Privacy-preserving collaborative machine learning in biomedical applications

Wonsuk Kim
School of Electrical Engineering,
Korea University,
Seoul, South Korea
Email: won425@korea.ac.kr

Junhee Seok
School of Electrical Engineering,
Korea University,
Seoul, South Korea
Email: jseok14@korea.ac.kr

***Abstract*—Machine learning (ML) algorithms are now widely used to tackle computational problems in diverse domains. In biomedicine, the rapidly growing amounts of experimental data increasingly necessitate the use of ML to discern complex data patterns. However, biomedical data is often considered sensitive, and the privacy of individuals behind the data is increasingly put at risk as a result. Traditional methods such as anonymization and pseudonymization are not always applicable and have limited effectiveness with respect to risk mitigation. Privacy researchers are actively developing alternative approaches to privacy protection, including strategies based on cryptography, such as homomorphic encryption and secure multiparty computation. This paper discusses recent advances in biomedical applications of these privacy techniques. We first review the key privacy techniques, then provide an overview of their applications in biomedical machine learning. Finally, we highlight the remaining challenges of current approaches and suggest directions for future work.**

***Index Terms*—Privacy-Preserving Machine Learning, Collaborative Learning, Federated Learning, Secure Multi-party Computation**

I. Introduction

Machine learning algorithms have revolutionized the way we solve problems in many domains, including computer vision, natural language processing, physical simulations, stock and housing market predictions, and biomedicine [1], [2], [3], [4], [5]. Since machine learning is driven by data, the quantity and quality of the data determine the performance of these algorithms. However, especially in biomedicine, it is often difficult to gather large amounts of data due to privacy and intellectual property issues associated with data sharing.

A traditional approach to protecting the privacy of biomedical data is to apply de-identification techniques [6]. "Anonymization" and "pseudonymization" are the most widely used methods for de-identification of private data. The possibility of re-identification can be reduced by removing identifying data features (anonymization) or replacing them with a random identifier for each subject (pseudonymization). However, these techniques are limited because they are still vulnerable to re-identification attacks, e.g. when linked with additional datasets [7].

Recent advances in cryptographic frameworks to overcome traditional limitations enabled new approaches to protecting privacy. In this paper, we present a review of emerging techniques for enhancing privacy in machine learning workflows in biomedicine based on a survey of recent literature. We discuss the promises and challenges of these techniques, and conclude with an outlook on future developments.

II. Privacy-preserving collaborative machine learning

Privacy-preserving technologies are primarily used to allow multiple input parties to collaboratively train ML models without revealing sensitive data, and it mainly includes technologies such as differential privacy, homomorphic encryption, trusted execution environment, and secure multiparty computation.

A. Differential privacy

Differential privacy (DP) [8] is a theoretical framework for releasing statistics from a dataset while limiting the leakage of information associated with each individual by adding a controlled amount of noise to the released data. Owing to these advantages, DP is adopted by the US Census Bureau [9] and several major technology organizations, including Google [10], Apple [11] and Microsoft [12]. These companies adopted DP to get insight from user behavior without revealing individual users' browsing habits. DP is also used in research dealing with genetic data and clinical data [13], [14]. DP can be applied not only to the dataset, but also to the parameters of algorithms or algorithm updates during training [15], [16]. However, there are limitations in that it reduces the accuracy of the model and makes it ambiguous to define an appropriate noise level.

B. Homomorphic encryption

Homomorphic encryption (HE) [17] is a cyptographic technique that allows computations on encrypted data without first decrypting it. By ensuring that no one can read or modify the data, HE can keep data safe, even in untrusted environments such as public clouds or external parties. Owing to this advantage, HE has numerous applications in genomics and biomedicine, where data is mainly spread across multiple institutions. For example, secure logistic regression and secure statistical tests in genome-wide association studies (GWAS) have

been proposed using HE [18], [19]. However, since HE mainly supports only addition and multiplication operations [20], the major limitation is that it is difficult to develop complex AI models with non-linear operations such as deep neural networks (DNNs) using HE. CryptoNets [21], training neural networks using HE, approximated non-polynomial functions such as sigmoid and rectified linear units and adopted mean-pooling layers instead of max-pooling layers. Also, HE is computationally expensive because it operates with encrypted data [22].

