A Trustworthy Decentralized System for Health Data Integration and Sharing: Design and Experimental Validation

Ruichen Cong[1], Yaping Ye[2], Jianlun Wu[2], Yuxi Li[1], Yuerong Chen[2], Yishan Bian[2], Kiichi Tago[3], Shoji Nishimura[4], Atsushi Ogihara[4], and Qun Jin[4](✉)

[1] Graduate School of Human Sciences, Waseda University, Tokorozawa, Japan
carriecong@moegi.waseda.jp
[2] Zhejiang Chinese Medical University, Hangzhou, China
[3] Department of Information and Network Science, Chiba Institute of Technology, Narashino, Japan
[4] Faculty of Human Sciences, Waseda University, Tokorozawa, Japan
jin@waseda.jp

Abstract. Personal health data collected via wearable devices can be used for sharing and utilization to provide smart healthcare services. Since personal health data involves sensitive information, it is necessary to require a secure way to manage and use data with the consent of each individual. To integrate and share health data securely, many frameworks using federated learning and blockchain-based system have been proposed. However, the issues of ensuring data ownership and enhancing privacy protection remain to be solved. In this paper, we propose a trustworthy system for health data integration and sharing enabled by decentralized federated learning. We describe the major functions and features, including health data integration, doubling data ownership, data analysis via decentralized federated learning, and incentive mechanisms. We further introduce the experiment and assume two application scenarios for sharing and utilization of personal health data and visualization feedback to users. Various types of health data are collected and integrated into the system with decentralized data analysis while sharing results and models and reducing data transmission for privacy-preserving. The proposed system can be expected to provide an effective way to integrate and analyze personal health data for personalized smart healthcare.

Keywords: Personal health data · Healthcare · Blockchain · Privacy protection · Decentralized Federated Learning · Data integration

1 Introduction

In recent years, the growth of IoT services and other data collection and utilization practices has greatly enhanced our daily lives and brought many benefits in fields such as healthcare and medical services. The continuous production of personal health data

A. Moallem (Ed.): HCII 2023, LNCS 14045, pp. 125–134, 2023.
https://doi.org/10.1007/978-3-031-35822-7_9

(PHD) from wearable devices and sensors has led to its widespread use for personalized healthcare analysis and the promotion of health and well-being [1]. However, there are growing concerns about privacy-preserving and how to make the proper protection and management of collected and stored data. Therefore, individuals who utilize such information face the challenge of securing and maximizing the benefits of their data while also addressing privacy concerns.

To address these challenges, various privacy regulations have been established globally, including the European Union General Data Protection Regulation (GDPR) [2], the California Consumer Privacy Act (CCPA) [3], and Singapore's Personal Data Protection Commission (PDPC) [4], among others. These regulations highlight the increasing importance of privacy protection in data management. The GDPR, in particular, imposes significant legal penalties. Therefore, for sensitive data sharing such as PHD, it is necessary to require a secure way to manage and use data with the consent of each individual. However, there are still challenges to be overcome, including concerns with ensuring data ownership and enhancing privacy protection.

To solve these problems, many solutions using blockchain and privacy computing have been proposed [5, 6]. In our previous work [7–9], we proposed a novel model of Individual-Initiated Auditable Access Control (IIAAC) for privacy-preserved data sharing based on blockchain, Ciphertext-Policy Attribute-Based Encryption (CP-ABE) and InterPlanetary File System (IPFS). We further implemented the secure interoperation of blockchain and IPFS through a client application.

In this paper, we propose a trustworthy system enabled by decentralized federated learning in IIAAC. Federated learning is a method for collaborative machine learning, which only shares training models while maintaining all the target data on the decentralized nodes without centralized data collection [10]. By coordinating multiple nodes to execute machine learning, federated learning is suitable for health data sharing and utilization while enhancing privacy protection [11].

The remainder of this paper is organized as follows. In Sect. 2, related work on blockchain and federated learning are overviewed, and privacy protection issues on data sharing are identified. In Sect. 3, we introduce the functions and features of our proposed system and present the basic system architecture using decentralized federated learning. In Sect. 4, our simulated experiment is described, and two scenarios for health-related data sharing and utilization are assumed. Thereafter, a set of individualized visualization results are shown. Finally, this paper is summarized, and future directions are highlighted in Sect. 5.

2 Related Work

In this section, we briefly introduce the issues of health data sharing and utilization. Then, we present federated learning, blockchain, and previous related works conducted in data integration and sharing based on federated learning and blockchain platforms.

