

Optimized Medical Data Storage and Query Retrieval Using Cloud Based Multi Indexing

Jayanthi S^a, Maalini D^{id}^a, Kavitha Anbalazhagan T³, Priyanka M⁴, Kosalairaman T⁵
and Karuppasamy L⁶

¹Department of CSE, Anna University, BIT Campus, Tiruchirappalli, India

^{2,5,6}Department of Information Technology, V.S.B. Engineering College, Karur, India

³Department of CSE, Vivekananda College of Engineering for Women, Namakkal, India

⁴Department of Computer Science and Technology, Vivekananda College of Engineering for Women, Namakkal, India


Keywords: Cloud Storage, Blowfish Encryption, Block Chain Technology, Keyword Based Retrieval, Confidentiality, Integrity and Cloud Based Data Management.

Abstract: Cloud storage with searchable encryption enables document retrieval from remote databases but requires sharing search keywords with database owners, raising privacy concerns. To address this challenge, this project integrates Blowfish encryption with Block chain technology to ensure secure, reliable, and efficient medical data storage and access. Blowfish, a robust symmetric-key block cipher, encrypts sensitive medical data, ensuring privacy even in case of unauthorized access. Blockchain provides a decentralized, immutable ledger to log transactions and access requests, enhancing data integrity. The proposed approach employs cryptographic methods and index structures to enable secure and efficient keyword searches on encrypted data. A secure index is generated during encryption, facilitating quick retrieval without exposing plaintext data. Access control mechanisms ensure that only authorized users, who possess the correct decryption keys and have their identities verified, can access the data. Additionally, the key verification process notifies data owners of unauthorized attempts. This solution achieves a balance between data privacy and search functionality by enabling keyword-based retrieval of encrypted cloud-stored data while ensuring the protection of sensitive information. The method demonstrates a practical and secure technique for cloud-based medical data management, maintaining confidentiality and integrity against potential threats.

1 INTRODUCTION

With the growing adoption of cloud computing, a significant volume of private data—including emails, official documents, and personal health records—is now being stored in the cloud (Huang, Song, et al. , 2019). By utilizing cloud storage, data owners can avoid the hassle of maintaining and storing data while benefiting from on-demand data storage services (Wang, Li, et al. , 2020). However, since the cloud server and data owners do not share the same trusted domain, the cloud server may not be entirely trustworthy, putting outsourced data at risk (Xu, Li, et al. , 2020). Therefore, to protect data privacy and prevent unauthorized access, sensitive data should be encrypted before outsourcing (Miao, Liu, et al. ,

2019). Data encryption, while essential for privacy, poses a significant challenge for the effective use of data, particularly when multiple data files are outsourced (Li and Yang, 2018). Keyword-based search is a popular method for retrieving files, as recovering all encrypted files at once is impractical in cloud computing environments (Qiu, Wang, et al. , 2017). However, encryption limits the ability to perform keyword searches, rendering traditional plaintext search techniques ineffective (Chaudhari and Das, 2019). Additionally, keyword privacy must be maintained during data encryption (Li, Guo, et al. , 2015). Cloud computing enables the provision of a vast reservoir of adaptable computing resources, including servers, networks, storage, applications,

^a <https://orcid.org/0009-0008-0065-9346>

and services, all readily accessible as needed (Miao, Ma, et al. , 2017).