C. Trusted execution environment

Trusted execution environment (TEE) [23], also known as secure enclaves, provides hardware-level isolation and memory encryption on every server or device, which keeps application code and the data hidden from end users, credentialed insiders and third parties. For example, Intel Software Guard Extensions (SGX), ARM TrustZone and AMD Secure Processor allows the execution of the applications or code inside the enclaves that claims to be secure. With hardware-level isolated execution using TEE, secure implementation of genetic analytics has been applied [24], [25], [26], [27]. The major challenges for TEE are limited scalability of enclave memory and the need for additional development using software development kits(SDKs) from vendors providing TEE. Side-channel attacks are also a threat, such as timing attack, which is based on measuring the time taken to perform various computations [28]. To overcome these challenges, recent studies proposed efficient scalable frameworks based on TEE [29] and software timer manipulation to prevent side-channel attacks [30].

D. Federated Learning

Federated learning (FL) [31] is a machine learning technique that trains an algorithm across multiple decentralized edge devices or servers holding local data samples without exchanging them. This approach stands in contrast to traditional centralized machine learning techniques where all the local datasets are uploaded to one server. In FL, a server coordinates a network of nodes as demonstrated in Figure 1, each of which has training data that it cannot or will not share directly. The nodes each train a model, and they share the model or the updates of the model with the server. By not transferring the data itself, federated learning helps to ensure privacy and minimizes communication costs of sharing data.

Since the final model in FL is aggregated using models trained with local data, one might think that the model is biased or the performance is poor, but it can be overcome by regularization and frequent synchronizations of the models. Therefore, FL is mainly used for applications using clinical data as summarized in Table I. Secure implementation for survival analysis [32] and secure prediction of in-hospital mortality [33] using electrical health records (EHR) was proposed. Privacy-preserving structure for epileptic seizure detection [34] and arrhythmia detection [33] using electroencephalography (EEG) and electrocardiography (ECG) were also studied.

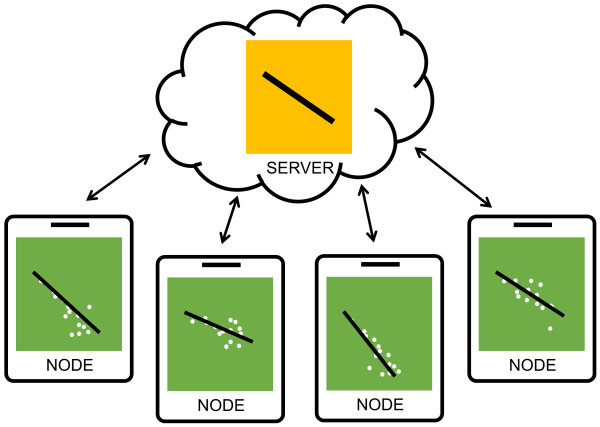

Fig. 1. The architecture of FL framework.

Secure diabetic retinal detection using retinal image [35], and prostate cancer diagnosis model using MRI data [36] were also proposed. The challenges in FL is that the data owners have to perform computations on the device or server that holds data and the performance is worse than centralized training. So, performance issues can occur if the data owners have limited computational capacity or small amount of local data. Since local data is not encrypted and the private data can be leaked only by the gradients of the model [37], [38] , privacy is not fully guaranteed.

E. Secure multiparty computation

Secure multiparty computation (SMPC) [39] is a cryptographic protocol that distributes computations across multiple parties, where individual parties cannot see the data of others. In other words, SMPC allows joint analysis of data without sharing the raw data. The architecture of SMPC is shown as Figure 2. SMPC protocol utilizes a well-established cryptographic concept called additive secret sharing. Each input party sends a different secret to each computation nodes. Each computation node computes the result on the secret shares from the input parties and shares the results with other computation nodes. Each computation node computes the final result by aggregating the results of all nodes. The SMPC-based privacy-preserving algorithm has been applied to medical diagnosing pneumonia and hepatitis [40], survival analysis [41], drug-target interaction using genetic data [42], [43] and quality control and population stratification for large-scale GWAS [44]. Although SMPC has a wider range of available operations than HE and is more efficient than HE in terms of computational cost, there are limitations in that it is a highly interactive computation and requires a large amount of communication between parties. The models implemented in SMPC are relatively simple compared to FL, as shown in Table I, because they have bottlenecks in performing non-linear operations such as ReLU, which are essential for implementing complex AI models. However, more privacy is

guaranteed because only unidentifiable secret shares remain on each node.

