
Research Paper**Decentralized based Security Mechanism for Linking Sensitive Data using Blockchain Technology****Ashwini Patil^{1*}**, **Vijay Shelake²**^{1,2}Dept. of Computer Engineering, Mumbai University, Thane, India*Corresponding Author: ashwinipatil090697@gmail.com**Received:** 20/Jul/2023; **Accepted:** 25/Aug/2023; **Published:** 30/Sept/2023. **DOI:** <https://doi.org/10.26438/ijcse/v11i9.2227>

Abstract: Privacy preservation practises intensify as data privacy infractions are an increasing source of worry. A lot of organisations gather a lot of data. Organisations occasionally utilise this data for a variety of activities. However, the data gathered may include sensitive or private information that has to be secured. If we disclose data for sharing purposes, privacy protection is a crucial problem. There are several methods for protecting privacy, but they are vulnerable to different kinds of assaults and data loss. In this research, we suggested a practical method for maintaining privacy via homomorphic encryption. Our method safeguards sensitive data with little information loss, improving data usefulness and guarding against many sorts of assault.**Keywords:** Privacy preservation, linking sensitive data, Homomorphic Encryption, Blockchain

1. Introduction

Increased study into techniques that securely compute meaningful information is a result of growing privacy and security awareness. When it comes to this data, it is a valuable asset, and this is especially true in this day and age when cloud computing, big data, and the Internet of Things are embracing one another. One of the cutting-edge technologies bringing about transformation in the field of data analytics is big data. Data security and privacy face significant problems in this historically unprecedented period of technology convergence. It has the ability to offer perceptions into the hidden facets of data analysis; they turn into a resource that can be used to make organisations smarter and therefore more successful. Online transactions, search queries, mobile devices, emails, movies, and other sources all contribute to the creation of big data. It is kept through partitioning among different servers. This suggests that there are several massive data sources and storage systems on the internet. Because big data analytics heavily relies on the online platform, security problems are a significant obstacle. Big data analytics that are not secure might result in significant losses for both individuals and organisations. This requirement has grown crucial in a variety of situations, such as integrating data on the Web and at businesses, developing online markets, sharing data for scientific research, exchanging data at government agencies, monitoring health emergencies, and enhancing homeland security. [1] Hackers and intruders may find the value buried in large data to be very valuable.

The use of social media, search engines, and other online platforms by individuals is tracked and included in big data. Data de-identification or anonymization is also useful for obscuring personal data. It involves altering data that will be utilised or disseminated in a way that makes it impossible to identify important information [3]. As an illustration, in 2013, Yahoo suffered a data breach that exposed the data of over 3 billion users, or about half of the world's population. And this incident is just one example of countless data breach events [1]. [6] However, much of the collected information may be sensitive private data, which raises privacy concerns. Blockchain technology has been widely employed in voting, supply chain, healthcare, IoT, and other applications due to its appealing properties, such as transparency, anonymity, autonomy, and tamper-proofness [4]. The blockchain prevents arbitrary data modification by participants by validating all transactions through a consensus process in an untrusted environment. Blockchain is designed to operate in a decentralised network of computing nodes, making it resistant to errors and intrusions. Additionally, the blockchain and smart contracts can improve the interoperability of health data. Additionally, it benefits from cheaper costs, better performance, and scalability [5]

From medical screening [5] to disease outbreak detection [32], advancements in machine learning models trained on sensitive real-world data are promised. Additionally, more sensitive and richer data is becoming accessible due to the increased usage of mobile devices [30]. Sensitive data collecting on a big scale is risky, though. Today's internet video streaming services are subject to restrictions imposed

by the statute that was implemented in reaction to the occurrence [28][4]. A recent examination [29] shown an increase in the annual amount of medical records that were made public. There are currently more than one healthcare data breaches every day. Therefore, the majority of healthcare providers and hospitals decide to construct their healthcare systems in a closed domain with a defensive perimeter, such as a private network outfitted with firewalls and intrusion detection systems, to improve security protections and prevent privacy leaks. [7][8]. The information of 3.5 million users to the social network "Cyworld" was exposed as a consequence of hacker attacks on the network and privacy leaks from Google, the CSDN network, and other sources. [2]. The urgent need is to provide systems that allow for mass data integration and exchange, especially in areas of national importance, while allowing individuals to easily and effectively maintain their privacy [1].