With the rapid growth of the Internet of Health Things (IoHT) in health infrastructures, large amounts of health-related data are being collected and processed in storage data centers [12]. However, securing a vast amount of health data presents a significant challenge, and it is necessary to implement innovation with privacy-preserving health data solutions.

In recent years, privacy computing has emerged as a promising privacy-preserving technology. Gartner has recognized privacy-enhancing computation as a top strategic technology trend for consecutive years 2021 and 2022 [13]. Privacy computing encompasses techniques, such as multi-party computation, differential privacy, and federated learning. Especially federated learning enables machine learning analysis without data aggregation and public transmission. In a federated learning platform, data remains stored locally while models are trained and shared in multiple nodes, preserving the privacy of the data owner [10]. Therefore, federated learning enables data processing and analysis locally while reducing the transmission of original data, allowing for secure and privacy-protected sharing of models.

However, federated learning may also bring privacy risks, such as malicious users can obtain sensitive information through the inference of data during the training process [14], and data leakage could occur when the central server or client is compromised [15].

To solve these problems, blockchain technology has been used as a complement to federated learning, which allows for trustworthy identity authentication, tamper-proof data storage, and decentralization. To counter global aggregation attacks and distributed poisoning attacks in federated learning, many frameworks and approaches using blockchain have been proposed. Heiss et al. [16] proposed a blockchain-based federated learning, which uses zero-knowledge proofs to verify off-chain computations and prove the correctness of parameters. It remains the federated learning framework with a central server. To address these several key issues, such as ensuring the reliability and quality of distributed data and considering how to motivate data owners to share data with others by using an incentive mechanism. In our previous work [8, 9], we proposed a novel model of Individual-Initiated Auditable Access Control (IIAAC) in a consortium blockchain-based system incorporating CP-ABE and IPFS. We further implemented secure interoperation of blockchain and IPFS through a client application. Based on our previous work, this paper focuses on data integration and sharing in a trustworthy way enabled by decentralized federated learning in IIAAC.

3 A Trustworthy Decentralized System for Health Data Integration and Sharing

In this section, we first introduce the functions and features of our proposed system. Then, we describe the prototype system using decentralized federated learning in IIAAC.

3.1 System Requirements

To integrate and share health data effectively and securely, a trustworthy decentralized system enabled by blockchain and decentralized federated learning is designed. The major functions and features of our proposed system are summarized as follows.

1) **Health data integration**

   Various types of health-related data are collected and integrated into the system, including health features collected via a wearable device related to bio indicators (e.g., heart rate and blood oxygen), sleep indicators (e.g., sleep score, deep sleep continuity, wake-up counts, and breathing quality) and activity indicators (e.g., step number, step distance, activity calories consumption, and moderate to high-intensity activity duration). In addition, health data in terms of Traditional Chinese Medicine (TCM) are obtained and used, which is important for predictive health risk analysis [17, 18]. These data are collected and integrated into the system for sharing and utilization in a trustworthy and secure way.

2) **Doubling data ownership**

Users in the system are classified into two types: the data owner, who generates the data, and the data requester, who uses the data. In certain situations, health data can be owned by plural users. For instance, an electronic medical record (e.g., TCM health data) is usually created by a medical staff or a medical device as the data owner. However, such a kind of health data is important for a person to manage his/her health through health data analysis. In our previous work, we allowed the system to set double ownership for the person, which is implemented by smart contracts in blockchain [19].

3) **Data analysis via decentralized federated learning**

By using the decentralized federated learning mechanism, machine learning algorithms are applied to data analysis while maintaining data decentralization. The local model and the global model are used for further personalized analysis.

4) **Incentive mechanism**

To motivate data owners to participate in data integration and sharing positively, we incorporate an incentive mechanism that provides individualized feedback to a data owner who shares the data. The individualized feedback includes comparative analysis results with a peer user or a group of users.

3.2 System Architecture

In this section, we describe the basic system architecture in IIAAC by using Hyperledger Fabric, a consortium blockchain, an IPFS distributed file system, a CP-ABE encryption mechanism, and a decentralized federated learning platform to interoperate with each other.

