International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 2, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v2n1.07
Journal homepage: https://wepub.org/index.php/IJCSIT/index

# The Utilization of Homomorphic Encryption Technology Grounded on Artificial Intelligence for Privacy Preservation

Guangze Su[1,*], Jiufan Wang[2], Xiaonan Xu[3], Yufu Wang[4], Cankun Wang[5]

[1]Infomation Science, Trine University, Phoenix, AZ, USA;

[2]Independent Researcher, William & Mary, Williamsburge, VA, USA;

[3]Independent Researcher, Northern Arizona University, Flagstaff, USA;

[4]Computer Science & Engineering, Santa Clara University, Santa Clara, CA, USA;

[5]Biomedical Informatics The Ohio State University Columbus, USA.

[*]Corresponding Author: jamsu665@gmail.com

## ABSTRACT

The advent of Artificial Intelligence (AI) and Machine Learning (ML), particularly deep learning, has escalated the demand for computing resources. However, the high hardware requirements pose challenges for companies, compelling them to outsource ML tasks to the cloud. Nevertheless, concerns about cloud trustworthiness limit such applications. Encrypting data before uploading it to the cloud is a straightforward solution to ensure data security. However, traditional encryption schemes render ciphertext data unable to participate in operations within the ciphertext domain, posing challenges for data analysis. This paper delves into the pivotal role of homomorphic encryption in addressing the critical issue of privacy protection in machine learning.

## KEYWORDS

Machine learning; Security and privacy; Homomorphic encryption; Artificial intelligence

## 1. INTRODUCTION

The emergence of Machine Learning (ML), especially deep learning, demands more and more computing resources. Because of the high demands of hardware, companies that can't afford it will outsource machine learning tasks to the cloud. But distrust of the cloud will limit such applications. One simple idea is to encrypt the data and then upload it to the cloud, so that the data can be kept safe[1]. However, after the traditional encryption scheme encrypts data, the ciphertext data cannot participate in operations in the ciphertext domain. Therefore, for the important application significance of information mining of encrypted data, the key technical issues of machine learning for privacy protection based on homomorphic encryption scheme are deeply studied and analyzed in this paper.

1) Machine learning for privacy protection based on secure multi-party computing [3].

Secure multi-party computing assumes that multiple service users who do not trust each other exchange data information through secure multi-party protocols, so that each service user can realize distributed machine learning tasks under the premise that they only know their own data[2]. Especially in the current research of privacy protection machine learning, the security comparison scheme is a very difficult problem, as far as we know, the security comparison scheme is basically based on multi-party security computing to achieve. Secure multi-party computing requires the

interaction of each user involved in the calculation, and many interactions are often required to complete a calculation, which increases the communication cost and reduces the computing efficiency.

2) Outsourced machine learning based on order preserving and homomorphic encryption [4-5].

At present, many of the known outsourcing computing, many of the data is encrypted using homomorphic encryption technology, and then the encrypted data and computing protocols are outsourced to the third-party cloud[3]. However, the order relationship between encrypted data is obtained under ciphertext, that is, the comparison problem under ciphertext. So some scholars introduced the sequence-preserving encryption technology. Under certain conditions, order preserving encryption technology can not only ensure the order relationship of data size, but also encrypt data to achieve the purpose of privacy protection. However, this encryption scheme does not support homomorphic operations, and those that support homomorphic operations do not support sequence-preserving encryption. Some scholars use the two techniques in combination for privacy-protecting machine learning purposes, but using multiple encryption techniques means more complex calculations. Therefore, this method is generally very inefficient and far from the purpose of actual use.

Therefore, this paper combines homomorphic encryption and machine learning to achieve privacy protection, thus forming network security[4-5]. The advantage of homomorphic encryption technology combined with machine learning to achieve privacy protection is that it allows the calculation of encrypted data without decryption, so as to realize data analysis and model training while protecting data privacy. Through homomorphic encryption, users can perform machine learning algorithms in the encrypted state, protect sensitive information and ensure data security and privacy, making it possible to perform privacy-protected data processing in scenarios such as cloud computing, data sharing and collaborative analysis

# 2. RELATED WORK

## 2.1. Federated learning and homomorphic encryption

Federated learning is a distributed machine learning algorithm, but in the learning process, the participants do not share their own training data, each federated learning uses multiple compute nodes for joint training, aiming to improve performance, protect private information, and make it scalable to larger training data and larger models. Federated learning breaks the "data silos" and achieves the balance between data privacy protection and shared analytics, that is, the "data available invisible" data application model[6][7]. Some federal agencies have strict privacy requirements, or are bound by regulations, and may require systems to add additional protection mechanisms to prevent the inference of personal information. In these cases, transmitting model updates in plain text could allow potential adversaries to infer personal data. Full homomorphic encryption (FHE) can help us reduce risk by hiding the final model and disclosing only the final aggregate results to the parties.