2. Related Work

A block chain technique employing the AES algorithm was proposed by Aditi Shinde et al. [1]. Each section of the report will be independently encrypted using AES to help safeguard the data and retain confidentiality. This solution not only made it easier to retain paper documents, but it also offered a blockchain-based security system.

A Privacy-Preserving Deep Learning was developed by Le Trieu Phong et al. [2] using Paillier encryption, LWE-based encryption, and additively homomorphic encryption. This system connects encryption with deep learning. Utilised, additively homomorphic encryption in conjunction with asynchronous stochastic gradient descent on neural networks. And it has been demonstrated that using encryption results in a reasonable overhead for a regular deep learning system.

K-anonymity Algorithm was introduced by Sang Ni et al. [3] for the protection of privacy. Author suggested a parallelized version of the clustering-based K-anonymity technique. The testing findings indicated that the algorithm outperformed the KACA and Incognito algorithms in terms of information loss and performance.

For the purpose of protecting privacy, Huda O. Mansour et al. [4] created a quasi-identifier recognition method. Attribute categorization and QID dimension identification made up the two primary steps of the proposed method. The algorithm's foundation for operation is the reidentification of risk rates for all characteristics and the dimensions of QIDs, from which it derives the correct QIDs and their appropriate dimensions. A actual dataset was used to test the suggested approach. In comparison to more current comparable algorithms, the findings showed that the suggested approach greatly lowers privacy leakage while maintaining the usefulness of the data.

Through the use of the best geometric transformations, M.A.P. Chamikara et al. [5] create an effective and scalable no reversible perturbation technique for protecting large data privacy. In order to solve the efficiency, Scalability, Privacy,

Usability, and Utility concerns of the existing data perturbation methods, a new perturbation method called PABIDOT was introduced. The suggested technique performs better than random rotation perturbation and geometric perturbation in terms of classification accuracy, coming very near to that of the original dataset. Additional benefits of the suggested algorithm are its composition and privacy guarantee. According to empirical findings, PABIDOT offers strong defence against a variety of privacy assaults.

With the completion of two processes, namely data sanitization and data restoration, Mohana Shivashankar et al. [6] established an enhanced model for privacy protection of massive data. Sensitive data held in large databases is protected by the data sanitization process by being hidden from unauthorised users. The process of retrieving or restoring data that has been sanitised at the sender side is known as data restoration. In terms of secrecy, a suitable key is required to conceal the sensitive data on both the sender's and receiver's sides. The same key is needed to recover the sanitised data after data sanitization. The best key generation is so essential to maintaining privacy protection. In this study, a modified Rider optimisation algorithm (ROA) called the Randomised ROA (RROA) model is used to select the best key.

By dividing a deep neural network into a classifier module that runs in the cloud and a feature extractor module that needs be installed on the user's device, Seyed Ali Osiaet al. [7] introduced a new hybrid architecture for effective privacy-preserving mobile analytics. It took advantage of DNNs' characteristics, particularly convolutional neural networks, to gain from their precision and layered design. Siamese fine-tuning was utilised to create a unique feature well-suited for the primary purpose but unfit for any other secondary duties in order to safeguard the data privacy against unauthorised tasks. In contrast, the properties of regular deep neural networks are general and may be used to a variety of applications. Users' privacy is protected by removing the unwanted sensitive information from the derived feature. The author examined several embedding approaches on various layers of pre-trained state-of-the-art models for gender categorization and activity identification and offered three ways to test the privacy of the proposed framework. finally shown that the framework was capable of achieving a respectable accuracy/privacy trade-off.

In order to construct PPML, Joon-Woo Lee et al. [8] adapted the RNS-CKKS scheme, a cutting-edge FHE strategy, to the common deep neural network ResNet-20. Bootstrapping and SoftMax functions have not been used on the PPML models due to the ReLU function's more accurate approximations. Up to this point, the author used a variety of adjusted settings to apply these strategies. And then demonstrated that the PPML models with the word-wise FHE scheme had the greatest classification accuracy among the ResNet-20 implemented with the RNS-CKKS scheme, which had almost the same outcome as the original ResNet-20.