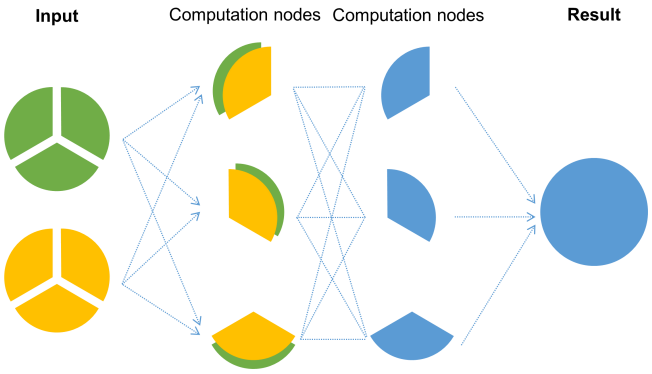

Fig. 2. The architecture of SMPC framework.

III. Outlook

In this paper, we investigated use cases applied to biomedical data, focusing on FL, which is flexible in terms of types of operations when training predictive models while keeping the data local and private, and SMPC, which encrypts both data and models as shown in Table I. Since SMPC still suffers from high overhead of cryptographic operations, it has been applied to EHR and genomic data, which are relatively low-dimensional among biomedical data, and implemented less complex models with fewer parameters compared to FL. To enhance the feasibility of SMPC, cryptography-side research has been conducted such as improving the encryption protocol [45] or using GPUs [46]. On the other hand, there have been ML-side approaches to reduce these bottlenecks. It is possible to speed up the runtime while maintaining the performance by reducing the number of ReLUs [47], [48] and using learnable scalars instead of the batch normalization layer [49].

Beyond the aforementioned issues and challenges, there are a number of practical concerns when applying FL and SMPC in production. Both technologies require a system design that considers the heterogeneity of the hardware level because computations are performed on multiple servers or devices. Since various environments such as hardware capacity, network connectivity, power, and physical location exist for each device, asynchronous communication [50] or fault-tolerant training methods [51] are necessary when applying FL, and not only semi-honest models but also malicious models [52] are required when applying SMPC.

Beyond the application of privacy-preserving techniques to supervised learning using biomedical data, there are many other problems that can be applied. For example, it can be applied to clustering for data analysis [53], reinforcement learning to infer treatment policies for patients [54], generative learning to impute missing genomic data [55] or to generate fake data to anonymize healthcare data [56]. Therefore, improving the runtime of these privacy-preserving techniques, solving the challenges that arise in production, and broadening the scope of application will be the future direction.

IV. Conclusion

With the advances in machine learning, various AI applications are emerging in biomedicine. However, the more AI models are trained on private biomedical data, the more the importance of the privacy and intellectual property issues increases. Thus, privacy-preserving techniques, such as DP, HE, TEE, FL and SMPC, are critical to making biomedical data more available and accessible. Among them, SMPC and FL were the most accurate methods, followed by HE and DP. A hybrid approach [57] (e.g., applying both FL and MPC) can be applied, and speeding up the operation in privacy technique to achieve a better-performing privacy-preserving model can be a future work.

Acknowledgment

This research was supported by the MOTIE (Ministry of Trade, Industry, and Energy) in Korea, under the Fostering Global Talents for Innovative Growth Program (P0008749) supervised by the Korea Institute for Advancement of Technology (KIAT) and National Research Foundation of Korea (NRF-2019R1A2C1084778).