Encrypting data is a crucial method for safeguarding data privacy, especially in scenarios where sensitive information needs to be transmitted or stored securely.[8]Homomorphic encryption is a specialized form of encryption that enables users to perform operations directly on the encrypted data, producing results that are equivalent to performing the same operations on the plaintext data. Figure 1 illustrates the fundamental process of homomorphic encryption. In this process, plaintext data is first encrypted using a homomorphic encryption scheme, resulting in ciphertext. Despite being encrypted, the ciphertext retains certain mathematical properties that allow specific operations to be performed on it without decrypting it. These operations can include addition, multiplication, or other mathematical functions.

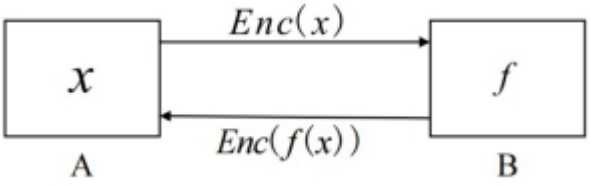


**Figure 1.** Homomorphic encryption implementation

After performing the desired operations on the ciphertext, the result is still encrypted and can be decrypted only by authorized parties possessing the corresponding decryption key. [9]This enables computations to be carried out on sensitive data while maintaining its confidentiality throughout the entire process.

Homomorphic encryption is particularly valuable in scenarios where privacy-preserving computations are necessary, such as in cloud computing, data sharing, or collaborative data analysis. By allowing computations to be performed directly on encrypted data, homomorphic encryption enhances data security and privacy, enabling organizations to leverage the benefits of data analysis and processing without compromising sensitive information.

## 2.2. Homomorphic encryption operation

Homomorphic encryption algorithms are divided into total homomorphic encryption algorithms and semi-homomorphic encryption algorithms.

- If a Homomorphic Encryption algorithm supports any form of ciphertext computation, it is called Fully Homomorphic Encryption (FHE).
- If it supports partial forms of ciphertext computation, such as only addition, only multiplication, or finite addition and multiplication, it is called semi-homomorphic encryption or partial homomorphic encryption.
- The English abbreviation is SWHE (Somewhat Homomorphic Encryption) or PHE (Partially Homomorphic Encryption).[10-11]

In general, since any computation can be constructed by addition and multiplication, a cryptographic algorithm can be said to satisfy total homomorphism if it satisfies both additive and multiplicative homomorphism.

The mathematical definition of homomorphic encryption is:

$$\mathrm{E}(m_1) \star \mathrm{E}(\mathrm{m}_2) = \mathrm{E}(\mathrm{m}_1 \star \mathrm{m}_2) \quad \forall\, m_1, m_2 \in M \tag{1}$$

Where E is the encryption algorithm and M is the set of all possible information. If the encryption algorithm E satisfies formula (1), then we say that E complies with homomorphic encryption in ★.

The current homomorphic encryption algorithm mainly supports two kinds of homomorphism on the operation: addition and multiplication.

It should be noted that the above formula (1) is only for us to understand the nature of homomorphic encryption more clearly, and the actual homomorphic encryption algorithm may be somewhat different. For example, the addition homomorphism of Paillier algorithm, then according to formula (1), the sum of the ciphertext should be equal to the sum ciphertext, but the actual situation is that the product of the ciphertext is equal to the sum ciphertext, so we generally only require that the ciphertext result obtained is the same as our expected calculation. However, there are no specific requirements on the ciphertext calculation (generally determined by the encryption algorithm).

### 2.3. Homomorphic encryption combined with machine learning

In the traditional recommendation system, users need to upload browsing history and evaluation information to achieve personalized recommendation, but these information are personal privacy data, direct upload will bring great security risks. In a federated recommendation system, each user keeps the data locally, uploading only specific model gradients[12]. Although this avoids direct disclosure of private data, it still reveals gradient information to the cloud server. In this case, homomorphic encryption must be used to protect the gradient uploaded by the user, that is, the user encrypts the gradient information using the addition homomorphic encryption algorithm before uploading the gradient, and then the cloud server aggregates (adds) the ciphertext gradient of all users, and then returns the updated model to each user for decryption and completion of training and updating.

