Proceedings of the 2023 International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/12/20230296

Research on endoscopic surgery based on SLAM

Zheng Zhang

School of Automotive Engineering, Harbin Institute of Technology (Weihai Campus), Weihai, 264209, China.

2200830127@stu.hit.edu.cn.

Abstract. The Simultaneous Localization and Mapping (SLAM) method is widely used in the positioning and mapping of robots. In the medical field, SLAM is used in auxiliary medical robots and surgical robots. In endoscopic surgery, SLAM performs endoscopic positioning and scene graph construction for the surgical environment based on the information collected by the endoscope. For research on endoscopic SLAM, this article will first introduce the application of SLAM in endoscopic surgery in recent years. This paper summarizes the innovations and future work of relevant literature in recent years and identifies existing problems in SLAM in endoscopic surgery. Next, this article will introduce the combination of deep learning and SLAM in endoscopic surgery and list some specific applications. Finally, this paper will give a prospect for the future application of SLAM in endoscopic surgery. The research in this paper will be of great value to applying SLAM in endoscopic surgery and conducive to the development of future endoscopic SLAM.

Keywords: endoscopy, SLAM, surgery.

1. Introduction

SLAM refers to a mobile robot starting from an unknown location, using sensors to collect information during the movement, locating its pose, and constructing a map of the surrounding environment. The input to SLAM is information collected by the robot's sensors during motion, usually camera images and inertial sensor information [1]. SLAM uses front end visual odometer to estimate camera motion between adjacent images and build local maps [1]. After that, the rear end uses the robot pose and loop closure detection information to perform non-linear optimization to obtain a global map [1]. When the robot moves to an area it has visited before, a loopback is detected, and the rear end processes the information [1]. (The currently widely used loop closure detection method is the bag of words model: the model uses machine learning methods such as k-means clustering to extract image features, builds a dictionary, and forms a bag of words vector. Loop closure detection can be done by comparing two vectors in the model.) Finally, the result of the mapping is output [1]. The classic SLAM framework is shown in Figure 1. Classical SLAM is very mature in environments with static, rigid bodies, no significant changes in lighting, and no human interference. SLAM mapping results include the Observational Models, Motion Models, Landmark, and Position & Orientation.

DOI: 10.54254/2755-2721/12/20230296

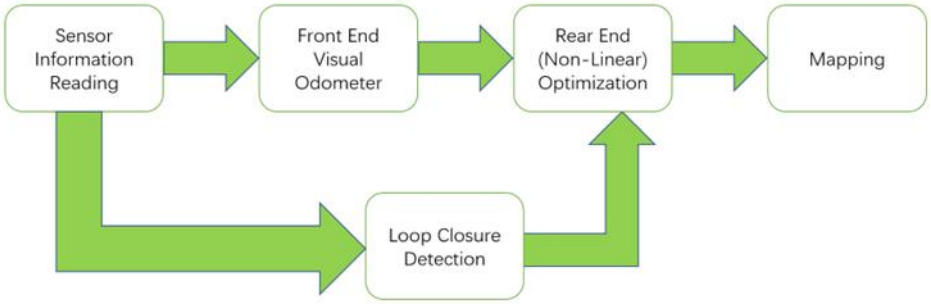

Figure 1. Classic SLAM framework [1].

The earliest application of SLAM in medical image processing is monocular endoscope registration computed tomography scanning. The traditional SLAM algorithm is suitable for environments with static, rigid bodies, no apparent changes in illumination, and no human interference. However, the surgical environment involves non-rigid deformation of tissues caused by breathing and heartbeat activities, clutter caused by surgical instruments, sudden movement of visual sensors, changes in lighting conditions, etc. Traditional SLAM methods do not work well. As a result, improved methods like Extended Kalman Filter SLAM (EKF-SLAM) were born, which are robust to environmental disturbances. Today, with the development of augmented reality and deep learning, these new technologies are combined with SLAM and serve medical care.

2. SLAM application in endoscopic surgery

Huo et al. proposed a real-time dense reconstruction method for binocular endoscopes based on StereoNet, Oriented FAST and Rotated BRIEF SLAM (ORB-SLAM) [2]. The method improves the binocular endoscope scene's reconstruction accuracy and point cloud density while ensuring computational efficiency and real-time performance [2]. The method meets real-time performance requirements on the stomach model and actual pig stomach [2].