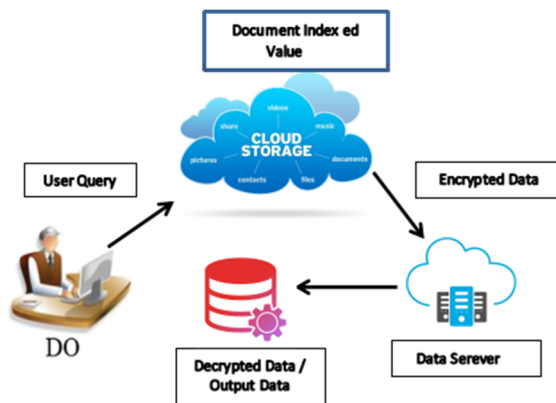


Figure 1: Cloud Keyword Search

In the Fig. 1. Cloud keyword search current digital economy, data security and privacy have grown to be key concerns for both individuals and businesses. Protecting sensitive data from breaches and unauthorized access is crucial as more and more information is stored on electronic devices. Encryption is one of the most dependable techniques for safeguarding data, and Blowfish encryption is a popular and trustworthy method used for this purpose (Maalini, and, Balraj, 2018), (Rayavel, Anbarasi, et al. , 2021). However, ensuring the safe storage and controlled distribution of encrypted data while retaining access control over it remains an ongoing challenge. The evolution of cloud computing has prompted data owners to transition from complex on-premises Document Management Systems (DMS) to commercial public cloud platforms to capitalize on enhanced cost-effectiveness and flexibility (Balraj, Maalini, et al. , 2018). As more people utilize the cloud to store documents and access data, the search requests must support multiple terms. The order in which the documents are returned determines their relevance to these terms. Comparable initiatives towards searchable encryption primarily focus on Blowfish-based, identity-based data sharing systems that combine data encryption with keyword storage and fully encrypted data (Maalini, and, Balraj, 2018), (Rayavel, Anbarasi, et al. , 2021).

A technique to effectively and precisely search for similarities in encrypted DNA data is proposed. The method begins by providing a general approach for rapidly estimating the edit distances between two sequences. Next, a novel Boolean search method is introduced that allows for the formulation of sophisticated logical queries (Xu, Li, et al. , 2020).

Additionally, the K-means clustering approach is utilized to enhance execution performance. To translate the edit distance computation problem into a symmetric set difference size approximation problem, a private approximation approach is presented (Wang, Li, et al. , 2020). By compressing each DNA sequence into a set of hash values, this method greatly reduces the number of elements in the cipher text that must match (Huang, Song, et al. , 2019). This approach is sufficient to execute the secure DNA similarity query function; thus, the data owner only needs to provide the search user with the key required to generate the encrypted index (which is different from the key used to encrypt the raw DNA sequences) (Miao, Liu, et al. , 2019). Therefore, under this paradigm, no unauthorized party will gain access to the raw DNA sequences belonging to the data owner. However, because encrypted indexes and trapdoors are one-way, the cloud server cannot retrieve the contents of other entities' indexes or trapdoors, even if it pretends to be a legitimate search user. Multiple users are permitted to view encrypted data as long as their attributes comply with the access control policy (Li and Yang, 2018). Only authorized users can access more complex queries, such as those requiring Boolean keyword expressions. The data owner maintains control over who has access to their encrypted data by outsourcing it to the cloud.

2 RELATED WORKS

Searchable Encryption (SE) aims to recover data where traditional encryption techniques fall short. Typically, SE methods work by creating an index with encryption. The data owner (DO) sends this index along with the encrypted data to the service provider. The service provider runs search algorithms and finds matches using the encrypted index and the search token that the data user (DU) supplies for a particular phrase. Earlier methods relied on linear scanning to return results, which reduced efficiency as the database size increased. However, many original SE schemes are rigid and cannot be easily modified. Updates for data stored on cloud servers are frequently required; hence dynamic SE technology has been developed to make SE schemes more accessible and adaptable (Maalini, Manivannan, et al. , 2024). Despite these advancements, dynamic SE introduces new security challenges. Most existing solutions assume an honest but curious cloud server, focusing less on security strategies against a malicious server (Manivannan, Gowda, et al. , 2024). When there is an internal setup issue or an external

attack, the cloud server can become malicious, leading to server modifications, encrypted data disclosure, or inaccurate query results (Li and Yang, 2018). In response to these challenges, this project aims to enhance SE technology, addressing both efficiency and security concerns, particularly against malicious servers. Attribute-Based Encryption (ABE) can be used for a variety of secure data distribution, search, and storage applications, two of which are keyword search and trapdoor-based storage (Rayavel, Anbarasi, et al. , 2021).