In addition to federation learning, another important application area of homomorphic encryption is dense state computing.[13-15] Unlike federated learning, dense-state computing does not require multiple parties to participate, but requires more complex calculations (more operators and more computation) than federated learning. The homomorphic encryption algorithms used in of dense state computation are mostly LHE and FHE. In fact, the original intention of full-homomorphic encryption research is to achieve secure cloud computing, that is, users who have needs for cloud computing power can encrypt all local data, and then upload it to the cloud, and then the cloud server can complete the calculation according to the user's instructions, and the user's data will not be leaked to the cloud during the whole process, so as to complete "absolutely safe" cloud computing services.

However, due to the low efficiency of FHE, the use of full-homomorphic encryption for cloud computing is far from reaching the level of application[16-19]. Machine learning has a broad market in cloud computing, and machine learning has two needs: training and reasoning. The training process generally has more data and a large amount of computation, while the reasoning process has a relatively small amount of data and a small amount of computation, so the current research mainly focuses on the dense state of machine learning reasoning, and there are already fast schemes.

## 3. METHODOLOGY

### 3.1. Privacy-Preserving Federated Learning

In the scenario of federated learning, participants typically possess their own private data but are reluctant to directly share raw data due to privacy and security concerns. Homomorphic encryption technology provides robust privacy protection mechanisms for federated learning by encrypting data before uploading it to the central server. Participants can encrypt their data using homomorphic encryption algorithms on their local computing devices, eliminating the need to transmit raw data to the central server[20-23]. Consequently, sensitive information remains protected throughout the data transmission and storage processes. For example, in healthcare federated learning, hospitals can encrypt patient data before uploading it to a cloud server for model training, ensuring patient privacy without compromising data security. Therefore, homomorphic encryption effectively enhances data privacy and security in the context of federated learning.

During the model update and aggregation process in federated learning, homomorphic encryption plays a crucial role, particularly in protecting sensitive gradients. Participants encrypt model gradients using homomorphic encryption algorithms on their local computing devices and then upload the encrypted gradients to the central server for aggregation. Since homomorphic encryption allows computations to be performed on ciphertext, the central server can aggregate gradients without decrypting any participant data. This ensures data privacy and security during the model update process because the central server cannot access or leak any participant's raw gradient information. For instance, financial institutions can utilize homomorphic encryption to encrypt customer transaction data, upload encrypted gradients to the central server for model updates, and enhance

prediction models without disclosing customer transaction details. Therefore, homomorphic encryption provides crucial privacy protection and security assurance for the model update and aggregation process in federated learning.

### 3.2. Experimental design

The experiment aims to evaluate privacy-preserving techniques using publicly available datasets due to the sensitive nature of electronic medical record (EMR) data. Two datasets were used for experimentation:

1. The Pima dataset from the National Institute of Diabetes and Digestive and Kidney Diseases contains features such as pregnancies, glucose concentration, blood pressure, insulin level, body mass index (BMI), and age, which can be utilized to predict whether a patient has diabetes. This dataset comprises over 700 samples with nine features.

2. The Heart Disease UCI dataset, jointly released by the Hungarian Institute of Cardiology, the University Hospital Zurich, and the University Hospital Basel, consists of features like resting blood pressure, fasting blood sugar, and maximum heart rate achieved. It includes over 300 samples with 14 features, which can be used to predict the presence of heart disease.

Based on this content analysis, the experiment aims to implement privacy-preserving techniques on these datasets, such as homomorphic encryption or secure multi-party computation, to ensure patient privacy while performing predictive analytics tasks. The goal is to demonstrate the feasibility and effectiveness of these techniques in protecting sensitive medical data during analysis and model training processes.

### 3.3. Experimental data and methods

The two data sets used in this paper all conform to the scenario of horizontal federation learning training, that is, the feature space of the data set is the same and the sample space is different. The experiment in this paper is to test the accuracy of the trained model under two scenarios of encrypted federated learning and unencrypted federated learning. In order to ensure the uniqueness of experimental environment variables, the client models of the two schemes are consistent. In the experiment, Pima data set was divided into two parts, 538 data for training and 238 data for testing. Similarly, this paper divides the heart disease data set into 190 training sets and 80 test sets in the experiment shown in Figure 2.