References

[1] W. Kim and J. Seok, "Indoor semantic segmentation for robot navigating on mobile," in *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, 2018, pp. 22–25.
[2] E. Alpaydin, *Introduction to machine learning*. MIT press, 2020.
[3] W. Kim and J. Seok, "Simulation acceleration for transmittance of electromagnetic waves in 2D slit arrays using deep learning," *Scientific reports*, vol. 10, no. 1, pp. 1–8, 2020.
[4] H. Cho, B. Berger, and J. Peng, "Compact integration of multi-network topology for functional analysis of genes," *Cell systems*, vol. 3, no. 6, pp. 540–548, 2016.
[5] H. Cho, D. Ippolito, and Y. W. Yu, "Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs," *arXiv preprint arXiv:2003.11511*, 2020.
[6] B. Berger and H. Cho, "Emerging technologies towards enhancing privacy in genomic data sharing," 2019.
[7] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305–311, 2020.
[8] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy." *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3-4, pp. 211–407, 2014.
[9] J. M. Abowd, "The US Census Bureau adopts differential privacy," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 2867–2867.
[10] Úlfar Erlingsson, V. Pihur, and A. Korolova, "Rappor: Randomized aggregatable privacy-preserving ordinal response," in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, 2014, pp. 1054–1067.
[11] A. G. Thakurta, A. H. Vyrros, U. S. Vaishampayan, G. Kapoor, J. Freudiger, V. R. Sridhar, and D. Davidson, "Learning new words," 2017.
[12] B. Ding, J. Kulkarni, and S. Yekhanin, "Collecting telemetry data privately," *arXiv preprint arXiv:1712.01524*, 2017.
[13] H. Cho, S. Simmons, R. Kim, and B. Berger, "Privacy-preserving biomedical database queries with optimal privacy-utility trade-offs," *Cell systems*, vol. 10, no. 5, pp. 408–416, 2020.
[14] M. Winslett, Y. Yang, and Z. Zhang, "Demonstration of damson: Differential privacy for analysis of large data," in *2012 IEEE 18th International Conference on Parallel and Distributed Systems*, 2012, pp. 840–844.

TABLE I

SUMMARY OF RECENT LITERATURE ON FEDERATED LEARNING AND SECURE MULTIPARTY COMPUTATION APPROACHES IN GENOMICS AND BIOMEDICINE. (FL: FEDERATED LEARNING, SMPC: SECURE MULTIPARTY COMPUTATION, CNN: CONVOLUTIONAL NEURAL NETWORK, MLP: MULTI-LAYER PERCEPTRON, LSTM: LONG SHORT-TERM MEMORY)

Privacy Technique	Model	Authors	Year	Application
FL	MLP, CNN	Baghersalimi et al. [34]	2021	epileptic seizure detection
FL	CNN	Karthik et al. [36]	2021	prostate segmentation, MRI diagnosis of cancer
FL	LSTM	Lee et al. [33]	2020	in-hospital mortality prediction
FL	CNN	Lee et al. [33]	2020	arrythmia detection
FL	CNN	Balachandar et al. [35]	2020	diabetic retinopathy detection
FL	cox regression	Dai et al. [32]	2020	survival analysis
SMPC	logistic regression	Li et al. [40]	2021	medical diagnosis - pneumonia, hepatitis
SMPC	log-rank test, Kaplan-Meier estimator	von Maltitz et al. [41]	2021	survival analysis
SMPC	MLP	Ma et al. [42]	2020	drug-target interaction
SMPC	MLP	Hie et al. [43]	2018	drug-target interaction
SMPC	quality control, population stratification	Cho et al. [44]	2018	genetic associations

[15] J. Dong, A. Roth, and W. J. Su, "Gaussian differential privacy," *arXiv preprint arXiv:1905.02383*, 2019.

[16] N. Papernot, S. Song, I. Mironov, A. Raghunathan, K. Talwar, and Úlfar Erlingsson, "Scalable private learning with pate," *arXiv preprint arXiv:1802.08908*, 2018.

[17] C. Gentry, *A fully homomorphic encryption scheme*. Stanford university, 2009.

[18] M. Kim, Y. Song, S. Wang, Y. Xia, and X. Jiang, "Secure logistic regression based on homomorphic encryption: Design and evaluation," *JMIR medical informatics*, vol. 6, no. 2, p. e8805, 2018.

[19] T. Morshed, D. Alhadidi, and N. Mohammed, "Parallel linear regression on encrypted data," in *2018 16th Annual Conference on Privacy, Security and Trust (PST)*, 2018, pp. 1–5.

[20] L. Morris, "Analysis of partially and fully homomorphic encryption," *Rochester Institute of Technology*, pp. 1–5, 2013.