Rodr´ıguez et al. used a new SLAM method of monocular deformation to track camera installation and scene deformation simultaneously [3]. The qualitative human colonoscopy results from the EndoMapper demonstrate that the method can successfully cope with deformation, low texture, and strong light variations in actual endoscopies [3].

Batlle et al. achieved in-vivo 3 Dimensions (3D) reconstruction of the human colon using photometric stereo imaging on a calibrated monocular endoscope using controlled illumination during colonoscopy [4]. It solves the problem that medical endoscopes can only provide monocular images due to space constraints, and the system lacks a real scale [4].

Liu et al. developed a SLAM system based on the combination of learned appearance, optimizable geometric priors, and factor graph optimization [5]. It tracks endoscopes and reconstructs the dense 3D geometry of the anatomy seen from monocular endoscopy videos [5]. The method is robust to the lack of texture and lighting variations in endoscopy, with good generalization [5].

Yang et al. propose a semantic description method using a scene graph, which combines contour features and Scale Invariant Feature Transform (SIFT) features and applies it to the Structure-from-Motion (SfM) reconstruction framework [6]. The new semantic feature description can reveal more accurate and dense feature correspondences and provide local semantic information in feature matching [6]. This method can achieve ideal performance in soft tissue deformation, smooth surfaces, lack of texture and other environments [6].

Wu et al. proposed an improved endoluminal SLAM method [7]. Combining with deep learning semantic segmentation to remove feature points on surgical instruments [7]. The method reduced the impact of dynamic surgical instruments [7]. The methodology results in more stable and accurate mapping than standard SLAM systems [7]. The method can avoid interference of surgical instrument motion in the chamber on the SLAM algorithm [7].

DOI: 10.54254/2755-2721/12/20230296

Recasens et al. utilize an approach for estimating camera poses of 6 degrees-of-freedom and dense 3D scene models from monocular endoscopy sequences, Endo-Depth-and-Motion [8]. The method was extensively experimentally evaluated on Hamlyn, producing 3D reconstructions of body cavities in poorly lit, unstable, texture-less scenes [8].

Lamarca et al. close-range sequences of deformed scenes are processed using DefSLAM, Shape-from-Template (SfT) and Non-Rigid Structure-from-Motion (NRSfM) techniques to generate accurate 3D models of the scene relative to a moving camera [9]. Compared with monocular ORB-SLAM, DefSLAM has better accuracy and robustness [9].

Xie et al. improved the classic SLAM algorithm regarding pose adjustment and spatial point positioning calculation [10]. The gastrointestinal tract SLAM algorithm framework is constructed by introducing the local pose optimization algorithm based on the visual correlogram and the minimum geometric distance triangulation algorithm [10]. The algorithm performs well regarding computational efficiency, localization, and feature map construction accuracy [10].

Yang et al. established a new monocular endoscopy depth estimation method with gradient loss [11]. They introduced average loss to explicitly express the sensitivity to periodic small structures [11]. A loss of geometric consistency is proposed to extend the spatial information to the whole sampling grid to constrain the overall geometric anatomy [11]. The method can generate consistent depth maps and plausible anatomy [11].

Table 1. Summary of SLAM literature in recent years.

advantage	future work	Ref.
Reconstruction accuracy; Cloud density; Computational efficiency; Real-time performance	Network structure; Parameter settings; More endoscopic scenes	[2]
Planar scenes; Deformation; Low texture; Strong lighting changes	Deformable SLAM system; Navigation; Augmented reality	[3]
Large-scale monocular SLAM	Photometric calibration; Depth estimation; Deep learning	[4]
Robust; Texture sparsity; Illumination changes	Representative dataset; Generalizability; Non-static scenes	[5]
Feature-matching methods; Soft tissue deformation; Smooth surfaces	Combined with existing SLAM	[6]
Semantic segmentation; Feature points; Surgical instruments	Intraluminal soft tissue deformation model; Densification surgery; Less intraluminal data	[7]
Robust photometric odometry; Poorly lit; Unstable; Texture-less scenes	SfM/SLAM; Stereo loss; Deformable models	[8]
Robustly initialized; Monocular sequence; Scene deformations; Scene estimates	Medical images; Long periods; Multiple moving; Deforming objects	[9]
Computational efficiency; Feature map construction; Accuracy	Semi-dense algorithm; Dense algorithm; Prior information; Increase frame rate; Compressive sensing	[10]
Depth estimation; Monocular endoscopy; Geometry-aware loss; Comprehensive geometric representation	Photometric model; Absolute scale; Scale ambiguity; Computational intelligence algorithms	[11]

