2023; Revised May 18, 2023; Accepted May 24, 2023.

[§]These authors contributed equally to this work.

**Address correspondence to:*

Wenchang Zhao, Key Laboratory for Nanomedicine, Guangdong Medical University, 1 Xincheng Avenue, Dongguan, Guangdong 523808, China.
E-mail: zhaowenchang@126.com

Mingbin Zheng, Institute of Neurology, National Clinical Research Center for Infectious Diseases, Third People's Hospital of Shenzhen, 29 Bulan Road, Shenzhen, Guangdong 518112, China.
E-mail: mingbinzheng@126.com

Hongzhou Lu, Department of Infectious Diseases, National Clinical Research Center for Infectious Diseases, the Third People's Hospital of Shenzhen, 29 Bulan Road, Shenzhen, Guangdong 518112, China.
E-mail: luhongzhou@fudan.edu.cn

Released online in J-STAGE as advance publication May 27, 2023.

Correspondence

DOI: 10.5582/bst.2023.01124

The priority for prevention and control of infectious diseases: Reform of the Centers for Disease Prevention and Control – Occasioned by "the WHO chief declares end to COVID-19 as a global health emergency"

Mingyu Luo[1,§], Fuzhe Gong[2,§], Jinna Wang[1,*], Zhenyu Gong[1,*]

[1] Department of Communicable Disease Control and Prevention, Zhejiang Provincial Center for Disease Prevention and Control, Hangzhou, Zhejiang, China;
[2] Zhejiang University School of Medicine, Hangzhou, Zhejiang, China.

SUMMARY The novel coronavirus disease 2019 (COVID-19) pandemic has revealed that infectious diseases will present a significant worldwide threat for a long time in the future. Centers for Disease Prevention and Control (CDCs) worldwide have developed for nearly 80 years to fight against infectious disease and protect public health. However, at the advent of the 21th century, the responsibility for prevention and control of infectious diseases has gradually been marginalized in the CDC system. The COVID-19 pandemic has also provided a glimpse into the overburdened operational process and inadequate personnel reserve of the current system of CDCs. In addition, a long-term multisectoral joint mechanism has not been created for sharing information and cooperation to facilitate public health. Reform of the system of CDCs or public health is very necessary. A global prevention and control system should be envisioned and implemented worldwide, and vertical management should be implemented throughout all levels of CDCs to improve their structure and administrative status. The WHO should expand its scope of responsibilities, especially with regard to mechanisms for joint prevention and control of infectious diseases, to substantially implement the "One Health" concept. The International Health Regulations (IHR) and relevant laws and regulations should enshrine the CDC's authority in administration and policy-making to deal with outbreaks or pandemics of infectious diseases.

Keywords reform, infectious disease; Center for Disease Prevention and Control, health policy

At the beginning of 2020, the novel coronavirus disease 2019 (COVID-19) pandemic began in China and then swept around the world. The World Health Organization (WHO) declared the COVID-19 outbreak a global pandemic on March 11, 2020. Three years later, the WHO finally announced that COVID-19 is no longer a global health emergency (*1*) (Figures 1 & 2). The new norm for COVID-19 may be smaller, more frequent, and less fatal outbreaks rather than seasonal outbreaks. Since the COVID-19 pandemic has passed, now is the time to enhance the public health system to respond to normalized management of COVID-19 and potential risks of other emerging infectious diseases (EIDs).

Characteristics of infectious diseases

The ability to spread is the most significant characteristic of infectious diseases. If effective prevention and control strategies are not implemented at critical points, the spread of an epidemic can rapidly reach critical proportions, resulting in substantial harm to public health, socioeconomics, and people's livelihoods.

In history, common infectious diseases have caused enormous harm. During the fourteenth century, the bubonic plague killed about 25 million people (*2*). The third plague pandemic began in Yunnan Province, the border area between China and Burma, and it continued in Yunnan province for 184 years (1772-1995), with nearly one million people dead (*3*). After Spanish colonists brought smallpox to the Americas, the native Indian population in central Mexico was decimated to one tenth of its original size by 1568 (*4*). During the seventh global cholera pandemic, the disease reached its worst level in 1991, and more than 594,000 cases were