An novel blockchain-based safe and privacy-preserving data exchange method for smart cities called "Privy Sharing" was introduced by Keith Bonawitz et al. [9]. The suggested approach makes sure that sensitive or personal user data is protected, processed securely, and disclosed to interested parties only, when necessary, in accordance with user-defined ACL rules included in smart contracts. Additionally, data owners receive compensation for sharing their information with stakeholders and other parties. Additionally, Privy Sharing conforms with a number of essential EU GDPR criteria, including sharing data assets, accessibility, and deletion with the owner's agreement. A multi-Ch blockchain solution scales better than a single Ch blockchain system, according to the testing data.

To safeguard shared model parameters in a Federated Learning-based system, Jaehyoung Park et al. [10] created the Privacy-Preserving Federated Learning system based on the Homomorphic Encryption technique. A method for the secure aggregation of local model parameters encrypted with several keys inside the same Federated Learning-based system was also put forth by the author. The computational and communication expenses necessary to increase security level in Federated Learning were theoretically analysed in the proposed system, and simulations were used to assess the performance of the proposed PPFL algorithm in terms of overhead.

For the purpose of protecting the data privacy for the Identical Generalisation Hierarchy (IGH) data, Waranya Mahanan et al. [11] devised a heuristic approach. The k-anonymity approach is the foundation of the proposed study. The suggested work generates the best solutions by traversing the generalisation lattice to discover the least amount of information loss. The technique takes use of the properties of IGH data, and it always finds the best solution at the lowest level of k-anonymous nodes of the generalisation lattice. Our approach may identify the best answer more quickly than other algorithms since the process has the potential to skip some nodes. The suggested work also employs an in-order traversal since it alternately searches the border of the generalised lattice between the upper and lower levels. As a result, there is a greater chance of finding the k-anonymous nodes, and the efficiency of the suggested approach may be as high as 21%.

Suneetha V. and others [12] The suggested method uses K-anonymization and L-diversity to mask the sensitive personal data, as well as Apache Spark to handle the large amounts of health care data quickly and effectively. This ensures that sensitive data is isolated before being sent to HDFS and that shared data does not disclose the real data.

A blockchain-based multi-Wireless Sensor Network (WSN) identity authentication technique was put up by Zhihua Cui et al. [13]. In order to create a private blockchain amongst cluster heads in a single WSN and add base stations from all WSNs to the public blockchain, it is necessary to combine the decentralisation of blockchain with the distributed structure of IoT nodes. A hybrid blockchain is built using the whole

network. In this paradigm, communication authentication between nodes and the registration of identity information between cluster head nodes and regular nodes are finished. The scheme had high security and efficiency, according to the security and performance study.

Unmanned Traffic Management (UTM)-Chain, a blockchain-based security solution for an unmanned traffic management system, was presented by AzzaAllouch et al. [14]. Using UTM components and blockchain technology, a reliable and secure traffic monitoring system has been created. A roadmap for the effective and secure handling of UAV flight data records is also provided by this study. A permissioned blockchain network serves as the foundation of the suggested design. The suggested architecture makes use of blockchain technology to manage access to flight data that is kept in a decentralised off-chain database by executing smart contracts, in addition to using it to store transactions. As centralised security risks are reduced, a decentralised off-chain database and decentralised blockchain network raise the level of security of the suggested design. Additionally, a number of tests were conducted to assess the UTM-Chain framework's performance using cAdvisor in terms of latency and resource use. The outcome demonstrates that utilising blockchain technology may enhance the functionality of the platform being used. Security research demonstrates that the UTM-Chain solution offers an unmanned traffic management system with a secure, effective, dependable, and tamper-resistant data security practise.

Haya R. Hasan and others [15] Infectious illness propagation is slowed down by the suggested strategy. This study offered four smart contracts that make use of on-chain events and alerts while utilising very little on-chain storage. With this strategy, the Author connected the participating entities' distinctive Ethereum addresses to their self-sovereign identities, re-encryption proxies, and related biometric data. This approach lays the way for effective solutions that can aid in preventing the spread of illnesses by accurately and promptly documenting occurrences in a tamper-proof manner.

A innovative conditional privacy-preserving authentication (CPPA) mechanism built on the blockchain was developed by Chao Lin et al. [16] with the goal of facilitating safe communication on VANETs. To build a revolutionary BCPPA protocol, the Author specifically combined blockchain and key derivation algorithms. Elliptic Curve Digital Signature Algorithm (ECDSA), which was utilised as the foundation for the proposed BCPPA protocol, can alternatively be substituted by some modified ECDSA with batch verification to increase performance. Additionally proved the usefulness and security of the suggested protocol.