A summary of the above literature is shown in Table 1. Through the collation of relevant literature, this paper summarizes the existing problems of SLAM in endoscopic surgery. Surgery has real-time requirements for SLAM. Usually, it needs to use the information collected by the endoscope for real-time positioning and mapping. In endoscopic surgery, SLAM faces complex surgical environments: lack of textured surfaces, changes in lighting conditions, deformation of soft tissues, movement of cameras and surgical instruments, etc. The complex environment puts forward requirements on the robustness of SLAM. SLAM must have the ability to migrate between different patients of the same

DOI: 10.54254/2755-2721/12/20230296

type of surgery, as well as the ability to migrate between different types of surgery. These two migration capabilities are requirements for the generalization of SLAM. The SLAM system must have a high enough accuracy to be applied in an actual surgical environment. Monocular SLAM will face the problems of scale ambiguity and scale offset, and it also involves estimating depth information. In recent years, relevant literature has improved SLAM for the above problems. However, it is not easy to establish a SLAM system with real-time performance, robustness, generalization and accuracy. Therefore, future work is establishing a SLAM system with the above performance. Apply a fully functional system to endoscopic surgery.

3. Deep learning SLAM and endoscopic surgery

Ali pointed out the overall framework and common tasks of deep learning in endoscopic surgery [12]. A neural network is composed of neurons [12]. Under the training set data, the network weights are continuously updated in iterations [12]. Various tasks can be performed by applying test set data to a fully trained network [12]. In endoscopic surgery, neural networks are usually used to perform classification tasks, detection and localization tasks, semantic segmentation tasks, instance segmentation tasks, depth estimation tasks [12]. The framework and tasks of deep learning in endoscopic surgery are shown in Figure 2.

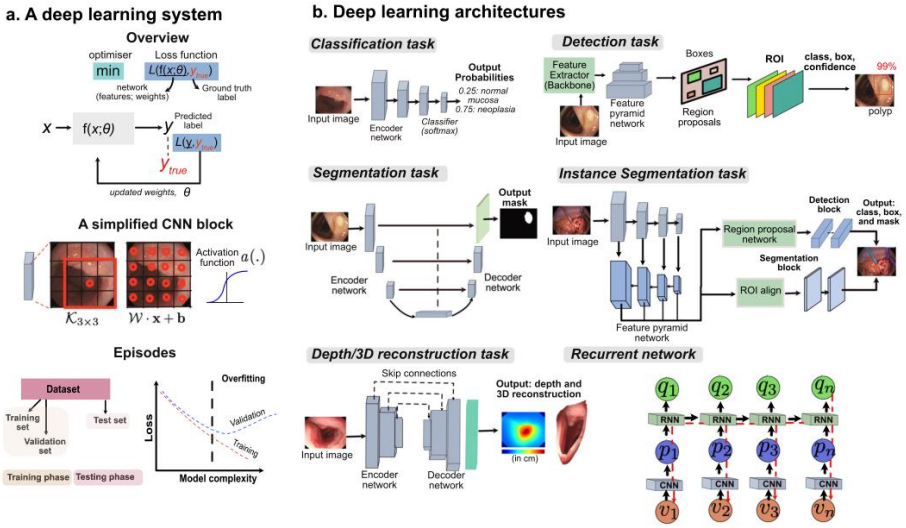

Figure 2. The framework and tasks of deep learning in endoscopic surgery [12].