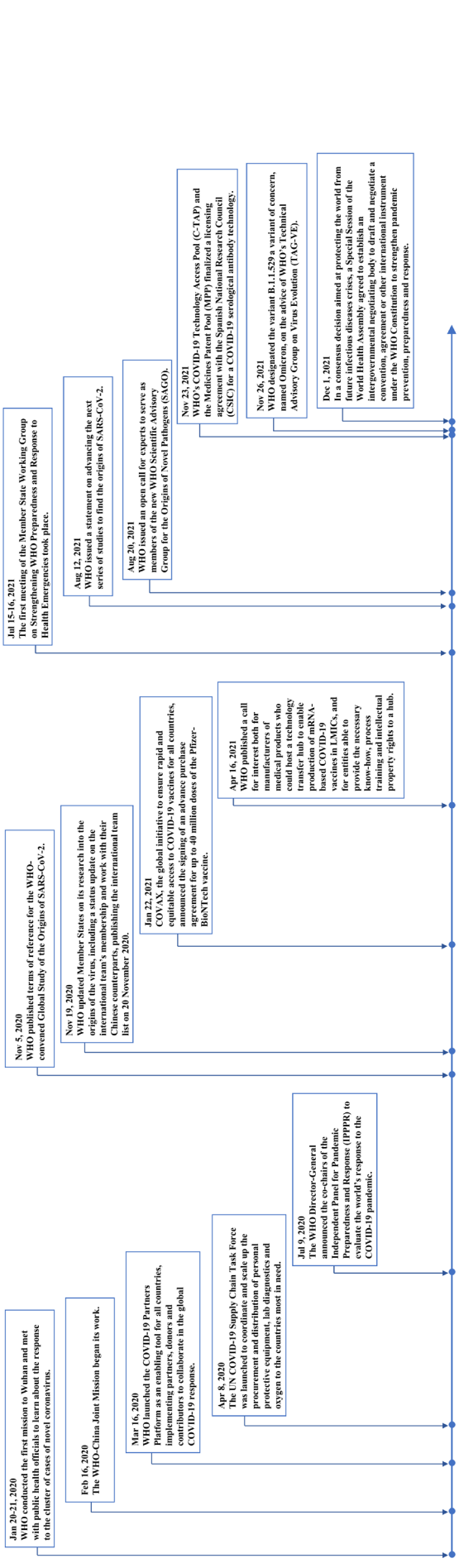

Figure 1. WHO's significant responses to COVID-19 since 2020.

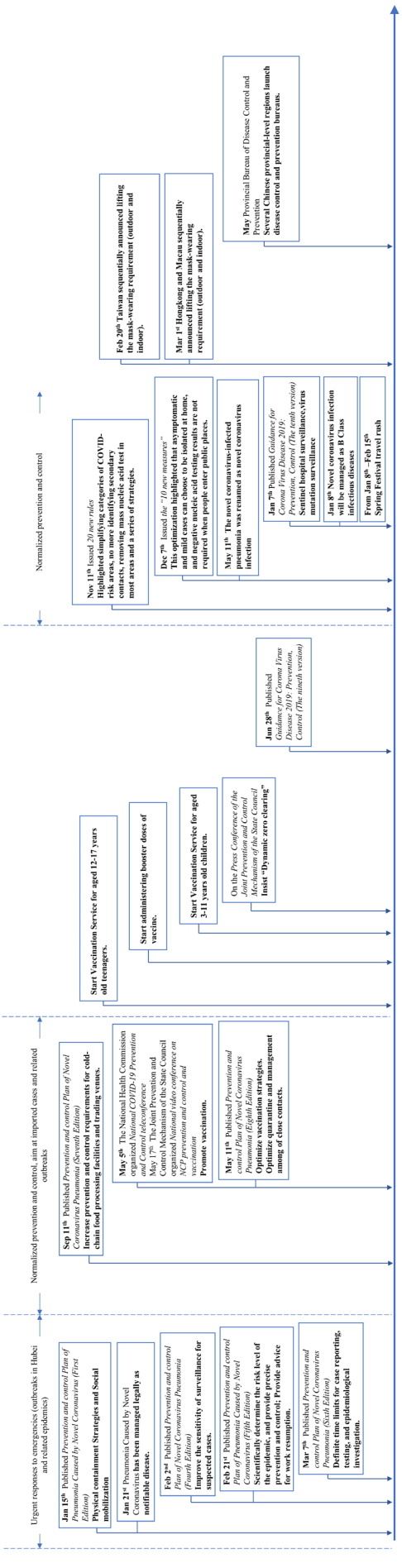

Figure 2. Prevention and control processes and strategies to combat COVID-19 in China since 2020.

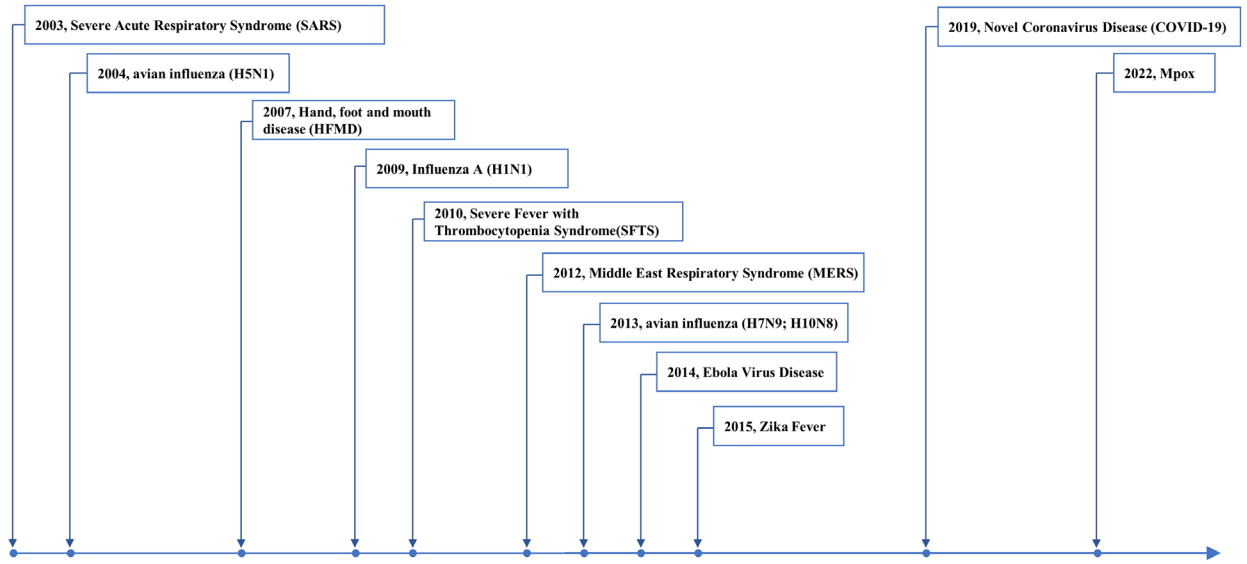

Figure 3. Emerging infectious diseases worldwide since 2000.

reported (*5*). The incidence of cholera remained high throughout the 1990s, and 140,000 to 590,000 cases were reported in 59 to 94 countries and regions around the world (*6*).