The method aims to provide a robust and effective framework for managing medical data while ensuring confidentiality and integrity by combining blockchain technology with the well-established Blowfish encryption algorithm. By encrypting sensitive medical data before it is stored or transmitted, Blowfish encryption offers strong protection, safeguarding the data even in the event of unauthorized access. The incorporation of blockchain technology, which provides a decentralized and immutable ledger system, complements this encryption technique. This ledger strengthens system security and accountability by acting as an open record of access requests and transactions related to the encrypted medical data (Qiu, Wang, et al. , 2017). Additionally, the system utilizes index structures and advanced cryptography to facilitate efficient keyword-based search operations on the encrypted data. This innovative approach ensures data privacy while allowing authorized users to access specific medical records. During the data outsourcing process, access rights are configured based on user identity. The system also employs strict key verification procedures to prevent unauthorized access attempts. Real-time notifications are triggered in the event of such attempts, alerting the data owner and enabling swift action to minimize potential security breaches. The proposed solution offers a comprehensive and carefully designed approach to the complex issue of securely transferring, storing, and using medical data in cloud environments. It strikes a balance between ensuring data privacy and enabling effective data retrieval, achieved by combining Blowfish's robust encryption with block chain technology's transparency and security features.

3 PROPOSED METHODOLOGY

The primary focus of this lecture is the infrastructure required to securely store. It comprises configuring the cloud storage environment, limiting access, and ensuring compliance with all relevant data protection

requirements. Disaster recovery and data redundancy solutions are also covered. The Data Encryption module uses the Blowfish encryption method to encrypt sensitive medical data before it is stored or sent to the cloud. This module controls the encryption process, generates encryption keys, and implements encryption methods to safeguard the privacy of the data. It also features key management and rotation methods to enhance security. The primary goal of the Index Creation module is to create safe indexes that provide effective keyword-based searches on the encrypted material. This module entails building cryptographic hashes or other indexing structures that preserve data privacy while enabling authorized users to retrieve encrypted information associated with particular keywords. Fig. 2. Architecture of proposed frameworks, it guarantees the secrecy of the stored data while maintaining search functionality.

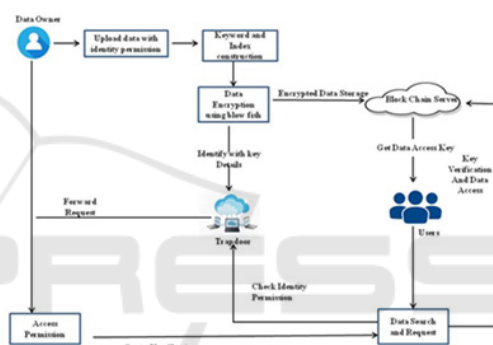


Fig.2.Architecture of Proposed Frameworks

3.1.1 Block chain Creation:

Using blockchain technology, the Blockchain Creation module creates an immutable and decentralized ledger system. In this module, a blockchain network that is customized to meet the needs of the medical data storage system is deployed and configured. By keeping track of all access requests, data exchanges, and significant events pertaining to the encrypted medical data on the blockchain, the module guarantees security and transparency. It strengthens the system's data integrity, auditability, and accountability while adding more resilience and confidence to the architecture as a whole.

3.1.2 Data Access Request

Authorized users requesting access to retrieve encrypted medical data from cloud storage are handled by the Data Access Request module. It has

features for requesting access, authenticating user credentials, and confirming the legitimacy of access authorizations. To improve security and privacy.

3.1.3 Identity Verification

Prior to allowing access to encrypted medical data, the Identity Verification module verifies users' identities. It securely verifies user identities by using multi-factor authentication methods including passwords, biometrics, or token-based authentication. This module is essential to maintaining compliance with data protection laws and avoiding unwanted access.

3.1.4 Secure Data Access

Authorized users can retrieve encrypted medical data securely with the help of the Secure Data Access module. It has features for granting access to the plaintext data, verifying access permissions, and decrypting the encrypted data using the proper decryption keys. This module guarantees the confidentiality and integrity of sensitive medical data during the data retrieval procedure.

3.1.5 Blowfish Encryption

It is suitable for application like file encryptors that operate automatically and communicate lines where the key is not changed often.

Quick: On large 32-bit processors, it encrypts data at a pace of 26 clock cycles per bytes.

Compact: It requires less than 5K of RAM to operate.

Easy to understand: 32-bit lookup tables, XOR, addition.

3.1.6 Secure

The default key length is 128-bits; however it can vary between 32 and 448-bits. It works well for applications like automatic file encryptors and communication links when the key is not changed frequently.

Blowfish Algorithm Steps:

Encryption:

Step 1: Dividing a message from 64 bits into 32-bits.

Step 2: The "left" 32-bits of the message are XORed with the first element of a P-array to create a value I'll call P'. After that, this value is put through the F transformation function and XORed

with the message's "right" 32 bits to produce a new value that I'll call F'.