In a decentralized federated learning mechanism, we use the local model and the global model for personal health data analysis and feedback on the analysis results to data owners. The system architecture designed in our proposed model mainly consists of two categories of users, which are data owner and data requester. In our proposed architecture, personal health data is kept on local devices (decentralized nodes) instead of being centralized on a server or transferred data to otherwhere. In addition, important information related to the training process and model parameters in the federated learning process can be securely stored on on-chain storage and off-chain storage via blockchain. The basic architecture of the prototype system is shown in Fig. 1.

4 Experiment and Application Scenarios

In this section, we first describe the simulated experiment. Then, we assume two application scenarios for sharing and utilization of personal health data and visualization feedback to users.

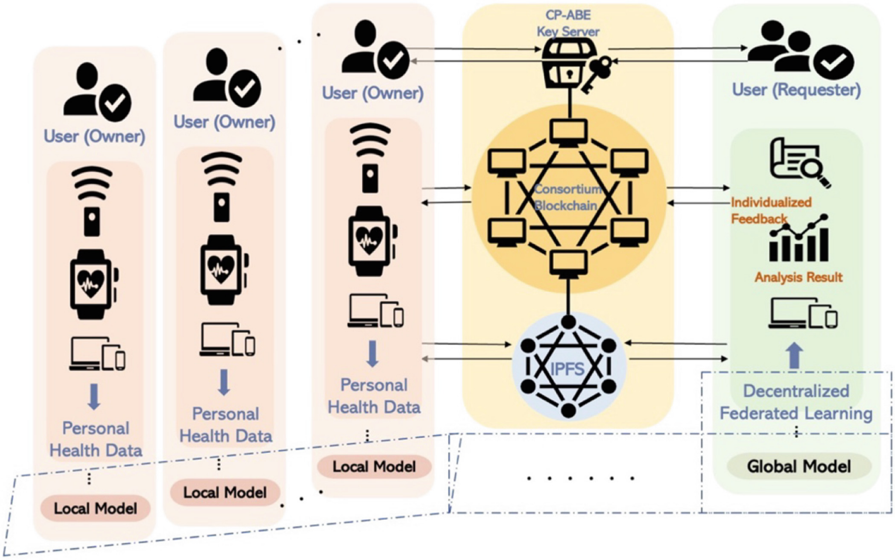


Fig. 1. Basic System Architecture

4.1 Experiment Overview

Wearable devices were used to monitor and record the daily health data of 22 recruited participants for 50 days and collecting data every day. Each participant was requested to record Self-assessed Subjective Health Score (SSHS) every day [20]. This study was conducted under the approval of the Ethics Review Committee on Research with Human Subjects of Waseda University, Japan (No. 2018-092), and all subjects for this experiment signed the informed consent.

In this paper, we design and conduct a simulated experiment based on the proposed system for health data integration and sharing using a part of the collected data, as mentioned above. The simulated experiment environment was built using a Blockchain-powered Verifiable PPC (Privacy-Preserving Computation) network, namely Delta Framework[1], which integrates blockchain and ZPK (Zero Knowledge Proof) to ensure it is verifiable that the computation is actually performed as designed on the required data in a privacy-preserving manner. Delta transforms the tasks into horizontal/vertical federated learning, or federated analytics task and executes it on the network.

For the experiment environment, we also implemented a dashboard interface on the Delta Framework for usability and use Jupyter Lab to construct these tasks, which are written in Python. In the experiment, we simulated three nodes on a computer with the Ubuntu OS. The specifications of the experiment computer are given in Table 1, and the versions of experiment platforms and tools are shown in Table 2.

In the experiment, we took three nodes as three users to simulate the integration and sharing of data in privacy-preserving computation. We put three datasets collected

[1] https://deltampc.com/en.

from three subjects into the three decentralized nodes to simulate the data integration and sharing, in which the data structure is the same. The result showed the feasibility of health data integration and sharing in the simulated experiment environment of decentralized federated learning with three nodes.

Table 1. Specifications of experiment computer

Item	Specification
OS	Ubuntu 22.04.1 LTS
CPU	Intel® Xeon (R) E5-2603 v4
RAM	32 GB
Disk capacity	2 TB