Unlike other diseases such as non-communicable chronic diseases (NCDs), the cause of an infectious disease can usually be identified as a single pathogenic microorganism. Vaccines targeting a specific pathogenic microorganism are the most crucial intervention to achieve herd immunity and prevent the spread of infectious diseases. In addition, if a pathogen has yet to be identified or while a vaccine is still be prepared, physical containment strategies such as patient isolation, contact tracing, and personal protective equipment can also intercept transmission and protect public health in the early stage of an outbreak. After thousands of years of efforts, mankind has accumulated vast experience in fighting infectious diseases. Numerous common infectious diseases such as smallpox and poliomyelitis have been effectively controlled and successfully eradicated (*7,8*).

However, EIDs have been attracting increasing attention in the 21st century. (Figure 3). In 2003, an outbreak of severe acute respiratory syndrome (SARS) was identified in China and then spread to four other countries. During the SARS pandemic, 5,327 cases and 349 deaths were reported on the Chinese mainland (*9*). The COVID-19 pandemic continues to ravage the world; as of May 24, 2023, over 766 million confirmed cases and over 6.8 million deaths have been reported globally (*10*) (Figure 4). The public needs to realize that infectious diseases such as respiratory infectious diseases and vector-borne diseases remain a major threat to public health. At the same time, governments should consider the enormous disruption that infectious diseases bring to economic and social stability. Thus, strategies to prevent and control infectious diseases affect not just public health but also political and national security.

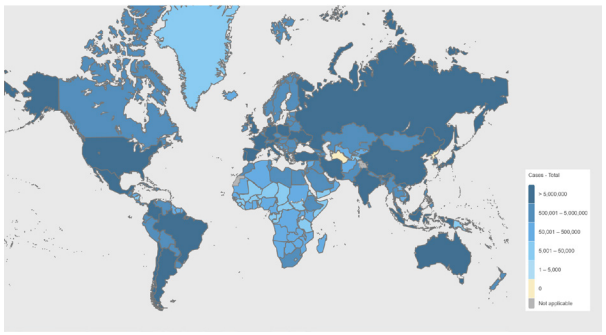

Figure 4. Global confirmed cases of COVID-19 reported to the WHO as May 24, 2023. Globally, as of 6:03 AM CEST, 24 May 2023, there have been 766,895,075 confirmed cases of COVID-19, including 6,935,889 deaths, reported to the WHO.

Current status of and problems with the system of Centers for Disease Prevention and Control

In a classic epidemiological case - the Broad Street cholera outbreak, John Snow made full use of epidemiological methodologies and he introduced comparison, probability, and other epidemiological concepts (*11*). In 1946, the Communicable Disease Center was established in Atlanta in the US. Its primary mission was simply to prevent malaria from spreading across the nation, and that was then expanded to other communicable diseases (*12*). In the 1950s, the Epidemiological Intelligence Service (EIS) was established, and the EIS began systematically training epidemiological field specialists.

In China, the Communist Party of China and the Government emphasize the prevention and control of

infectious diseases. Multiple levels of sanitation and epidemic control stations (at the provincial, prefectural, and county level) were established nationwide by 1953. Through the hard work of three generations of health workers, epidemics of serious infectious diseases such as the plague and cholera were successively controlled, and smallpox was eradicated. Public health has significantly improved in China.

In 1946, the Harvard School of Public Health became an independent, degree-granting body and was no longer affiliated with the medical school (*13*). Since then, the gulf between public health and clinical treatment has become ever wider. With the advent of the 21st century, academics generally believe that the highest proportion of deaths has changed from acute infectious diseases to chronic diseases, and especially in developed countries. The Center for Disease Prevention and Control (CDC) has begun to take on more responsibilities with regard to providing health education and monitoring risk factors for chronic diseases. The prevention and control of infectious diseases has gradually been marginalized, and related departments have been abolished or merged with other departments; departments which are mainly responsible for disinfection and vector control have been merged with other departments. The attrition of specialists and loss of expertise are serious.

Although a long-term multisectoral joint mechanism has prescribed by the Law on the Prevention and Treatment of Infectious Diseases in the People's Republic of China, this mechanism has not been fully created. An uninterrupted mechanism of sharing public health information, such as sharing information between the clinical medical system and CDCs, a channel for sharing information between health administrative departments and other departments (such as the environmental protection department and the agricultural department) has not been created. In addition, CDCs are responsible for disease surveillance and data reporting, but they are not authorized to announce information on the status of infectious diseases.

Suggested reforms for the system of CDCs

This COVID-19 pandemic has provided a glimpse into the overburdened operational process and inadequate personnel reserve of the current system of CDCs. Public health is essential to national stability and essential to global economic development, global political security, and even globalization. Reform of the system of CDCs or public health is very necessary.