Step 3: Next, the "left"-half of the message is replaced with F', and the "right"-half is replaced with P, loop it 15 times.

Step 4: Ultimately, the final two entries (17 and 18) in the P-array are XORed with the resultant P' and F' to produce the 64-bit cipher text.

Decryption:

Step 1: The S-array is indexed using the four bytes derived from a 32-bit input by the function.

Step 2: The output is generated by ORing and XORing results.

Since Blowfish uses a symmetric method, the encryption and decryption processes follow the same procedures. The only distinction is that ciphertext is used in decryption, whereas plaintext is used in encryption.

Based on the user's key, Blowfish precomputes the P-array and S-array values. After the key is successfully converted into the S-array and P-array, the original key can be thrown away. The S-array and P-array do not need to be recomputed as long as the key is the same; nonetheless, they need to be kept private.

3.1.7 Block chain Algorithm

Blockchain is an open, trusted, shared ledger of transactions that is not controlled by any one person but is accessible to all. It is a distributed database that keeps an ever-expanding list of transaction data records safe from alteration and tampering via cryptography. There exist three distinct varieties of blockchain technology: consortium, private, and public. Public blockchains like Bitcoin and Ethereum, allow anybody, anywhere, to join and receive relief whenever they choose. This is demonstrated by the intricate mathematical operations. The company's internal public ledger is called the private blockchain, and access to it is only authorized by the blockchain's owner. Block creation and mining speed are much faster on the private blockchain than on the public one since there are less nodes there. In contrast, a corporation or group of companies uses the consortium blockchain, and membership standards are employed to govern blockchain transactions more efficiently than a consensus.

3.1.8 Hashing

The process of hashing converts an arbitrary, variable-sized input into an output with a fixed size. Various functions are available for doing hashing at

different levels. The MD5 algorithm yields a hash value that is 128 bit, or 32 symbols long, and is commonly used for hashing. The series' most recent algorithm is MD5, however earlier iterations included Md2, Md3, and Md4. Although the technique was intended to be used as a cryptographic hashing algorithm, it has certain vulnerabilities because of issues that limit the number of unique hashes that it can produce. Secure Hashing technique (SHA), another cryptographic hash technique, generates a 160-bit hash output with 40 hexadecimal characters. Since the algorithm could not withstand collusion attacks, its application has declined. During this period, two new algorithms have been proposed: SHA 3 and SHA 256. The SHA 2 series of algorithms was developed by the US National Security Agency. The recently developed hash algorithms, SHA 256 and SHA 512, are regarded as safe in other contexts and do not yet have problems with collusion.

Each block in a blockchain is made up of the headers listed below:

Previous Hash: The hash address used to locate the previous block.

Transaction Details: Information regarding each transaction that occurs.

Nonce: An arbitrarily assigned integer, determined through cryptography, used to differentiate the hash address of a block.

3.1.9 Block Hash Address

Therefore mentioned information, such as the nonce, transaction specifics, and prior hash, is sent using a hashing technique. This generates an output with a length of 64 characters (256 bits), which is referred to as the unique "hash address".

All throughout the world, a lot of people try to employ computational processes to figure out what hash value is right in order to meet a predetermined requirement. The transaction is complete when the predetermined condition is met.

4 PERFORMANCE ANALYSIS

The encryption and decryption periods are crucial factors in evaluating the effectiveness of the Blowfish encryption method. Fig .3. Comparison chart for Encryption Time shows Encryption time, the time required converting plaintext data into, and the time required to reverse the process and return ciphertext to plaintext is referred to as decryption time. In this case, several symmetric cryptographic algorithms were compared to the Blowfish algorithm. The

suggested Blowfish technique performs better in terms of encryption and decryption time, Fig .4 Comparison chart for Decryption Time as seen in the graph below.