[21] R. Gilad-Bachrach, N. Dowlin, K. Laine, K. Lauter, M. Naehrig, and J. Wernsing, "Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy," in *International conference on machine learning*, 2016, pp. 201–210.

[22] C. Moore, M. O'Neill, E. O'Sullivan, Y. Doröz, and B. Sunar, "Practical homomorphic encryption: A survey," in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2014, pp. 2792–2795.

[23] M. Sabt, M. Achemlal, and A. Bouabdallah, "Trusted execution environment: what it is, and what it is not," in *2015 IEEE Trustcom/BigDataSE/ISPA*, vol. 1, 2015, pp. 57–64.

[24] N. Dokmai, C. Kockan, K. Zhu, X. Wang, S. C. Sahinalp, and H. Cho, "Privacy-preserving genotype imputation in a trusted execution environment," *Cell Systems*, 2021. [Online]. Available: https://doi.org/10.1016/j.cels.2021.08.001

[25] F. Chen, S. Wang, X. Jiang, S. Ding, Y. Lu, J. Kim, S. C. Sahinalp, C. Shimizu, J. C. Burns, and V. J. Wright, "Princess: Privacy-protecting rare disease international network collaboration via encryption through software guard extensions," *Bioinformatics*, vol. 33, no. 6, pp. 871–878, 2017.

[26] F. Chen, C. Wang, W. Dai, X. Jiang, N. Mohammed, M. M. A. Aziz, M. N. Sadat, C. Sahinalp, K. Lauter, and S. Wang, "Presage: privacy-preserving genetic testing via software guard extension," *BMC medical genomics*, vol. 10, no. 2, pp. 77–85, 2017.

[27] C. Kockan, K. Zhu, N. Dokmai, N. Karpov, M. O. Kulekci, D. P. Woodruff, and S. C. Sahinalp, "Sketching algorithms for genomic data analysis and querying in a secure enclave," *Nature methods*, vol. 17, no. 3, pp. 295–301, 2020.

[28] W. Wang, G. Chen, X. Pan, Y. Zhang, X. Wang, V. Bindschaedler, H. Tang, and C. A. Gunter, "Leaky cauldron on the dark land: Understanding memory side-channel hazards in SGX," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 2421–2434.

[29] J.-B. Truong, W. Gallagher, T. Guo, and R. J. Walls, "Memory-Efficient Deep Learning Inference in Trusted Execution Environments," *arXiv preprint arXiv:2104.15109*, 2021.

[30] W. Huang, S. Xu, Y. Cheng, and D. Lie, "Aion Attacks: Manipulating Software Timers in Trusted Execution Environment."

[31] J. Koneỳ, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," *arXiv preprint arXiv:1610.05492*, 2016.

[32] W. Dai, X. Jiang, L. Bonomi, Y. Li, H. Xiong, and L. Ohno-Machado, "VERTICOX: Vertically Distributed Cox Proportional Hazards Model Using the Alternating Direction Method of Multipliers," *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[33] G. H. Lee and S.-Y. Shin, "Federated learning on clinical benchmark data: Performance assessment," *Journal of medical Internet research*, vol. 22, no. 10, p. e20891, 2020.

[34] S. Baghersalimi, T. Teijeiro, D. Atienza, and A. Aminifar, "Personalized Real-Time Federated Learning for Epileptic Seizure Detection," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[35] N. Balachandar, K. Chang, J. Kalpathy-Cramer, and D. L. Rubin, "Accounting for data variability in multi-institutional distributed deep learning for medical imaging," *Journal of the American Medical Informatics Association*, vol. 27, no. 5, pp. 700–708, 2020.

[36] K. V. Sarma, S. Harmon, T. Sanford, H. R. Roth, Z. Xu, J. Tetreault, D. Xu, M. G. Flores, A. G. Raman, and R. Kulkarni, "Federated learning improves site performance in multicenter deep learning without data sharing," *Journal of the American Medical Informatics Association*, vol. 28, no. 6, pp. 1259–1264, 2021.

[37] A. Wainakh, F. Ventola, T. Mü, J. Keim, C. G. Cordero, E. Zimmer, T. Grube, K. Kersting, and M. Mühlhäuser, "User Label Leakage from Gradients in Federated Learning," *arXiv preprint arXiv:2105.09369*, 2021.