(1) A global prevention and control system should be devised and implemented worldwide (*14*). Governments should be mindful of improving the administrative status of CDCs, and vertical management should be implemented throughout all levels of CDCs. In addition, global public health education and creation of teams of public health specialists should be enhanced, and a public health system that includes integration of medical care and prevention in colleges should be created. Clinicians should enhance the awareness of public health; especially when treating patients suggestive of infectious disease, clinicians should pay attention to sources of infection and further proliferation. This would enhance the prevention of infectious diseases in the field of clinical medicine rather than simply enhancing methodologies of clinical epidemiology.

The WHO should expand its scope of responsibilities, especially with regard to joint mechanisms for prevention and control of infectious diseases. The new International Health Regulations (IHR) should be revised to enshrine a joint mechanism of preventing and controlling infectious diseases in the law. The "One Health" concept should be substantially implemented by the WHO, the World Organization for Animal Health (OIE), the Food and Agriculture Organization of the United Nations (FAO), and other authorities to facilitate interdisciplinary, cross-departmental, and cross-regional cooperation and to ensure harmony among human health, animal health, and the state of the environment (*15,16*). CDCs should also search for supports of social forces and explore public mobilization for prevention and control of infectious diseases. For example, the construction of "mosquito-free villages" by Zhejiang Provincial CDC since 2016 has provided a sustainable mechanism supported by whole local villagers to prevent mosquito-borne diseases. Globally, mechanisms of cooperation and sharing information should also be enhanced between CDCs and clinical healthcare systems to combat infectious diseases.

The IHR and the Law on the Prevention and Treatment of Infectious Diseases should enshrine CDC's authority in administration and policy-making to deal with outbreaks or pandemics of infectious diseases.

(2) In China, the mechanism for internal management of the current CDCs should be improved. The goal of the reformed CDCs should be to focus on improving prevention of, detection of, and responses to infectious diseases and to protect public health and socioeconomic development from the threats of infectious diseases (especially EIDs).

(*i*) An elite team should be created for prevention and control of infectious diseases, and the members of this elite team could be individually drawn from current CDCs. The current CDC can be merged into the Administration of Disease Prevention and Control. The "National Directorate General of Health and Epidemic Prevention" can be established at the national level; each province can set up bureaus of Health and Epidemic Prevention (just like combat zones), and each county or district can establish county-level sub-bureaus that directly answer to provincial leadership. Major cities such as sub-provincial cities and cities with separate state planning can establish agencies with personnel dispatched from the province (sub-bureaus).

Hiring requirements should be improved. Quality

field epidemiological investigations, disinfection, vector control and prevention, epidemic analysis and assessment, emergency management, and emergency detection can be performed with specialists in acute infectious disease surveillance at their core; chemical, nuclear, and biological warfare units can be added. This elite team should fall under the direct leadership of the Party Central Committee or be subordinate to the armed police force and have military vertical management. This team should participate in daily efforts to achieve a "Healthy China" and serve society, and it should be supplied efficiently and nationwide after the outbreak of an epidemic.

To ensure that this team is focused on infectious disease surveillance, outbreak detection, and rapid response, it must be subject to preferential policies and have more authority. An independent mechanism for performance evaluation should be created, and the staff's performance should not be evaluated based on areas, papers, or income generated to encourage staff members to improve their ability to identify and solve problems in practice. On March 23, 2023, the General Office of the Central Committee of the Communist Party and the General Office of the State Council published opinions on further improving the medical and health care system. During the handling of outbreaks or pandemics, laws and regulations should give CDCs the administrative right to announce epidemics, to quarantine epidemic sites and areas, and to isolate patients and contacts. Laws and regulations should authorize public health physicians to treat, investigate, and deal with infectious diseases.

(*ii*) National and provincial preventive medicine research needs to be enhanced. Since the COVID-19 pandemic began, the National CDC has made great achievements in virus sequencing, isolation, and elucidation of disease transmission dynamics. However, applied research on epidemic analysis and assessment, decision-making, and handling are still inadequate.

More quality research projects in preventive medicine should be implemented at the national and provincial levels to improve CDCs' ability to rapidly detect pathogens, to screen diagnostic reagents, and to develop vaccines and to provide technical support to control infectious diseases. In 2021, the Zhejiang Provincial CDC and Hangzhou Medical College jointly established a school of public health (*17*). Since 2022, a course on "Prevention and Control of Infectious Disease" has been offered to undergraduates and a course on "Field Epidemiology" has been offered to graduate students. This is a significant reform, introducing the latest developments and infectious disease prevention and control practices in university.

(*iii*) Surveillance and studies on chronic diseases, nutrition, and mental health, and related health education can be conducted by hospitals and colleges. General testing can be left to third-party agencies. Public education about science can also be left to third-party agencies and colleges.

Infectious diseases will present a significant worldwide threat for a long time in the future. To protect the public from infectious diseases, the system of CDCs should lead the way and stand at the forefront of the clinical medical system. Qualitative improvements should be implemented to reform the system of CDCs. The public is looking for an effective and long-term multisectoral mechanism to jointly respond to infectious diseases that can be implemented soon.

Funding: None.