Table 1: Table for Encryption Time based on Various Symmetric Cryptography Algorithm

File Size	Blowfish	DES	AES
10 KB	1.5	2	2
13 KB	2	2.5	2
39 KB	3	6.5	3.5
56 KB	3.7	9.3	4.5

Encryption Time Comparison

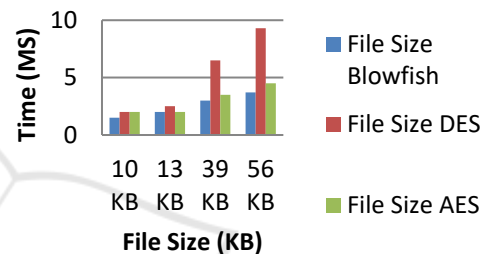


Figure 3: Comparison chart for Encryption Time

Table 2: Table for Decryption Time based on Various Symmetric Cryptography Algorithm

File Size	Blowfish	DES	AES
10 KB	1.3	1.7	2
13 KB	1.7	2	2
39 KB	3.2	6.8	4
56 KB	3.5	8.5	4.7

Decryption Time Comparison

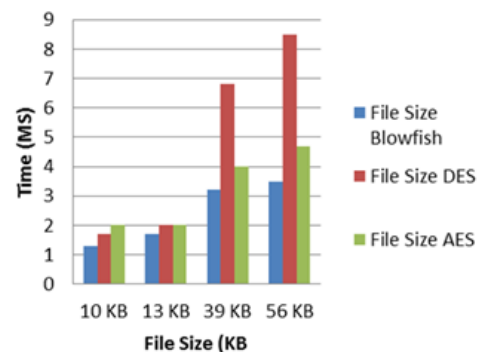


Figure 4: Comparison chart for Decryption Time

5 CONCLUSION

The integration of Blowfish encryption with blockchain technology, coupled with secure indexing and identity management, provides a robust solution for securely managing and sharing sensitive medical data. The use of Blowfish encryption ensures that medical data remains confidential even in the case of unauthorized access, while blockchain technology guarantees data integrity by providing a tamper-proof ledger of transactions and access requests. The implementation of secure indexing enables efficient and privacy-preserving keyword searches on encrypted data, maintaining the confidentiality of sensitive information. Furthermore, the key verification process and access control mechanisms add an extra layer of security by preventing unauthorized access and notifying data owners of any suspicious activities. This approach successfully balances data protection with usability, ensuring that authorized users can retrieve necessary information without compromising privacy. The proposed method offers an effective and secure solution for cloud-based medical data management, addressing both data privacy concerns and the need for efficient data retrieval, making it a valuable tool in safeguarding sensitive healthcare data in today's digital age.

