Federated Learning Approaches for Privacy Preserving Data Science in Healthcare

Abstract

Federated Learning (FL) is a technique that trains learning algorithms on decentralized data, particularly in situations where sharing raw data is challenging due to privacy concerns. An example of this data is Electronic Health Records (EHRs), which encompass sensitive patient information. In FL, local models are trained, and the model parameters are then combined on a central server instead of sharing sensitive data. However, this approach has privacy issues, so before disclosing the model parameters, privacy protection measures such as data anonymization must be put in place. A key component of FL research is striking a balance between privacy and utility because incorporating privacy algorithms may affect the utility. This research work will propose an efficient privacy-preserved federated learning model for heart disease prediction. During the first stage, data from various medical devices or facilities will be combined and locally preprocessed to ensure privacy and legal compliance. Data quality will be ensured by methods like feature encoding, normalization, and handling missing values. These datasets will be used to train local deep learning models; only the encrypted model weights or gradients will be sent to a central server. Federated averaging and differential privacy techniques will be used by the central server to safely aggregate updates, ensuring statistical indistinguishability and privacy. The accuracy, precision, recall, F1-score, and AUC will be used to iteratively refine and validate a global model.

Keywords:*Data Analytics, Data Mining, Healthcare Sector, Deep Learning, Federated Learning, Data Security. Encryption*

1. Introduction

Big Data is a category of data that is challenging for conventional data processing software to handle. Big data was originally associated with the fundamental concepts of volume, variety, and velocity; it encompasses data dimensions that surpass the processing capabilities of software within an acceptable timeframe. Big data analysis offers a number of benefits and ensures that healthcare has enormous potential for transformation [1]. The preservation of huge data is undoubtedly being improved daily. Healthcare associations are also developing a focused strategy to identify confidentiality requirements for association protection. The methods of associations to manage, analyze, and regulate data have been altered by big data. Big data is crucial for improving patient outcomes, predicting trends, facilitating early interventions, preventing diseases, reducing healthcare delivery costs, and enhancing quality of life.

Data mining is the process of collecting and identifying information from large amounts of data, using techniques that allow for the maintenance of data from simple problems [2]. Additionally, a wide range of fields, including marketing, medical diagnosis, weather forecasting, and state security, were aided by this extracted data. Data mining that aims to protect private information from uninvited disclosure is referred to as privacy-preserving. Data mining techniques were used to arithmetically aggregate and evaluate a model set of data since privacy preservation involves guarding against entity data records being revealed.

Privacy preservation is crucial for effectively transmitting data to consumers through data extraction methodologies such as clustering and classification [3]. Data privacy is a critical issue when sharing information. When gathering and keeping personal information, individual privacy must be respected. Financial, business, and medical information are examples of personal data. Personal information privacy upholds individuals' right to decide what personal information they can share. Personal data can be protected using methods such as authentication, authorization, and encryption.Numerous new data mining methods have been developed as computer storage capabilities have advanced. All social organizations are capable of acquiring an increasing amount of data. Traditional privacy protection techniques are unable to meet the pressing demand for privacy protection in data mining because they block access to the knowledge contained in the data when they protect sensitive information. Privacy protection in data mining primarily addresses two facets. Sensitive original information data is either modified or removed from the original database. The objective of this action is to protect individual privacy from harmful data acquisition. The next step involves utilizing the data in a more advantageous manner. The sensitive rule should be removed since the service data excavation algorithm may ruin data privacy due to the sensitive information it uncovers from the database.

The topic of privacy protection research is determined by the practical implementation of various privacy protection requirements. General privacy preservation techniques focus on data protection at a diminished privacy level, achieving privacy preservation through the implementation of statistical and probabilistic models. The primary goal of privacy preservation in data mining is to ensure the protection of various data attributes in high-level data [4]. Data release-based privacy protection serves as a standard privacy protection technique across numerous applications, making the built privacy algorithm adaptable.