Table 2. Versions of experiment platforms and tools

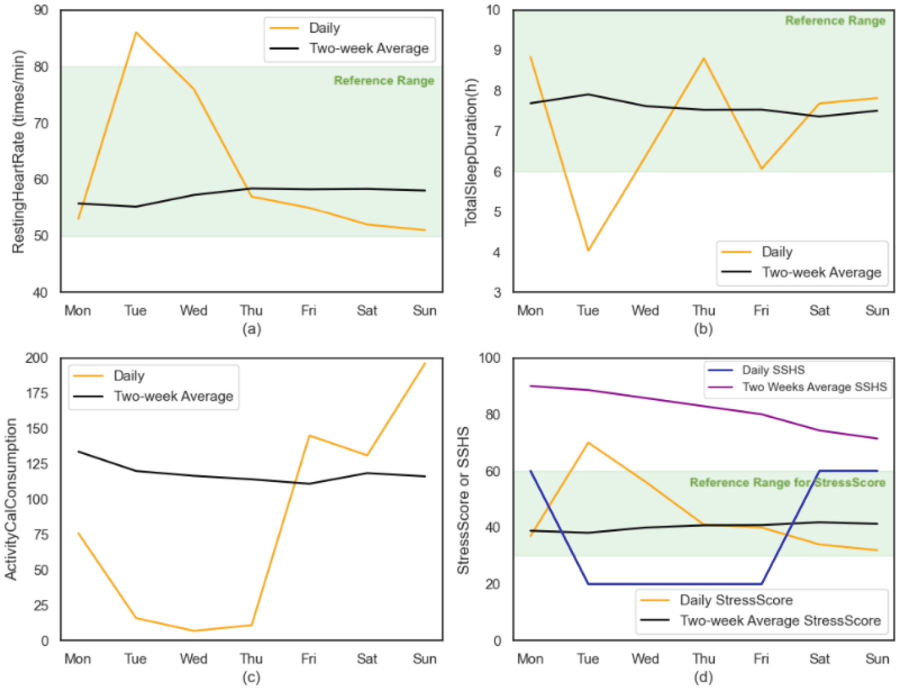
Platforms/Tools	Delta Framework	Node.js	Docker-container	Docker-compose	Python	Jupyter Lab
Version	*v 0.8.0*	*v 8.15.0*	*v 20.10.11*	*v 1.29.2*	*v 3.8.10*	*v 3.1.6*

4.2 Application Scenarios

In our previous work [21], we proposed two types of comparisons for visualized feedback of personal health data analysis results, namely temporal comparison, and horizontal comparison. Temporal comparison is designed to show the current health indicators versus the features obtained from the data of the past. On the other hand, the horizontal comparison provides a measure to let users know where they stand in relation with others, e.g., a peer of the same gender, or a group of the same age, which is also considered to be a good way for us to know ourselves. We describe two application scenarios by showing the visualization feedback which our proposed system aims to provide as follows.

Scenario for Temporal Comparison. A female student wants to know her health indicators for the past week from January 2 (Monday) to 8 (Sunday), 2023, comparing with the averages in the last two weeks up to the day before. She selected five health indicators, i.e., resting heart rate, total sleep duration, activity calorie consumption, stress score, and SSHS (the last two as one pair). The results of these temporal comparisons are shown in Fig. 2, in which the reference range is highlighted in light green color. The user can observe from the graphs how her health features changed, to what extent they are different from the averages, and whether the selected resting heart rate, total sleep duration, and stress score are within the reference range or not.

In Fig. 2, we can see that the daily indicators in the last one week have significant fluctuations compared to the averages in the last two weeks up to the day before. Moreover, the resting heart rate and the total sleep duration on Tuesday are greatly out of the reference range and the other one is less than the reference range, which resulting a higher stress score. And the SSHS is also very low, which is kept low for continuous four days. By continuously observing the bio indicators, such as the resting heart rate, or the sleep indicators, such as the total sleep duration, the user can understand their physical strength improvement, accumulated fatigue, or stress, and their relation to her health status.