[38] Z. Wang, M. Song, Z. Zhang, Y. Song, Q. Wang, and H. Qi, "Beyond inferring class representatives: User-level privacy leakage from federated learning," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, 2019, pp. 2512–2520.

[39] Y. Lindell, "Secure multiparty computation for privacy preserving data mining," *Encyclopedia of Data Warehousing and Mining*, pp. 1005–1009, 2005.

[40] D. Li, X. Liao, T. Xiang, J. Wu, and J. Le, "Privacy-preserving self-serviced medical diagnosis scheme based on secure multi-party computation," *Computers & Security*, vol. 90, p. 101701, 2020.

[41] M. von Maltitz, H. Ballhausen, D. Kaul, D. F. Fleischmann, M. Niyazi, C. Belka, and G. Carle, "A Privacy-Preserving Log-Rank Test for the Kaplan-Meier Estimator With Secure Multiparty Computation: Algorithm Development and Validation," *JMIR medical informatics*, vol. 9, no. 1, p. e22158, 2021.

[42] R. Ma, Y. Li, C. Li, F. Wan, H. Hu, W. Xu, and J. Zeng, "Secure multiparty computation for privacy-preserving drug discovery," *Bioinformatics*, vol. 36, no. 9, pp. 2872–2880, 2020.

[43] B. Hie, H. Cho, and B. Berger, "Realizing private and practical pharmacological collaboration," *Science*, vol. 362, no. 6412, pp. 347–350, 2018.

[44] H. Cho, D. J. Wu, and B. Berger, "Secure genome-wide association analysis using multiparty computation," *Nature biotechnology*, vol. 36, no. 6, pp. 547–551, 2018.

[45] D. Escudero, S. Ghosh, M. Keller, R. Rachuri, and P. Scholl, "Improved primitives for MPC over mixed arithmetic-binary circuits," in *Annual International Cryptology Conference*, 2020, pp. 823–852.

[46] S. Tan, B. Knott, Y. Tian, and D. J. Wu, "CRYPTGPU: Fast Privacy-Preserving Machine Learning on the GPU," *arXiv preprint arXiv:2104.10949*, 2021.
[47] N. K. Jha, Z. Ghodsi, S. Garg, and B. Reagen, "DeepReDuce: Relu reduction for fast private inference," *arXiv preprint arXiv:2103.01396*, 2021.
[48] I. Helbitz and S. Avidan, "Reducing ReLU Count for Privacy-Preserving CNN Speedup," *arXiv preprint arXiv:2101.11835*, 2021.
[49] S. De and S. Smith, "Batch Normalization Biases Residual Blocks Towards the Identity Function in Deep Networks," *Advances in Neural Information Processing Systems*, vol. 33, 2020.
[50] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.
[51] C. Xie, O. Koyejo, and I. Gupta, "Generalized byzantine-tolerant sgd," *arXiv preprint arXiv:1802.10116*, 2018.
[52] J. B. Nielsen, P. S. Nordholt, C. Orlandi, and S. S. Burra, "A new approach to practical active-secure two-party computation," in *Annual Cryptology Conference*, 2012, pp. 681–700.
[53] A. Hegde, H. Möllering, T. Schneider, and H. Yalame, "SoK: Efficient Privacy-preserving Clustering." PETS, 2021.
[54] A. Raghu, M. Komorowski, I. Ahmed, L. Celi, P. Szolovits, and M. Ghassemi, "Deep reinforcement learning for sepsis treatment," *arXiv preprint arXiv:1711.09602*, 2017.
[55] R. Viñas, T. Azevedo, E. R. Gamazon, and P. Liò, "Gene expression imputation with generative adversarial imputation nets," *bioRxiv*, 2020.
[56] E. Piacentino and C. Angulo, "Generating fake data using gans for anonymizing healthcare data," in *International Work-Conference on Bioinformatics and Biomedical Engineering*, 2020, pp. 406–417.
[57] D. Froelicher, J. R. Troncoso-Pastoriza, J. L. Raisaro, M. Cuendet, J. S. Sousa, J. Fellay, and J.-P. Hubaux, "Truly Privacy-Preserving Federated Analytics for Precision Medicine with Multiparty Homomorphic Encryption," *bioRxiv*, 2021.