Federated learning is a decentralized machine learning method that constructs a model across multiple remote locations before consolidating it on a central server. This process is known as collaborative learning, wherein a local model is developed at each site prior to the aggregation and training of all models at the central server [5]. Each local model update is accessible to the aggregating server. However, in terms of raw data transmission, this model assumes that there is no communication between the local models. The major aim of the proposed research work is to design and develop a novel federated learning approach for privacy preserving in healthcare data.Federated Learning (FL) is an innovative method in privacy-preserving data science, especially significant in the healthcare sector. Federated learning facilitates collaborative model training among decentralized healthcare organizations without the need to transfer sensitive patient data to a central repository. This decentralized platform facilitates strong AI development while guaranteeing data confidentiality and privacy and adhering to strict healthcare standards. This platform is significant as it empowers medical professionals to develop diagnostic tools and predictive models using a diverse range of large-scale, heterogeneous datasets, thereby enhancing accuracy and generalization. FL addresses significant privacy issues, promotes inter-institutional collaboration, and speeds up progress in customized medicine, disease prediction, and treatment improvement.

2. Literature Review

Mohammed Abaoud *et al.* [6] proposed an innovative methodology for privacy-preserving federated learning models specifically designed for healthcare applications. The proposed system allowed healthcare organizations to collaboratively train machine learning models on decentralized data while also safeguarding the confidentiality of individual patient information. In the model aggregation phase, the suggested system secured sensitive data by using advanced privacy-preserving techniques, such as secure multi-party computation and differential privacy. The simulation results demonstrated that the proposed methodology outperformed existing approaches in offering more utility and assuring strong privacy protections. The proposed method demonstrated the viability of secure and privacy-preserving collaboration on healthcare data, acting as a strong testament to its practicality and efficacy.

Shynu Padinjappurathu Gopalan *et al.* [7] proposed an effective privacy-preserving (PP) strategy for patient healthcare data obtained from IoT devices, intended for disease prediction in contemporary healthcare systems. The suggested method employed Log of Round value-based Elliptic Curve Cryptography (LR-ECC) to augment security during data transmission following the initial authentication phase. Authorized healthcare personnel can safely download patient data on the hospital's premises. The Herding Genetic Algorithm-based Deep Learning Neural Network (EHGA-DLNN) can analyze this data using the trained system to predict diseases. The experimental results indicate that the suggested strategy enhanced prediction accuracy, privacy, and security relative to existing techniques.

Nirmala Devi Kathamuthu *et al.* [8] developed a deep Q-learning-based neural network with a privacy preservation approach (DQ-NNPP) to protect data transmission from external threats while minimizing encryption and decryption time. This technique was employed to manage patient data, hence diminishing network traffic. This procedure also diminishes the expenses and inaccuracies associated with communication. The simulation results demonstrated that the proposed technique attained an accuracy of 93.74%, sensitivity of 92%, specificity of 92.1%, communication overhead of 67.08%, encryption time of 58.72 ms, and decryption time of 62.72 ms.

Ke Wang *et al.* [9] developed a trapdoor permutation technique and introduced a verified forward searchable encryption framework. The original system replaced the incrementing technique with a private key and the trapdoor replacement function, which transformed the relevant state counter after each insertion update. Additionally, a multi-keyword search verification mechanism utilizing a pseudo-random function was proposed. The experimental findings demonstrated that the proposed scheme is appropriate for IoT-enabled healthcare systems.

Yi Sun *et al.* [10] introduced a privacy-preserving medical record searching technique, PMRSS, enabling secure retrieval of diagnosis reports through merely two rounds of interaction, ensuring no additional information is disclosed between the parties involved. In the proposed approach, the patient can safely make self-helped medical diagnoses by accessing past case databases and safely comparing the blinded abstracts of current data and prior records, as well as by blinding the patient's healthy data and the intelligent doctor's database, respectively. The simulation results demonstrated that the proposed approach attained bilateral security.