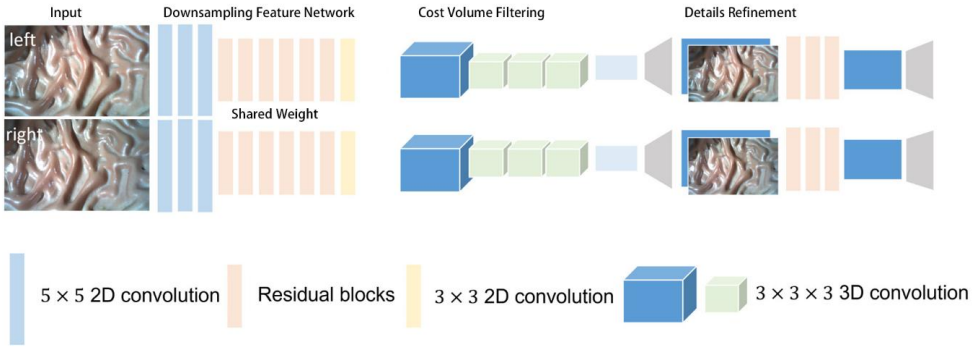

Figure 3. StereoNet for depth prediction [2].

DOI: 10.54254/2755-2721/12/20230296

StereoNet uses low-resolution feature maps and edge-aware upsampling modules to achieve high-precision real-time disparity prediction and obtains predicted depth maps, which perform well in weakly textured areas and are computationally efficient [2]. StereoNet for depth prediction is shown in Figure 3.

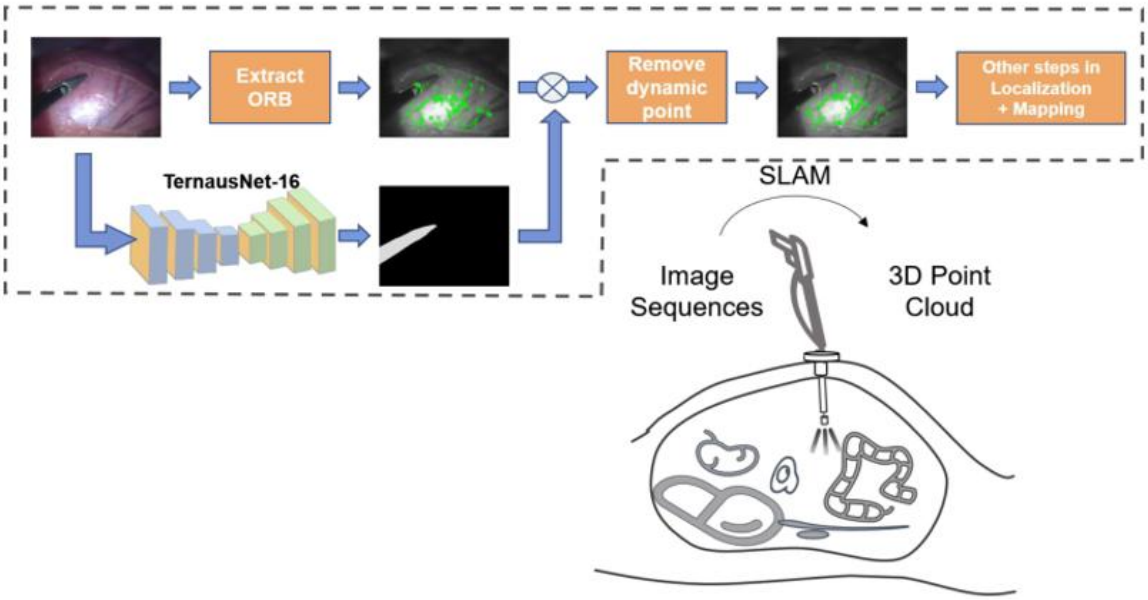

Figure 4. SLAM framework combined with semantic segmentation [7].

In the SLAM process, the semantic segmentation based on the convolutional neural network is introduced to ensure that the dynamic feature points on the surgical instrument are eliminated in the tracking module and that the wrong data association information is avoided [7]. The method was located and mapped using only characteristic points from areas other than surgical instruments [7]. The impact of dynamic objects, such as surgical instruments, on the system can be significantly reduced [7]. SLAM framework combined with semantic segmentation is shown in Figure 4.