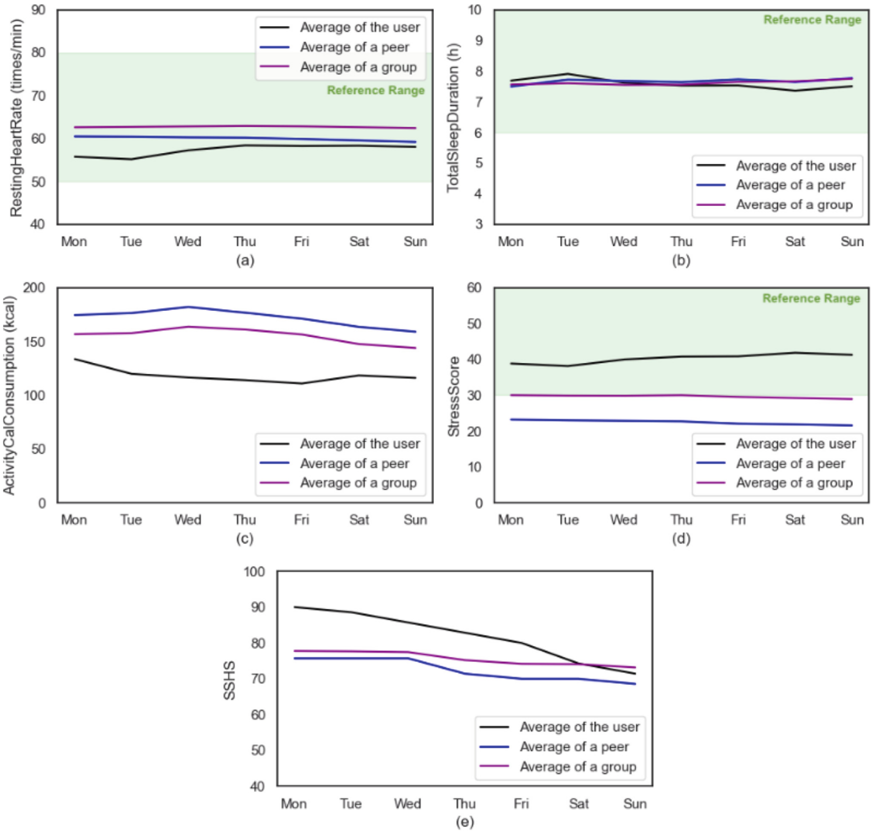


Period: from January 2 (Monday) to 8 (Sunday), 2023
Average based on the data in the last two weeks up to the day before.

Fig. 2. Temporal Comparison

Scenario for Horizontal Comparison. The female student wants to know her selected five health indicators during the period of January 2 (Monday) to 8 (Sunday), 2023, compared in terms of the averages in the last two weeks up to the day before with one of her good friends as well as a female group whose members all agree to share their data. The results of these horizontal comparisons of five selected health indicators are shown in Fig. 3, in which the reference range is highlighted in light green color. From the figures, she can observe the trends and the changes to find similarities or differences between herself and her friend or a group of chosen female users.

In Fig. 3, we can see that the trends of the average of the user, the average of a peer, and the average of a group are similar and stable. Among them, her resting heart rate and activity calorie consumption (represented in black lines) are lower than the peer and



Period: from January 2 (Monday) to 8 (Sunday), 2023
Average based on the data in the last two weeks up to the day before.

Fig. 3. Horizontal Comparison

the group, while their total sleep duration is almost the same, about 7.5 h. Less activity calorie consumption of the user may imply that she spent less time doing exercises, which may result in her higher stress scores than the peer and the group, although it is still within the reference range. From Fig. 3(d), we can also see that the average SSHS of the user has a declining trend from weekdays to weekends.

5 Conclusion

In this study, we proposed a trustworthy decentralized system with the blockchain and federated learning for privacy-preserving data integration and sharing. The system introduced in this paper can be expected to realize the processes of data life cycle management and utilization with trustworthiness.

In this paper, we described the functions and features of our proposed system in terms of health data integration, doubling data ownership, data analysis via decentralized federated learning, and incentive mechanism. Then, we explained the system architecture in IIAAC. We further described the simulated experiment and assumed two scenarios for sharing and utilization of personal health data. The individualized feedback was given to a user in a visualized way, in comparison with the averages calculated from the data of a peer or a group, which can also be used as an incentive for positive data sharing.

For our future work, we will implement the proposed trustworthy decentralized system. We will conduct the validation and performance evaluation experiment on the proposed system using decentralized federated learning models to analyze the tasks in [20]. We further plan to compare the proposed system with other related works for benchmark analysis.

Acknowledgement. The work was supported in part by 2022–2024 Masaru Ibuka Foundation Research Project on Oriental Medicine, 2020–2025 JSPS A3 Foresight Program (Grant No. JPJSA3F20200001), 2022 Waseda University Grants for Special Research Projects (Nos. 2022C-225 and 2022R-036), 2020–2021 Waseda University-NICT Matching Funds Program, 2022 JST SPRING (Grant No. JPMJSP2128), and 2022 Waseda University Advanced Research Center for Human Sciences Project (Grant No. BA080Z000300).