Guangjun Wu *et al.* [11] suggested a blockchain-based smart healthcare system that offered fine-grained privacy protection for secure data exchange and sharing among various users. Multi-level smart contracts are created on a blockchain platform to facilitate dynamic access control between publishers and requesters, hence fulfilling the needs of an anonymous transaction process. The EMR properties were categorized into various privacy levels, and corresponding privacy budgets of Local Differential Privacy (LDP) were configured to attain the objective of attribute-based differential privacy protection. The prototype system successfully established patient-specific privacy settings at the patient's location while simultaneously providing error-free statistics at the requester's location.

Sascha Welten *et al.* [12] used a paradigm known as the Personal Health Train (PHT), which makes use of distributed analytics (DA) techniques, to enable studies on sensitive patient data while also adhering to local data protection laws. This work introduced the PHT paradigm, which functioned with a restricted number of communication channels while maintaining the independence and autonomy of the data producers. The simulation results demonstrated that the proposed approach facilitated the training of data models utilizing distributed data sources.

Hongliang Bi *et al.* [13] developed a data analytics system for privacy preservation and data analysis in IoT-enabled healthcare with deep learning, capable of isolating privacy information included in raw data and analyzing health-related data. The proposed system functions by separating privacy-sensitive content and extracting and identifying non-privacy data. At the cloud endpoint, health-related data was evaluated without any of users' private information, and an advanced security module was developed utilizing convolutional neural networks. The effectiveness and resilience of the suggested system were confirmed by evaluating its performance in various scenarios.

Jiachun Li *et al.* [14] introduced ADDETECTOR, a privacy-preserving smart healthcare system for the cost-effective detection of Alzheimer's Disease (AD). ADDETECTOR utilized audio from smart devices as input and a DP-based mechanism alongside an FL-based framework to safeguard against the leaking of raw data and model specifics during data transmission. Furthermore, the federated learning framework employed an asynchronous privacy-preserving aggregation module to secure the model updates. Experimental results indicated that ADDETECTOR attained outstanding accuracy while maintaining a robust level of security protection.

J. Andrew Onesimu *et al.* [15] suggested a privacy-preserving data-collecting framework for IoT-based healthcare service systems. A clustering-based anonymity model was employed to provide an effective privacy-preserving strategy that addressed privacy standards and safeguarded healthcare IoT against various privacy threats. The threat model was established as client-server-to-user to ensure anonymity at both ends. The client side employed a modified clustering-based k-anonymity model with dissociation to anonymize data generated from the IoT nodes. A bottom-up clustering mechanism secured base-level privacy by generating record clusters based on privacy criteria. The server-side employed the cluster-combination method UPGMA to save communication expenses and enhance privacy levels. The simulation findings demonstrated that the proposed system effectively achieved a healthy balance between privacy and data utility.

Harsh Kasyap and Somanath Tripathy [16] introduced a privacy-preserving blockchain architecture within a collaborative learning framework. The suggested architecture established decentralized trust by authenticating participants' identities through hospitals serving as verified issuers. The architecture functioned without aggregators, computing the global model at the gateways by averaging all updates recorded as transactions in the blockchain. A privacy-leaking dense network had an accuracy of 98.20%, and a privacy-preserving, dense, centralized network had a 99.60% accuracy. The suggested learning model got it right 97.80% of the time.

Syed Atif Moqurrab *et al.* [17] developed a fog-enabled, privacy-preserving strategy that utilized deep learning to enhance the healthcare system. A Convolutional Neural Network with Bidirectional-LSTM and Medical Entity Recognition is the foundation of the suggested model. The experimental findings indicated that the suggested model surpassed the current models, with a recall of 91.14%, precision of 92.63%, and an F1 score of 92%. The proposed model demonstrated a 28.77% enhancement in utility preservation relative to the state of the art.

Afsoon Abbasi and Behnaz Mohammadi [18] proposed a methodology utilizing the K-means++ clustering technique to attain an optimal k-anonymity procedure. The normal distribution function was employed to exclude infrequent data, hence enhancing the quality of anonymized data. Comprehensive testing has shown that the suggested method reduces information loss by 1.5 times and execution time by 3.5 times in comparison to the AKA and GCCG algorithms.