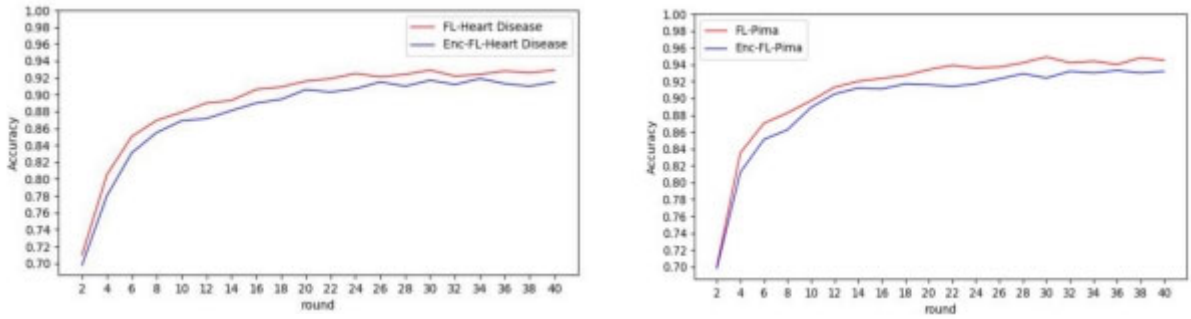


**Figure 2.** Data encryption accuracy curve of two groups

It can be seen from the above test experiments on the development data to realize encryption, the accuracy of the model trained by cryptographic federation learning is slightly lower than that of the model without cryptographic training, but this is within the acceptable range of the experiment. Because the client model parameter is a floating point number, and the encryption model parameter requires the integer operation of the model parameter. Although the calculation loss is reduced as much as possible in this paper, the accuracy of the trained model is still slightly decreased.

## 4. CONCLUSION

The integration of homomorphic encryption with machine learning holds significant promise for addressing privacy concerns in data analysis and model training processes. By encrypting data before uploading it to the cloud, organizations can ensure data security while harnessing the power of artificial intelligence and machine learning[24]. However, challenges persist in balancing privacy protection with computational efficiency, particularly in scenarios involving secure multi-party computing and outsourced machine learning. Despite these challenges, the combination of homomorphic encryption and machine learning offers a viable solution for achieving network security and protecting sensitive information in various applications such as cloud computing, data sharing, and collaborative analysis.

Moreover, have demonstrated experiments demonstrate the feasibility and effectiveness of privacy-preserving techniques, including homomorphic encryption and secure multi-party computation, in safeguarding patient privacy during predictive analytics tasks[25]. While there may be slight decreases in model accuracy due to encryption-related computational losses, these trade-offs are acceptable within the context of preserving data privacy. Moving forward, further research and advancements in homomorphic encryption algorithms and machine learning techniques are essential to enhance the efficiency and scalability of privacy-preserving methodologies. By leveraging these advancements, organizations can confidently adopt privacy-preserving technologies to safeguard sensitive data and uphold privacy standards in the era of AI and ML-driven analytics.

## ACKNOWLEDGEMENT

The completion of this paper cannot be achieved without the support and help from many aspects. I would like to express my special thanks to Mr. Yulu Gong for his valuable literature resources and core inspiration for this study[21]. Mr. Yulu Gong's article "Unveiling the Future: Navigating Next-Generation AI Frontiers and Innovations in Application "is published in the International Journal of Computer Science and Technology Information Technology, on December 1, 2023 (1), pages 147-156, https://doi.org/10.62051/ijcsit.v1n1.20, the research direction and depth of understanding of this article provided great help and inspiration. I would like to express my heartfelt thanks to Mr. Yulu Gong for his selfless sharing and profound insights. Your contribution is one of the important factors for the smooth progress and results of this research.

## REFERENCES

[1] Chen, Jianhang, et al. "One-stage object referring with gaze estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

[2] Zhang, Quan, et al. "Application of the AlphaFold2 Protein Prediction Algorithm Based on Artificial Intelligence." Journal of Theory and Practice of Engineering Science 4.02 (2024): 58-65.

[3] K. Tan and W. Li, "Imaging and Parameter Estimating for Fast Moving Targets in Airborne SAR," in IEEE Transactions on Computational Imaging, vol. 3, no. 1, pp. 126-140, March 2017, doi: 10.1109/TCI.2016.2634421.

[4] K. Tan and W. Li, "A novel moving parameter estimation approach offast moving targets based on phase extraction," 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 2015, pp. 2075-2079, doi: 10.1109/ICIP.2015.7351166.

[5] Tan, Kai, et al. “Integrating Advanced Computer Vision and AI Algorithms for Autonomous Driving Systems”. Journal of Theory and Practice of Engineering Science, vol. 4, no. 01, Jan. 2024, pp. 41-48, doi:10.53469/jtpes.2024.04(01).06.

[6] “Unveiling the Future Navigating Next-Generation AI Frontiers and Innovations in Application”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 147-56, https://doi.org/10.62051/ijcsit.v1n1.20.

[7] “The Application of Artificial Intelligence to The Bayesian Model Algorithm for Combining Genome Data”. Academic Journal of Science and Technology, vol. 8, no. 3, Dec. 2023, pp. 132-5, https://doi.org/10.54097/ykhccb53.