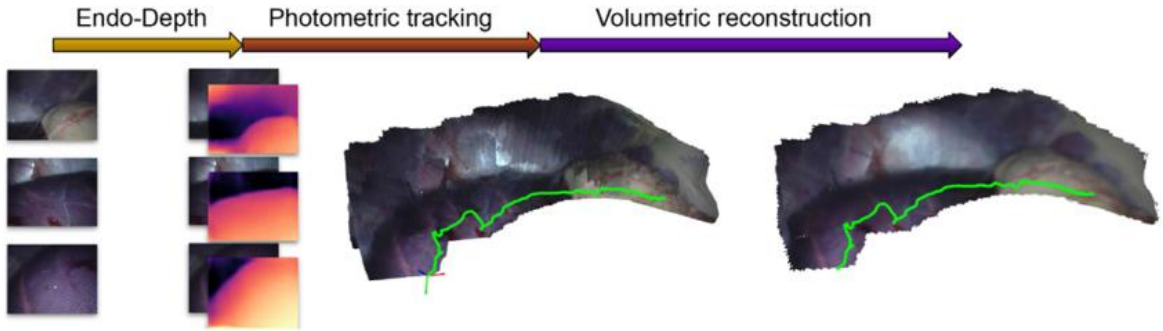

Figure 5. Endo-Depth-and-Motion for depth prediction [8].

Pixel depth prediction for monocular video keyframes using deep neural networks [8]. The movement of each image for the nearest keyframe is estimated by minimizing the photometric error, robustized using image pyramids and a robust error function [8]. Furthermore, fuse the depth map of the keyframe in the volumetric representation based on Truncated Signed Distance Function (TSDF) [8]. Endo-Depth-and-Motion for depth prediction is shown in Figure 5.

DOI: 10.54254/2755-2721/12/20230296

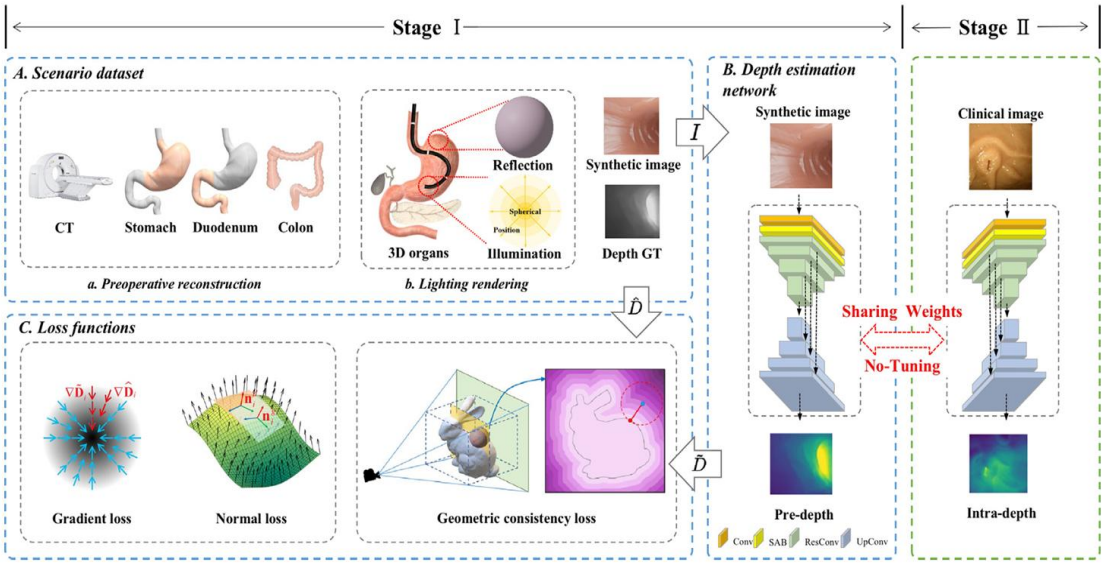

Figure 6. Network for the Geometry-Informed depth estimation system [11].

The proposed framework consists of two stages: the scene dataset, depth estimation network, and loss function [11]. The second stage is a zero-shot application of the trained deep network on clinical images, and the second stage shares weights with the first stage [11]. Network for the Geometry-Informed depth estimation system is shown in Figure 6.

4. Conclusion

As an essential technology in robotics, SLAM plays a vital role in practically applying robots. In the current medical field, robot-assisted medicine. In endoscopic surgery, SLAM uses the information gathered by the endoscope to position the endoscope and maps the surgical environment. In the actual surgical environment, there are many problems, such as the deformation of human soft tissues, the clutter of surgical instruments, sudden movement of visual sensors, and changes in lighting conditions. Traditional SLAM methods are not applicable. Therefore, it is necessary to develop SLAM suitable for the surgical environment.