Li Zhang *et al.* [19] offered a federated learning framework for privacy preservation in IoT-based healthcare applications. A weighted average technique predicated on data quality was introduced to replace the conventional weight calculation method reliant on data volume. A masking approach utilizing homomorphic encryption and secure multi-party computation was presented for federated learning. This study also examined categorization accuracy and privacy protection measures. The simulation results suggested that the proposed method was capable of detecting lesion cell types with an accuracy of 76.9%.

Xu Yu *et al.* [20] suggested a Privacy-Preserving Cross-Domain Healthcare Wearables Recommendation system utilizing domain-dependent and domain-independent feature fusion(PPCDHWRec).The algorithm comprehensively retrieved both domain-dependent and domain-independent variables using the latent factor model and integrated auxiliary-domain and target-domain information through the FM algorithm, thereby significantly enhancing the performance of healthcare wearables predictions. Moreover, the suggested approach can successfully provide cross-domain data privacy preservation by merely transferring users' latent characteristics. Experiments conducted on two groups of auxiliary domains, characterized by high and low correlations with the target domain, demonstrated the efficacy of PPCDHWRec.

3. Research Gap

The major limitations of the existing approaches include:

- Risk of a single point of failure and attack.
- Reliance on a third party for handling sensitive data.
- Control over the data and risk of a privacy breach.
- Performance is reduced due to delays andimposed storage space overhead in all the blockchain nodes (asthey all store a copy of the data).
- Data is still storedcentrally and challenges of central storage of data exist.
- Users still do not have control of their data.
- Compared to the centralized approaches thedelay in accessing the health records and storage overhead hasbeen increased.
- Privacy information and health-relatedinformation are mixed, it is difficult for people to distinguishthem directly. Even if the privacy information can be separated,the health-related data may be distorted and difficult to extract.
- Huge information loss andprone to skewness attack.

- Huge correlation loss and prone tosensitivity attack.
- More robustand lightweightencryption algorithmis needed.
- The choice of activation function is critical becauseencrypted data only supports addition and multiplication.
- There is a time complexity issue when using theextreme learning machine algorithm for privacy protection.
- The privacy-preserving control variable ‘s value influencesthe model-based scheme ‘s stability.
- Conventional privacy-preserving classifiers are inflexibleand inefficient.
- Scalability, decentralization (to avoid single points of failure),durability, high accessibility, control mechanisms for data ownership andsharing, suitable structures that increase data exchange costs.

4. Motivation of the Study

During the digital transformation period, healthcare has experienced a significant shift towards data-driven decision-making, utilizing advanced technologies like artificial intelligence (AI). These technologies possess the capacity to transform healthcare through accurate diagnosis, tailored treatment strategies, and optimal resource distribution. The sensitive nature of healthcare data, including patient demographics, clinical records, genetic information, and others, poses considerable privacy and security challenges. Data science methods that protect privacy, like federated learning, differential privacy, homomorphic encryption, and secure multi-party computation, have come up as important ways to deal with these issues. These methods seek to derive insights from data while safeguarding patient privacy and adhering to rigorous rules, including the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA). The need to strike a compromise between the ethical duty to protect patient anonymity and innovation in healthcare analytics represents what inspired this work.The need for safe data-sharing frameworks has increased as healthcare increasingly depends on cooperative efforts involving numerous stakeholders, including hospitals, research institutes, and pharmaceutical corporations. Privacy-preserving methodologies safeguard sensitive information while simultaneously reducing risks related to data breaches, identity theft, and information misuse. Moreover, these methodologies empower businesses to fully leverage remote and fragmented healthcare data while maintaining data integrity and ownership. The need to investigate innovative approaches that integrate privacy-preserving strategies into practical healthcare applications motivates this research. It seeks to tackle critical issues including scalability, computing burden, and the balance between data utility and privacy. The goal of this work is to empower data-driven innovation, build trust among stakeholders, and advance privacy-preserving data science in order to create safe and fair healthcare systems across the world.