REFERENCES

- Huang, Y., Song, X., Ye, F., Yang, Y. and Li, X., 2019. "Fair and efficient caching algorithms and strategies for peer data sharing in pervasive edge computing environments". *IEEE Transactions on Mobile Computing*, 19(4), pp.852-864.
- Wang, C., Yang, Y. and Zhou, P., 2020. Towards efficient scheduling of federated mobile devices under computational and statistical heterogeneity. *IEEE Transactions on Parallel and Distributed Systems*, 32(2), pp.394-410.
- Xu, G., Li, H., Ren, H., Lin, X. and Shen, X., 2020. DNA similarity search with access control over encrypted cloud data. *IEEE Transactions on Cloud Computing*, 10(2), pp.1233-1252.
- Miao, Y., Liu, X., Choo, K.K.R., Deng, R.H., Li, J., Li, H. and Ma, J., 2019. Privacy-preserving attribute-based keyword search in shared multi-owner setting. *IEEE Transactions on Dependable and Secure Computing*, 18(3), pp.1080-1094.
- Li, Z. and Yang, Y., 2018. RRect: A novel server-centric data center network with high power efficiency and availability. *IEEE Transactions on Cloud Computing*, 8(3), pp.914-927.
- Qiu, S., Wang, B., Li, M., Liu, J. and Shi, Y., 2017. Toward practical privacy-preserving frequent itemset mining on encrypted cloud data. *IEEE Transactions on Cloud Computing*, 8(1), pp.312-323.
- Chaudhari, P. and Das, M.L., 2019. Privacy preserving searchable encryption with fine-grained access control. *IEEE Transactions on Cloud Computing*, 9(2), pp.753-762.
- Li, P., Guo, S., Yu, S. and Zhuang, W., 2015. Cross-cloud mapreduce for big data. *IEEE Transactions on Cloud Computing*, 8(2), pp.375-386.
- Miao, Y., Ma, J., Liu, X., Li, X., Jiang, Q. and Zhang, J., 2017. Attribute-based keyword search over hierarchical data in cloud computing. *IEEE Transactions on Services Computing*, 13(6), pp.985-998.
- He, K., Guo, J., Weng, J., Weng, J., Liu, J.K. and Yi, X., 2018. Attribute-based hybrid boolean keyword search over outsourced encrypted data. *IEEE Transactions on Dependable and Secure Computing*, 17(6), pp.1207-1217.
- Wang, H., Ning, J., Huang, X., Wei, G., Poh, G.S. and Liu, X., 2019. Secure fine-grained encrypted keyword search for e-healthcare cloud. *IEEE Transactions on Dependable and Secure Computing*, 18(3), pp.1307-1319.
- Shu, J., Jia, X., Yang, K. and Wang, H., 2018. Privacy-preserving task recommendation services for crowdsourcing. *IEEE Transactions on Services Computing*, 14(1), pp.235-247.
- Ma, R. and Du, L., 2022. Attribute-based blind signature scheme based on elliptic curve cryptography. *IEEE Access*, 10, pp.34221-34227.
- Bhatt, S., Pham, T.K., Gupta, M., Benson, J., Park, J. and Sandhu, R., 2021. Attribute-based access control for AWS internet of things and secure industries of the future. *IEEE Access*, 9, pp.107200-107223.
- Fugkeaw, S., 2021. A lightweight policy update scheme for outsourced personal health records sharing. *IEEE Access*, 9, pp.54862-54871.
- Khan, S., Khan, S., Zareei, M., Alanazi, F., Kama, N., Alam, M. and Anjum, A., 2021. ABKS-PBM: Attribute-based keyword search with partial bilinear map. *IEEE Access*, 9, pp.46313-46324.
- Ra, G., Kim, D., Seo, D. and Lee, I., 2021. A federated framework for fine-grained cloud access control for intelligent big data analytic by service providers. *IEEE Access*, 9, pp.47084-47095.
- Zhao, J., Zeng, P. and Choo, K.K.R., 2021. An efficient access control scheme with outsourcing and attribute revocation for fog-enabled E-health. *IEEE Access*, 9, pp.13789-13799.
- Sahi, A., Lai, D. and Li, Y., 2021. A Review of the State of the Art in Privacy and Security in the eHealth Cloud. *Ieee Access*, 9, pp.104127-104141.
- Khan, A.W., Khan, M.U., Khan, J.A., Ahmad, A., Khan, K., Zamir, M., Kim, W. and Ijaz, M.F., 2021. Analyzing and evaluating critical challenges and practices for software vendor organizations to secure big data on cloud computing: An AHP-based systematic approach. *IEEE Access*, 9, pp.107309-107332.
- Maalini, D. and Balraj, E., 2018. Secured and Energy Efficient Packet Transmission in Wireless Sensor

- Networks using Flooding protocol and AES Algorithm. *Journal of Computational Information Systems*, 14 (4) 7, 13.
- Balraj, E. and Maalini, D., 2018. A survey on predicting student dropout analysis using data mining algorithms. *International Journal of Pure and Applied Mathematics*, 118(8), pp.621-626.
- Rayavel, P., Anbarasi, N., Renukadevi, B. and Maalini, D., 2021, July. Raspberry pi based secured cloud data. In *Journal of Physics: Conference Series* (Vol. 1964, No. 4, p. 042101). IOP Publishing.
- Manivannan, K., Ramkumar, K. and Krishnamurthy, R., 2024. Enhanced AI Based Diabetic Risk Prediction Using Feature Scaled Ensemble Learning Technique Based on Cloud Computing. *SN Computer Science*, 5(8), p.1123.
- Mrs Maalini Dharmaraj, Mr Mohanraj Seetharam, Mr Venkatesh Annamuthu, Mr Rajesh Veerasamy "Fingerprint based Anti Theft for Two Wheelers Authentication of Vehicle users", 2023.
- Mary, P.A., Maalini, D., Manivannan, K., Umamaheswari, K., Balaji, S. and Deepak, R., 2024, October. A Secure Aware Routing Protocol for Efficient and Reliable Fpanets. In *2024 First International Conference on Software, Systems and Information Technology (SSITCON)* (pp. 1-6). IEEE.
- Manivannan, K., Gowda, V.D., Pavan, B.V., Aravindh, S., Nithisha, C. and chaithanya Tanguturi, R., "Enhanced Agricultural Methods and Sustainable Farming Through IoT and AI Technology," in *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, pp. 1206-1212, IEEE, 2024.