References

1. Wang, Z., et al.: From personalized medicine to population health: a survey of mHealth sensing techniques. IEEE Internet Things J. **9**(17), 15413–15434 (2022)
2. General Data Protection Regulation (GDPR). https://gdpr-info.eu/. Accessed 9 Feb 2023
3. California Consumer Privacy Act (CCPA). https://oag.ca.gov/privacy/ccpa. Accessed 9 Feb 2023
4. Personal Data Protection Commission Singapore (PDPC). https://www.pdpc.gov.sg/. Accessed 9 Feb 2023
5. Liang, W.: PDPChain: a consortium blockchain-based privacy protection scheme for personal data. IEEE Trans. Reliab. **72**(2), 586–598 (2022). https://doi.org/10.1109/TR.2022.3190932
6. Ali, M., Tariq, M., Naeem, F., Kaddoum, G.: Federated learning for privacy preservation in smart healthcare systems: a comprehensive survey. IEEE J. Biomed. Health Inform. **27**(2), 778–789 (2022)
7. Ito, K., Tago, K., Jin, Q.: i-Blockchain: a blockchain-empowered individual-centric framework for privacy-preserved use of personal health data. In: Proceedings of the 2018 9th International Conference on Information Technology in Medicine and Education (ITME), pp. 829–833 (2018)
8. Cong, R., Liu, Y., Tago, K., Li, R., Asaeda, H., Jin, Q.: Individual-initiated auditable access control for privacy-preserved IoT data sharing with blockchain. In: Proceedings of the 2021 IEEE International Conference on Communications Workshops (ICC Workshops), pp. 1–6 (2021)
9. Cong, R., et al.: Secure interoperation of blockchain and IPFS through client application enabled by CP-ABE. In: Moallem, A. (ed.) HCII 2022. LNCS, vol. 13333, pp. 30–41. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-05563-8_3
10. Elayan, H., Aloqaily, M., Guizani, M.: Sustainability of healthcare data analysis IoT-based systems using deep federated learning. IEEE Internet Things J. **9**(10), 7338–7346 (2021)
11. Nguyen, D.C., et al.: Federated Learning for Smart Healthcare: A Survey. ACM Comput. Surv. (CSUR) **55**, 1–37 (2021)

12. Sarosh, P., Parah, S., Bhat, G., Heidari, A.A., Muhammad, K.: Secret sharing-based personal health records management for the internet of health things. Sustain. Cities Soc. **74**, 103129 (2021)
13. Gartner Identifies Top Security and Risk Management Trends for 2022. https://www.gartner.com/en/newsroom/press-releases/2022-03-07-gartner-identifies-top-security-and-risk-management-trends-for-2022. Accessed 9 Feb 2023
14. Salim, S., Turnbull, B., Moustafa, N.: A blockchain-enabled explainable federated learning for securing internet-of-things-based social media 3.0 networks. IEEE Trans. Comput. Soc. Syst. (2021). https://doi.org/10.1109/TCSS.2021.3134463
15. Peng, Z., et al.: VFChain: enabling verifiable and auditable federated learning via blockchain systems. IEEE Trans. Netw. Sci. Eng. **9**(1), 173–186 (2022)
16. Heiss, J., Grünewald, E., Tai, S., Haimerl, N., Schulte, S.: Advancing blockchain-based federated learning through verifiable off-chain computations. In: Proceedings of 2022 IEEE International Conference on Blockchain (Blockchain), pp. 194–201 (2022)
17. Tago, K., Wang, H. Jin, Q.: Classification of TCM pulse diagnoses based on pulse and periodic features from personal health data. In: Proceedings of 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, pp. 1–6 (2019)
18. Tago, K., Nishimura, S., Ogihara, A., Jin, Q.: Improving diagnosis estimation by considering the periodic span of the life cycle based on personal health data. Big Data Res. **23**, 100176 (2021)
19. Wang, Y., et al.: Multi-ledger coordinating mechanism by smart contract for individual-initiated trustworthy data sharing. In: Moallem, A. (ed.) HCII 2023, LNCS, vol. 14045, pp. xx–yy. Springer, Cham (2023)
20. Wu, J., et al.: Multidimensional data integration and analysis for youth health care during the Covid-19 pandemic. In: Moallem, A. (ed.) HCII 2023, LNCS, vol. 14045, pp. xx–yy. Springer, Cham (2023)
21. Li, Z., Jin, Q.: Visualization design based on personal health data and persona analysis. In: Proceedings of 2019 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), Fukuoka, Japan, pp. 201–206 (2019)