5. Objectives of the Study

The major objectives of the proposed study include:

- To design and implement a federated learning model that ensures data privacy and regulatory compliance by keeping raw data local to individual clients while enabling secure and efficient aggregation of model updates.
- To build a deep learning-based model capable of accurately predicting heart disease risk using diverse and distributed datasets while addressing challenges such as data imbalance through techniques like oversampling and weighted loss functions.
- To integrate novel encryption methods and differential privacy approaches for secure aggregation of client updates, ensuring patient-level data protection while maintaining model performance.

6. Scope of the Study

The goal of this research is to enable privacy-preserving data science in the healthcare industry by investigating and putting into practice federated learning (FL) methodologies. Traditional machine learning models typically utilize centralized training data, posing significant privacy issues, particularly in the healthcare sector that utilizes confidential patient data. Federated learning inherently facilitates decentralized model training by retaining data on client devices or servers and transmitting only model changes. This study emphasizes the utilization of federated learning in diverse healthcare applications, including disease prediction, customized treatment recommendations, and patient risk evaluation. The use of federated learning (FL) techniques in the medical field represents a major advancement in tackling important issues related to the security, privacy, and scalability of data-driven solutions. Healthcare data is frequently dispersed among various institutions, complicating the aggregation of extensive datasets due to legal and ethical issues related to patient privacy. Federated learning offers a decentralized framework where multiple nodes or institutions cooperatively develop machine learning models without exchanging raw data. This ensures adherence to privacy standards while effectively utilizing the aggregated insights of varied datasets. FL reduces the danger of data breaches by keeping private patient data on local devices or servers. This is important because attacks on healthcare systems are becoming more frequent. The potential for federated learning to transform patient care delivery makes it significant in the healthcare industry in ways that go beyond its technical advantages. By utilizing federated learning, healthcare providers can formulate individualized treatment methods and enhance decision-making processes while maintaining patient confidentiality. The democratization of data science ensures that progress in artificial intelligence is accessible beyond wealthy institutions, spreading throughout the global healthcare ecosystem. As healthcare moves toward value-based care models, federated learning becomes an important part of making systems that are accurate, efficient, moral, and long-lasting. This makes sure that advances in AI and machine learning help people while protecting their privacy rights.

7. Proposed Methodology

Federated Learning (FL) allows several decentralized clients to work together to train a global model without exchanging raw data. A central server receives just the updated model parameters from each client, which trains the model locally on its own private dataset. To enhance the global model, the server combines these updates using algorithms such as Federated Averaging. This iterative process shares the modified global model with the clients for further improvement. Privacy-preserving strategies ensure the privacy of individual data at all times. This research work will propose an efficient privacy-preserved federated learning model for heart disease prediction. The detailed block diagram of proposed methodology is shown in Figure 1.