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

References

1. United Nations. WHO chief declares end to COVID-19 as a global health emergency. UN News (2023) *https://news.un.org/en/story/2023/05/1136367* (Accessed May 5, 2023)
2. Kathryn A. Glatter PF. History of the plague: An ancient pandemic for the age of COVID-19. Am J Med.2021; 134:176-181.
3. Xingqi D. Reflection and strategies after the control of plague epidemic in southern China. Chinese J Endem. 2008; 27:593-594.
4. Borah W. America as Model: The Demographic Impact of European Expansion Upon the Non-European World. 1964; 378–387.
5. World Health Organization. Cholera in 1991. Wkly Epidemiol Rec.1992; 67:253-260.
6. Chaochen Luo, Yijun Xie JZ. The situation of cholera epidemics and problems in prevention and control of cholera in China. Infect Dis Inf. 2008; 21:153-154,192.
7. WHO. Smallpox. *https://www.who.int/health-topics/smallpox#tab=tab_1* (Accessed March 24, 2023)
8. WHO. Poliomyelitis. *https://www.who.int/news-room/fact-sheets/detail/poliomyelitis* (Accessed March 24, 2023)
9. Zhao L, Feng D, Ye RZ, Wang HT, Zhou YH, Wei JT, de Vlas SJ, Cui XM, Jia N, Yin CN, Li SX, Wang ZQ, Cao WC. Outbreak of COVID-19 and SARS in mainland China: a comparative study based on national surveillance data. BMJ Open. 2020; 10:e043411.
10. WHO. WHO Coronavirus (COVID-19) Dashboard. 2023. *https://covid19.who.int/ (Accessed May 24, 2022*)
11. Prabhu M. John Snow and the Pump Handle of Public Health. VaccinesWork *https://www.gavi.org/vaccineswork/john-snow-and-pump-handle-public-health* (Accessed March 24, 2023)
12. Centers for Disease Prevention and Control. Our History-Our Story. 2018. *https://www.cdc.gov/about/history/index.html* (Accessed March 14, 2023)
13. TH Chan School of Public Health, Harvard University. Harvard T.H. Chan School of Public Health: About. *https://www.hsph.harvard.edu/about/* (Accessed March 24, 2023)
14. Luo M, Sun J, Gong Z, Wang Z. What is always necessary throughout efforts to prevent and control COVID-19 and other infectious diseases? A physical containment strategy and public mobilization and management. Biosci Trends. 2021; 15:188-191.

15. Luo M, Liu Q, Wang J, Gong Z. From SARS to the Omicron variant of COVID-19: China's policy adjustments and changes to prevent and control infectious diseases. Biosci Trends.2021; 15:418-423.
16. WHO. Quadripartite call to action for One Health for a safer world. 2023. *https://www.who.int/news/item/27-03-2023-quadripartite-call-to-action-for-one-health-for-a-safer-world* (Accessed April 4, 2023)
17. New Education. Hangzhou Medical College and Zhejiang Provincial CDC have jointly established a school of public health and research institute. Hourly News (2021) *https://www.thehour.cn/news/450831.html* (Accessed May 20, 2023) (in Chinese)

Received May 30, 2023; Revised June 17, 2023; Accepted June 20, 2023.

[§]These authors contributed equally to this work.
**Address correspondence to:*
Jinna Wang and Zhenyu Gong, Department of Communicable Disease Control and Prevention, Zhejiang Provincial Center for Disease Prevention and Control, Hangzhou, Zhejiang 310051, China. E-mail: jnwang@cdc.zj.cn (JW), zhygong@cdc.zj.cn (ZG)

Released online in J-STAGE as advance publication June 22, 2023.

Correspondence

DOI: 10.5582/bst.2023.01104

Emerging infectious diseases never end: The fight continues

Yang Yang, Liping Guo, Hongzhou Lu*

National Clinical Research Center for Infectious Diseases, State Key Discipline of Infectious Diseases, Shenzhen Third People's Hospital, Second Hospital Affiliated with the Southern University of Science and Technology, Shenzhen, China.

SUMMARY Emerging infectious diseases have accompanied the development of human society while causing great harm to humans, and SARS-CoV-2 was only one in the long list of microbial threats. Many viruses have existed in their natural reservoirs for a very long time, and the spillover of viruses from natural hosts to humans *via* interspecies transmission serves as the main source of emerging infectious diseases. Widely existing viruses capable of utilizing human receptors to infect human cells in animals signal the possible outbreak of another viral infection in the near future. Extensive and close collaborative surveillance across nations, more effective wildlife trade legislation, and robust investment into applied and basic research will help to combat the possible pandemics of new emerging infectious diseases in the future.

Keywords SARS-CoV-2, coronavirus, emerging infectious diseases, interspecies transmission, zoonosis

For 12,000 years, ever since human hunter-gatherers settled in villages to domesticate animals and cultivate crops, emerging diseases like the "Black Death" (1340s, ~50 million deaths), "Spanish influenza" (1918, ~50 million deaths), and H1N1 "swine flu" (2009, 284,000 deaths) have been threatening human beings and have killed substantial proportions of the population (*1*). Over the past 40 years, constantly emerging and re-emerging infectious diseases have been found almost every year, such as acquired immune deficiency syndrome (AIDS, a sexually transmitted disease), severe acute respiratory syndrome (SARS, a respiratory disease), H1N1 "swine" influenza (a respiratory disease), Middle East respiratory syndrome (MERS, a respiratory disease), and Zika fever (mosquito-borne disease) (*1,2*). The ongoing coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in particular has led to unprecedented challenges to public health and devastating economic losses (*3*). As of 3 May, 2023, SARS-CoV-2 has accounted for more than 765 million infections and 6.9 million deaths globally (*https://covid19.who.int/*). Based on the decrease in COVID-19 deaths, the decline in COVID-19 related hospitalizations and intensive care unit admissions, and the high levels of population immunity to SARS-CoV-2, the World Health Organization (WHO) declared that COVID-19 is now an established and ongoing health issue that no longer constitutes a public health emergency of international concern (PHEIC) on May 5, 2023 (*4*). However, RNA viruses are inherently genetically unstable which allows their rapid evolution, resulting in the predominance of new variants of SARS-CoV-2 with a higher capacity for immune evasion and transmissibility (*5*). Moreover, the zoonotic characteristics of SARS-CoV-2 increase the risk of virus spread and mutation (*6*).