[8] Qian, Wenpin, et al. "NEXT-GENERATION ARTIFICIAL INTELLIGENCE INNOVATIVE APPLICATIONS OF LARGE LANGUAGE MODELS AND NEW METHODS." OLD AND NEW TECHNOLOGIES OF LEARNING DEVELOPMENT IN MODERN CONDITIONS (2024): 262.

[9] Zhang, Chenwei, et al. "SegNet Network Architecture for Deep Learning Image Segmentation and Its Integrated Applications and Prospects." Academic Journal of Science and Technology 9.2 (2024): 224-229.

[10] Pan, Yiming, et al. “Application of Three-Dimensional Coding Network in Screening and Diagnosis of Cervical Precancerous Lesions”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 61-64, https://doi.org/10.54097/mi3VM0yB.

[11] He, Yuhang, et al. “Intelligent Fault Analysis With AIOps Technology”. Journal of Theory and Practice of Engineering Science, vol. 4, no. 01, Feb. 2024, pp. 94-100, doi:10.53469/jtpes.2024.04(01).13.

[12] “Exploring New Frontiers of Deep Learning in Legal Practice: A Case Study of Large Language Models”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 131-8, https://doi.org/10.62051/ijcsit.v1n1.18.

[13] Wei, Kuo, et al. “Strategic Application of AI Intelligent Algorithm in Network Threat Detection and Defense”. Journal of Theory and Practice of Engineering Science, vol. 4, no. 01, Jan. 2024, pp. 49-57, doi:10.53469/jtpes.2024.04(01).07.

[14] Shen, Zepeng, et al. "EDUCATIONAL INNOVATION IN THE DIGITAL AGE: THE ROLE AND IMPACT OF NLP TECHNOLOGY." OLD AND NEW TECHNOLOGIES OF LEARNING DEVELOPMENT IN MODERN CONDITIONS (2024): 281.

[15] Su, Jing, et al. "Large Language Models for Forecasting and Anomaly Detection: A Systematic Literature Review." arXiv preprint arXiv:2402.10350 (2024).

[16] Wang, Yong, et al. "Construction and application of artificial intelligence crowdsourcing map based on multi-track GPS data." arXiv preprint arXiv:2402.15796 (2024).

[17] Duan, Shiheng, et al. “Prediction of Atmospheric Carbon Dioxide Radiative Transfer Model Based on Machine Learning”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 132-6, https://doi.org/10.54097/ObMPjw5n.

[18] Chen , Jianfeng, et al. “Implementation of an AI-Based MRD Evaluation and Prediction Model for Multiple Myeloma”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 127-31, https://doi.org/10.54097/zJ4MnbWW.

[19] “Machine Learning Model Training and Practice: A Study on Constructing a Novel Drug Detection System”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 139-46, https://doi.org/10.62051/ijcsit.v1n1.19.

[20] Duan, Shiheng, et al. "THE INNOVATIVE MODEL OF ARTIFICIAL INTELLIGENCE COMPUTER EDUCATION UNDER THE BACKGROUND OF EDUCATIONAL INNOVATION." The 2nd International scientific and practical conference “Innovations in education: prospects and challenges of today”(January 16-19, 2024) Sofia, Bulgaria. International Science Group. 2024. 389 p.. 2024.

[21] Gong, Yulu, et al. "RESEARCH ON A MULTILEVEL PRACTICAL TEACHING SYSTEM FOR THE COURSE'DIGITAL IMAGE PROCESSING." OLD AND NEW TECHNOLOGIES OF LEARNING DEVELOPMENT IN MODERN CONDITIONS (2024): 272.

[22] Liang, Penghao, et al. "Enhancing Security in DevOps by Integrating Artificial Intelligence and Machine Learning." Journal of Theory and Practice of Engineering Science 4.02 (2024): 31-37.

[23] W. Sun, W. Wan, L. Pan, J. Xu, and Q. Zeng, “The Integration of Large-Scale Language Models Into Intelligent Adjudication: Justification Rules and Implementation Pathways”, Journal of Industrial Engineering & Applied Science, vol. 2, no. 1, pp. 13–20, Feb. 2024.

[24] Zhou, Yanlin, et al. "Utilizing AI-Enhanced Multi-Omics Integration for Predictive Modeling of Disease Susceptibility in Functional Phenotypes." Journal of Theory and Practice of Engineering Science 4.02 (2024): 45-51.

[25] Wang, Yong, et al. "Autonomous Driving System Driven by Artificial Intelligence Perception Fusion." Academic Journal of Science and Technology 9.2 (2024): 193-198.