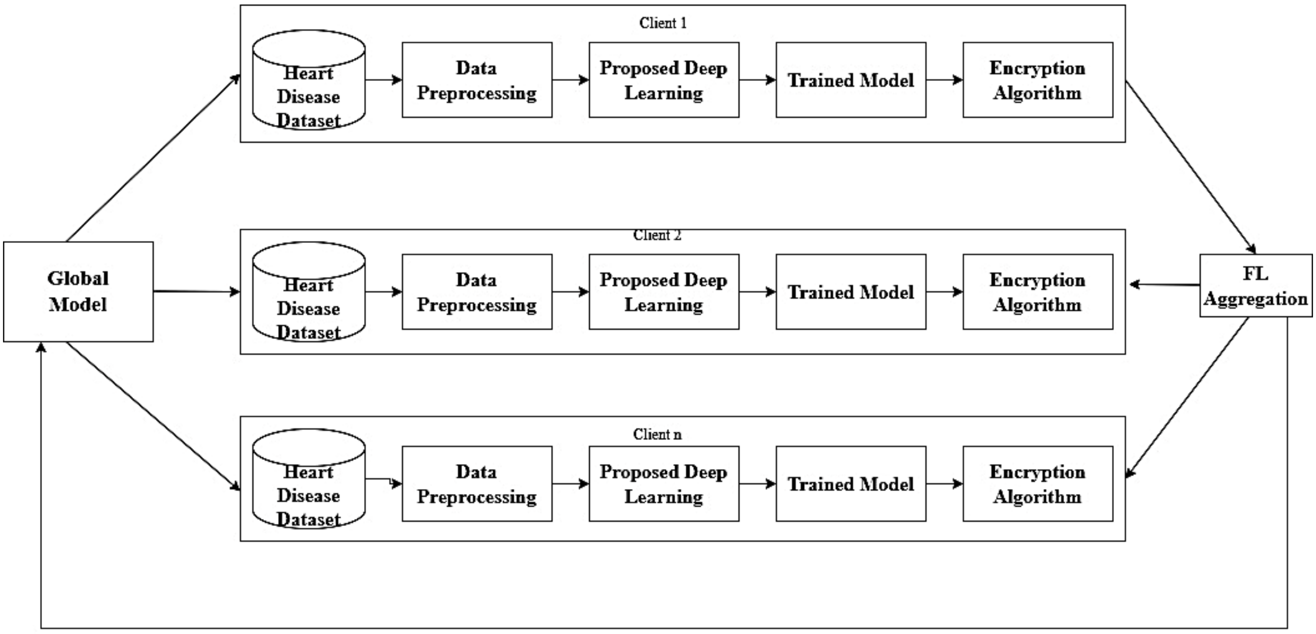

Figure 1: Block Diagram of Proposed Methodology

The initial phase of the suggested privacy-preserving federated learning model for heart disease prediction will entail the aggregation and preprocessing of data from various healthcare facilities or devices. Each institution or device (client) will maintain its own dataset, often encompassing patient characteristics such as age, gender, cholesterol levels, blood pressure, and other cardiovascular risk factors. These datasets will originate from publicly accessible sources or hospital databases. The raw data will, however, stay local to each client due to privacy concerns, ensuring regulatory compliance. Handling missing values, normalization, and categorical feature encoding are all part of data preprocessing. The preprocessed data will be fed as input to the proposed deep learning-based local models. After that, clients will be ready for training, in which local models will be loaded using shared architectures but trained separately using their own data.In this methodology, the safe aggregation and training of the central model will be the essential component of the federated learning process. After training, clients will only communicate the model weights (or gradients) to the central server, and they will update their models autonomously using local data. A novel encryption technique will be introduced to secure the privacy of these updates, enabling the server to aggregate the updates without disclosing individual model parameters. The Federated Averagingtechnique will be used to aggregate the model updates. This approach will generate a global model by combining the weights or gradients of the models from all participating clients. Differential privacy approaches will be used to further improve the aggregation process by adding noise to the updates. This will ensure that the model updates are statistically indistinguishable from one another, preserving patient-level data. The central server will then send the modified global model back to the clients for additional improvement, ensuring that the process maintains privacy at all times.The federated model will be assessed and validated in the last stage to ensure its effectiveness and resilience in predicting heart disease. The global model will be assessed using aggregated datasets from a holdout set or through cross-validation methods that do not necessitate access to individual patient data following multiple iterations of model updates. The model's performance in identifying heart disease risk will be evaluated using advanced evaluation metrics like accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). When dealing with extremely unbalanced datasets, methods like oversampling or weighted loss functions will be used to increase the robustness of the model. Lastly, the federated learning cycle will be used to deploy the model for real-time heart disease prediction, continuously improving as new data becomes available. This strategy will improve individualized healthcare outcomes without sacrificing data security by protecting privacy while utilizing a variety of datasets to develop a more accurate and generalized heart disease prediction model.

8. Implementation Feasibility

The proposed methodology will be designed and developed on Google Collaboratory platform with Python language.