Studies have indicated that the spike (the main target of neutralizing antibodies) protein of SARS-CoV-2 is evolving at twice the rate of a similar protein in seasonal influenza and about ten times as quickly as "seasonal" coronaviruses (*7*). A variety of evidence has also shown that newly emerging variants are better able to evade vaccine-induced, infection-induced, or 'hybrid' immunity and that the protection provided by vaccination was short-lasting (*8,9*). Therefore, scientists believe that the combination of rapid mutation and short-lived human immunity will result in SARS-CoV-2 circulating in mini-waves rather than seasonal surges in the future, so explosive, hospital-filling COVID-19 waves are unlikely to return (*7*). China has experienced one wave dominated by BA.5.2 and BF.7 variants following the adjustment of prevention and control policies in December, 2022 (*10*). As studies have shown that XBB related variants possessed strong antibody escape against breakthrough infections with previous variants including BA.5 and BF.7 (11-13), it seems that the second wave dominated by XBB related variants is ongoing according to the data from our hospital (Figure 1) and GISAID database (*https://www.gisaid.*

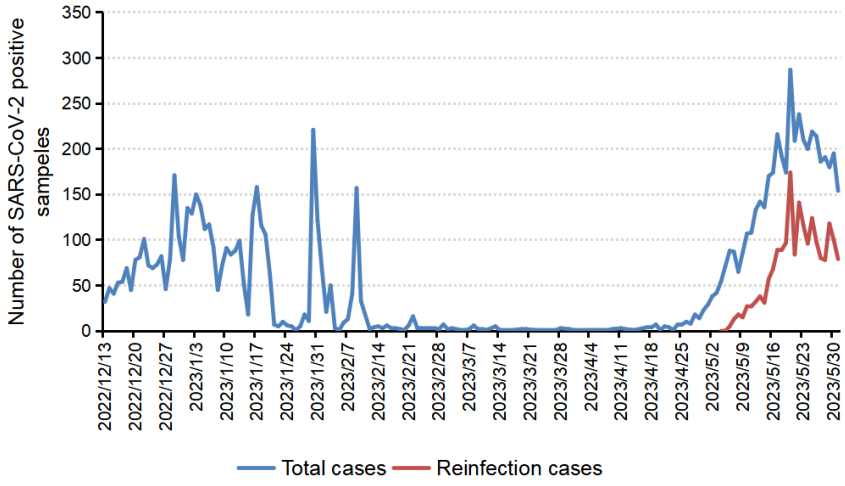

Figure 1. Dynamics of SARS-CoV-2 infections (blue line) and reinfections (red line) in patients diagnosed at Shenzhen Third people's hospital during Dec 13, 2022 and May 31, 2023.

org/). Undoubtedly, the elderly and individuals with an immune deficiency or underlying medical condition should be the focus of the medical system, and more effective vaccines and oral antivirals still serve as key weapons in the battle against SARS-CoV-2 (*14*).

Interspecies transmission plays a key role in the emergence of new human infectious diseases like human coronavirus infections and human avian influenza virus (AIV) infection (Figure 2) (*15,16*). Coronaviruses, for example, are characterized by a higher rate of mutation and recombination than other viruses, facilitating host adaptability and interspecies transmission (*17*). Notably, all known human coronaviruses are believed to be zoonotic pathogens that crossed the species barriers to infect humans based on current sequence databases (*5,15*). SARS-CoV, MERS-CoV, HCoV-NL63, and HCoV-229E are considered to have originated in bats and HCoV-OC43 and HKU1 in rodents (*5,15,18*). Although the accurate reservoir of SARS-CoV-2 has not yet been determined, there is no doubt of its zoonotic origins *via* the wildlife trade (*19,20*). Many coronaviruses that are capable of utilizing human receptors to infect human cells were found to circulate in bats and pangolins during the last decades, such as SARS-CoV, MERS-CoV and SARS-CoV-2 related coronavirus (*21-28*). These coronaviruses are potentially likely to spilled over from an intermediate host, and domestic animals in particular, into human beings (*29*). A novel HKU2-related bat coronavirus named swine acute diarrhea syndrome coronavirus (SADS-CoV) that has been found to infect human cells emerged in Guangdong Province, China in 2016, causing large-scale mortality in piglets on several farms (*30,31*). Recently, human infections with a novel recombinant canine coronavirus were found in East Malaysia and Haiti (*32,33*). These findings signal the possible outbreak of another coronavirus in the near future.