In order to solve the current problems of endoscopic SLAM, this paper summarizes the relevant articles that have improved SLAM in terms of real-time, robustness, generalization, and accuracy in recent years. The merits of related articles and their future work are reviewed. In addition, since the combination of deep learning and SLAM technology has become a future development trend, this paper conducts an in-depth discussion on the combination of deep learning technology and endoscopic SLAM. The application method of deep learning technology in SLAM-based endoscopic surgery is analyzed.

The research in this paper is of great significance for applying SLAM in endoscopic surgery. In the future, researchers will continue to improve SLAM. Emerging fields such as deep learning and augmented reality will be highly integrated with SLAM. SLAM will increasingly fit the endoscopic surgical environment. When a kind of endoscopic surgical SLAM with many advantages such as real-time, robustness, generalization, and accuracy is born, it will serve as a milestone in the SLAM and medical fields, bringing great convenience to surgery.

References

[1] Gao, X. (2017). 14 Lectures on Visual SLAM: From Theory to Practice. (Beijing: Publishing House of Electronics Industry)

[2] Huo, J., Zhou, C., Yuan, B., Yang, Q., & Wang, L. (2023). Real-Time Dense Reconstruction

with Binocular Endoscopy Based on StereoNet and ORB-SLAM. Sensors, 23(4), 2074. https://doi.org/10.3390/s23042074

[3] Gomez Rodriguez, J. J., Montiel, J. M. M., & Tardos, J. D. (2022). Tracking monocular camera pose and deformation for SLAM inside the human body. 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 5278–5285. https://doi.org/10.1109/IROS47612.2022.9981203

[4] Batlle, V. M., Montiel, J. M. M., & Tardos, J. D. (2022). Photometric single-view dense 3D reconstruction in endoscopy. 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 4904–4910. https://doi.org/10.1109/IROS47612.2022.9981742

[5] Liu, X., Li, Z., Ishii, M., Hager, G. D., Taylor, R. H., & Unberath, M. (2022). SAGE: SLAM with Appearance and Geometry Prior for Endoscopy. 2022 International Conference on Robotics and Automation (ICRA), 5587–5593. https://doi.org/10.1109/ICRA46639.2022.9812257

[6] Yang, Z., Pan, J., Li, R., & Qin, H. (2022). Scene-graph-driven semantic feature matching for monocular digestive endoscopy. Computers in Biology and Medicine, 146, 105616. https://doi.org/10.1016/j.compbiomed.2022.105616

[7] Wu, H., Zhao, J., Xu, K., Zhang, Y., Xu, R., Wang, A., & Iwahori, Y. (2022). Semantic SLAM Based on Deep Learning in Endocavity Environment. Symmetry, 14(3), 614. https://doi.org/10.3390/sym14030614

[8] Recasens, D., Lamarca, J., Facil, J. M., Montiel, J. M. M., & Civera, J. (2021). Endo-Depth-and-Motion: Reconstruction and Tracking in Endoscopic Videos Using Depth Networks and Photometric Constraints. IEEE Robotics and Automation Letters, 6(4), 7225–7232. https://doi.org/10.1109/LRA.2021.3095528

[9] Lamarca, J., Parashar, S., Bartoli, A., & Montiel, J. M. M. (2021). DefSLAM: Tracking and Mapping of Deforming Scenes From Monocular Sequences. IEEE Transactions on Robotics, 37(1), 291–303. https://doi.org/10.1109/TRO.2020.3020739

[10] Xie, C., Yao, T., Wang, J., & Liu, Q. (2020). Endoscope localization and gastrointestinal feature map construction based on monocular SLAM technology. Journal of Infection and Public Health, 13(9), 1314–1321. https://doi.org/10.1016/j.jiph.2019.06.028

[11] Yang, Y., Shao, S., Yang, T., Wang, P., Yang, Z., Wu, C., & Liu, H. (2023). A geometry-aware deep network for depth estimation in monocular endoscopy. Engineering Applications of Artificial Intelligence, 122, 105989. https://doi.org/10.1016/j.engappai.2023.105989

[12] Ali, S. (2022). Where do we stand in AI for endoscopic image analysis? Deciphering gaps and future directions. Npj Digital Medicine, 5(1), 184. https://doi.org/10.1038/s41746-022-00733-3