References

1. Zakir, J., Seymour, T., & Berg, K. (2015). Big data analytics. Issues in Information Systems, 16(2).
2. Mining, W. I. D. (2006). Introduction to data mining (pp. 2-12). New Jersey: Pearson Education, Inc.
3. Xiao, X., & Tao, Y. (2006, June). Personalized privacy preservation. In Proceedings of the 2006 ACM SIGMOD international conference on Management of data (pp. 229-240).
4. Bertino, E., Fovino, I. N., & Provenza, L. P. (2005). A framework for evaluating privacy preserving data mining algorithms. Data Mining and Knowledge Discovery, 11, 121-154.
5. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2021). Advances and open problems in federated learning. Foundations and trends® in machine learning, 14(1–2), 1-210.
6. Abaoud, M., Almuqrin, M. A., & Khan, M. F. (2023). Advancing federated learning through novel mechanism for privacy preservation in healthcare applications. IEEE Access, 11, 83562-83579.
7. Padinjappurathu Gopalan, S., Chowdhary, C. L., Iwendi, C., Farid, M. A., & Ramasamy, L. K. (2022). An efficient and privacy-preserving scheme for disease prediction in modern healthcare systems. Sensors, 22(15), 5574.
8. Kathamuthu, N. D., Chinnamuthu, A., Iruthayanathan, N., Ramachandran, M., & Gandomi, A. H. (2022). Deep Q-learning-based neural network with privacy preservation method for secure data transmission in internet of things (IoT) healthcare application. Electronics, 11(1), 157.

9. Wang, K., Chen, C. M., Tie, Z., Shojafar, M., Kumar, S., & Kumari, S. (2021). Forward privacy preservation in IoT-enabled healthcare systems. IEEE transactions on industrial informatics, 18(3), 1991-1999.
10. Sun, Y., Liu, J., Yu, K., Alazab, M., & Lin, K. (2021). PMRSS: Privacy-preserving medical record searching scheme for intelligent diagnosis in IoT healthcare. IEEE Transactions on Industrial Informatics, 18(3), 1981-1990.
11. Wu, G., Wang, S., Ning, Z., & Zhu, B. (2021). Privacy-preserved electronic medical record exchanging and sharing: A blockchain-based smart healthcare system. IEEE journal of biomedical and health informatics, 26(5), 1917-1927.
12. Welten, S., Mou, Y., Neumann, L., Jaberansary, M., Ucer, Y. Y., Kirsten, T., ... & Beyan, O. (2022). A privacy-preserving distributed analytics platform for health care data. Methods of information in medicine, 61(S 01), e1-e11.
13. Bi, H., Liu, J., & Kato, N. (2021). Deep learning-based privacy preservation and data analytics for IoT enabled healthcare. IEEE Transactions on Industrial Informatics, 18(7), 4798-4807.
14. Li, J., Meng, Y., Ma, L., Du, S., Zhu, H., Pei, Q., & Shen, X. (2021). A federated learning-based privacy-preserving smart healthcare system. IEEE Transactions on Industrial Informatics, 18(3).
15. Onesimu, J. A., Karthikeyan, J., & Sei, Y. (2021). An efficient clustering-based anonymization scheme for privacy-preserving data collection in IoT based healthcare services. Peer-to-Peer Networking and Applications, 14(3), 1629-1649.
16. Kasyap, H., & Tripathy, S. (2021). Privacy-preserving decentralized learning framework for healthcare system. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 17(2s), 1-24.
17. Moqurrab, S. A., Tariq, N., Anjum, A., Asheralieva, A., Malik, S. U., Malik, H., ... & Gill, S. S. (2022). A deep learning-based privacy-preserving model for smart healthcare in Internet of medical things using fog computing. Wireless Personal Communications, 126(3), 2379-2401.
18. Abbasi, A., & Mohammadi, B. (2022). A clustering-based anonymization approach for privacy-preserving in the healthcare cloud. Concurrency and Computation: Practice and Experience, 34(1), e6487.
19. Zhang, L., Xu, J., Vijayakumar, P., Sharma, P. K., & Ghosh, U. (2022). Homomorphic encryption-based privacy-preserving federated learning in IoT-enabled healthcare system. IEEE Transactions on Network Science and Engineering, 10(5), 2864-2880.

20. Yu, X., Zhan, D., Liu, L., Lv, H., Xu, L., & Du, J. (2021). A privacy-preserving cross-domain healthcare wearables recommendation algorithm based on domain-dependent and domain-independent feature fusion. IEEE Journal of Biomedical and Health Informatics, 26(5), 1928-1936.