Many viruses have existed in their natural reservoirs

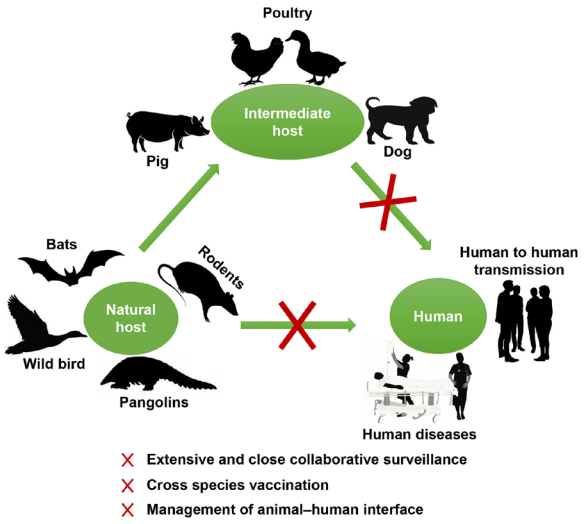

Figure 2. Typical cycle of interspecies transmission and possible countermeasures. Mutation and recombination of viruses during their circulation in natural hosts enable host adaptability and interspecies transmission via an intermediate host or direct contact. Extensive and close collaborative surveillance, cross-species vaccination, and management of animal-human interface are important countermeasures against interspecies transmission.

for a very long time and the spillover of viruses from natural hosts to humans and other animals serves as the main source of emerging infectious diseases (Figure 2) (*1,5,34*). The spread of infectious diseases is not limited by geographic boundaries, so extensive and close collaborative surveillance among nations will be crucial to the prevention and control of future pandemics (*2*). The identification of key reservoir species could enable targeted prophylactic cross-species vaccination campaigns or intervention strategies to minimize contact between natural reservoirs and humans (*28,35*). Over the past three decades, most new human pathogens with substantial impacts on human health or economics

have originated in wildlife, so enactment of legislation effectively addressing the wildlife trade, protection of habitats, and reduction of the wildlife-livestock-human interface is another effective way to prevent future zoonoses (*36,37*). Basic science has brought us many life-saving drugs, vaccines, and diagnostics as it did during the COVID-19 pandemic (*38*). Therefore, robust investment into applied and basic research on virus pathogenesis, interspecies transmission, vaccines, and antivirals is needed to combat possible pandemics of new emerging infectious diseases in the future.

Funding: This work was supported by a grant from Major National Science and Technology Projects (2021YFC2301800) and a grant from the Shenzhen Fund for Guangdong Provincial High-level Clinical Key Specialties (SZGSP011).

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

References

1. Morens DM, Fauci AS. Emerging Pandemic Diseases: How We Got to COVID-19. Cell. 2020; 182:1077-1092.
2. Gao GF. From "A"IV to "Z"IKV: Attacks from Emerging and Re-emerging Pathogens. Cell. 2018; 172:1157-1159.
3. Yang J, Gong Y, Zhang C, Sun J, Wong G, Shi W, Liu W, Gao GF, Bi Y. Co-existence and co-infection of influenza A viruses and coronaviruses: Public health challenges. Innovation (Camb). 2022; 3:100306.
4. WHO. Statement on the fifteenth meeting of the International Health Regulations (2005) Emergency Committee regarding the coronavirus disease (COVID-19) pandemic. *https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic* (accessd May 7 2023).
5. Xia S, Wang L, Zhu Y, Lu L, Jiang S. Origin, virological features, immune evasion and intervention of SARS-CoV-2 Omicron sublineages. Signal Transduct Target Ther. 2022; 7:241.
6. Qiu X, Liu Y, Sha A. SARS-CoV-2 and natural infection in animals. J Med Virol. 2023; 95:e28147.
7. Callaway E. COVID's future: mini-waves rather than seasonal surges. Nature. 2023.
8. Callaway E. The next generation of coronavirus vaccines: a graphical guide. Nature. 2023; 614:22-25
9. Bobrovitz N, Ware H, Ma X, *et al*. Protective effectiveness of previous SARS-CoV-2 infection and hybrid immunity against the omicron variant and severe disease: a systematic review and meta-regression. Lancet Infect Dis. 2023; 23:556-567.
10. Pan Y, Wang L, Feng Z, Xu H, Li F, Shen Y, Zhang D, Liu WJ, Gao GF, Wang Q. Characterisation of SARS-CoV-2 variants in Beijing during 2022: an epidemiological and phylogenetic analysis. Lancet. 2023; 401:664-672.
11. Mykytyn AZ, Rosu ME, Kok A, Rissmann M, van Amerongen G, Geurtsvankessel C, de Vries RD, Munnink BBO, Smith DJ, Koopmans MPG, Lamers MM, Fouchier RAM, Haagmans BL. Antigenic mapping of emerging SARS-CoV-2 omicron variants BM.1.1.1, BQ.1.1, and XBB.1. Lancet Microbe. 2023; 4:e294-e295.
12. Zhu A, Wei P, Man M, Liu X, Ji T, Chen J, Chen C, Huo J, Wang Y, Zhao J. Antigenic characterization of SARS-CoV-2 Omicron subvariants XBB.1.5, BQ.1, BQ.1.1, BF.7 and BA.2.75.2. Signal Transduct Target Ther. 2023; 8:125.
13. Yang Y, Gong X, Tang Y, Liu J, Zeng L, Kuang J, Wang F, Lu H, Liu Y. Naive and breakthrough infections with BA.2, BA.5 and BF.7 variants provide poor cross protection against XBB related variants. J Infect. 2023.
14. Lu H. More effective vaccines and oral antivirals: Keys for the battle against Omicron. Biosci Trends. 2022; 16:1-3.
15. Lu G, Wang Q, Gao GF. Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. Trends Microbiol. 2015; 23:468-478.
16. Shi Y, Wu Y, Zhang W, Qi J, Gao GF. Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza A viruses. Nat Rev Microbiol. 2014; 12:822-831.
17. Khamassi Khbou M, Daaloul Jedidi M, Bouaicha Zaafouri F, Benzarti M. Coronaviruses in farm animals: Epidemiology and public health implications. Vet Med Sci. 2021; 7:322-347.
18. Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, Liu W, Bi Y, Gao GF. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. Trends Microbiol. 2016; 24:490-502.
19. Pekar JE, Magee A, Parker E, *et al*. The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. Science. 2022; 377:960-966.
20. Garry RF. The evidence remains clear: SARS-CoV-2 emerged *via* the wildlife trade. Proc Natl Acad Sci U S A. 2022; 119:e2214427119.
21. Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, Wong SS, Leung SY, Chan KH, Yuen KY. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci U S A. 2005; 102:14040-14045.
22. Li W, Shi Z, Yu M, *et al*. Bats are natural reservoirs of SARS-like coronaviruses. Science. 2005; 310:676-679.
23. Ge XY, Li JL, Yang XL, *et al*. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature. 2013; 503:535-538.
24. Lau SKP, Fan RYY, Zhu L, *et al*. Isolation of MERS-related coronavirus from lesser bamboo bats that uses DPP4 and infects human-DPP4-transgenic mice. Nat Commun. 2021; 12:216.
25. Xiong Q, Cao L, Ma C, *et al*. Close relatives of MERS-CoV in bats use ACE2 as their functional receptors. Nature. 2022; 612:748-757.
26. Chen J, Yang X, Si H, *et al*. A bat MERS-like coronavirus circulates in pangolins and utilizes human DPP4 and host proteases for cell entry. Cell. 2023; 186:850-863.e816.
27. Lam TT, Jia N, Zhang YW, *et al*. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature. 2020; 583:282-285.
28. Letko M, Seifert SN, Olival KJ, Plowright RK, Munster VJ. Bat-borne virus diversity, spillover and emergence. Nat Rev Microbiol. 2020; 18:461-471.
29. Temmam S, Vongphayloth K, Baquero E, *et al*. Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. Nature. 2022; 604:330-336.

30. Zhou P, Fan H, Lan T, *et al*. Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. Nature. 2018; 556:255-258.
31. Yang YL, Qin P, Wang B, Liu Y, Xu GH, Peng L, Zhou J, Zhu SJ, Huang YW. Broad Cross-Species Infection of Cultured Cells by Bat HKU2-Related Swine Acute Diarrhea Syndrome Coronavirus and Identification of Its Replication in Murine Dendritic Cells In Vivo Highlight Its Potential for Diverse Interspecies Transmission. J Virol. 2019; 93:e01448-19.
32. Vlasova AN, Diaz A, Damtie D, Xiu L, Toh TH, Lee JS, Saif LJ, Gray GC. Novel Canine Coronavirus Isolated from a Hospitalized Patient With Pneumonia in East Malaysia. Clin Infect Dis. 2022; 74:446-454.
33. Lednicky JA, Tagliamonte MS, White SK, Blohm GM, Alam MM, Iovine NM, Salemi M, Mavian C, Morris JG. Isolation of a Novel Recombinant Canine Coronavirus From a Visitor to Haiti: Further Evidence of Transmission of Coronaviruses of Zoonotic Origin to Humans. Clin Infect Dis. 2022; 75:e1184-e1187.
34. Tian J, Sun J, Li D, Wang N, Wang L, Zhang C, Meng X, Ji X, Suchard MA, Zhang X, Lai A, Su S, Veit M. Emerging viruses: Cross-species transmission of coronaviruses, filoviruses, henipaviruses, and rotaviruses from bats. Cell Rep. 2022; 39:110969.
35. Warimwe GM, Francis MJ, Bowden TA, Thumbi SM, Charleston B. Using cross-species vaccination approaches to counter emerging infectious diseases. Nat Rev Immunol. 2021; 21:815-822.
36. Borzée A, McNeely J, Magellan K, *et al*. COVID-19 Highlights the Need for More Effective Wildlife Trade Legislation. Trends Ecol Evol. 2020; 35:1052-1055.
37. Gallo-Cajiao E, Lieberman S, Dolšak N, *et al*. Global governance for pandemic prevention and the wildlife trade. Lancet Planet Health. 2023; 7:e336-e345.
38. Edwards AM, Baric RS, Saphire EO, Ulmer JB. Stopping pandemics before they start: Lessons learned from SARS-CoV-2. Science. 2022; 375:1133-1139.

Received May 8, 2023; Revised June 9, 2023; Accepted June 13, 2023.

**Address correspondence to:*
Hongzhou Lu, National Clinical Research Center for Infectious Diseases, State Key Discipline of Infectious Diseases, Shenzhen Third People's Hospital, Second Hospital Affiliated with the Southern University of Science and Technology, Shenzhen, Guangdong 518100 , China.
E-Mail: luhongzhou@fudan.edu.cn

Released online in J-STAGE as advance publication June 16, 2023.

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