Jiafu Wan et al [17] Developed an innovative blockchain-based IIoT architecture to help build a more secure and reliable IIoT system. Author introduced a new IIoT architecture and give a detailed analysis of all architecture layers. Also, introduced BLP model as well as Biba model to design secure assurance in theory. On this basis, Author

described the key technologies, the flow, and the defense mechanisms of the proposed architecture.

Chao Lin and co. A smart house was used to demonstrate how blockchain may be used in conjunction with other methods to provide mutual authentication between users and the home gateway. The suggested approach uses a GS and MAC, respectively, to authenticate a requestor without disclosing information about the particular member or the home gateway with complete forward secrecy. The method also makes it possible to locate any users who are later discovered to be acting inappropriately. eventually proved the usefulness and security of the suggested system.

Using a cluster structure and a novel routing protocol, Abbas Yazdinejad et al. [19] developed an energy-efficient and secure blockchain-enabled architecture of Software Defined Networking (SDN) controllers for IoT networks. By eliminating Proof-of-Work (POW) and using an effective authentication method with distributed trust, the architecture uses public and private blockchains for peer-to-peer (P2P) communication between IoT devices and SDN controllers. This makes the blockchain suitable for resource-constrained IoT devices. According to the experimental findings, the routing protocol based on the cluster structure has a faster throughput, a shorter latency, and consumes less energy. In other words, it is shown that the suggested design performs better than traditional blockchain.

A framework was developed by Ayesha Shahnaz et al. [20] that might be utilised to apply blockchain technology for Electronic Health Records (EHR) in the healthcare industry. The suggested framework aims to use blockchain technology for EHR in the first place and to offer safe storage of electronic data by outlining specific access guidelines for users. The advantages of having a scalable, secure, and integrative blockchain-based solution are provided by this framework for the EHR system.

3. Proposed Research Methodology

The main aspect of the research is to safeguard the privacy data, normally; the input data is gathered and fed into the Quasi identifier, which is a subset for attributes to determine the unique entities in the real world. Quasi identifier has the capability to separate the data as sensitive and non-sensitive. Non-sensitive data doesn't need any security purposes but the sensitive data required to secure data from theft under various regions. Homomorphic encryption (HE) is used to protect the data through privacy preservation, and then for further improvements in the cornered sensitive data this HE is induced through the hybridized form of optimization such as sparrow as well as Crow search algorithm.

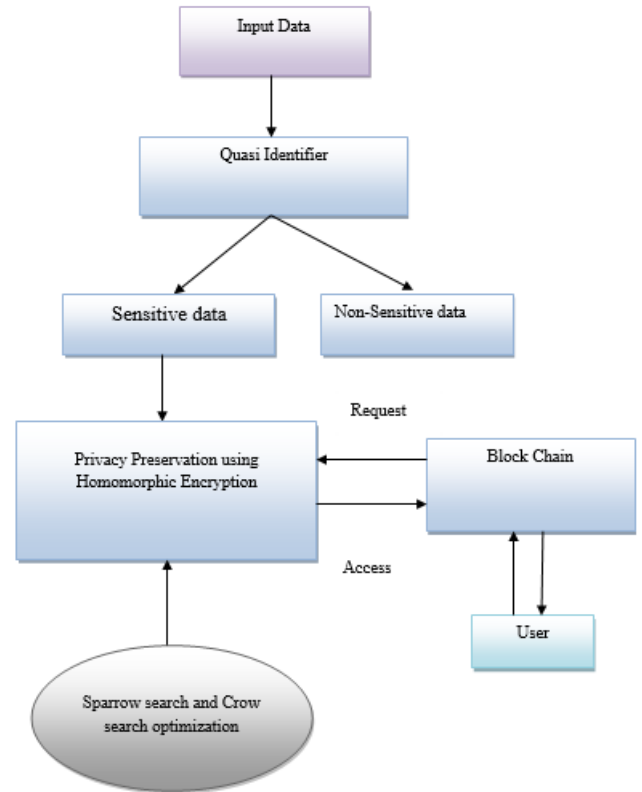


Figure 1.1: System Architecture

Blockchain is considered as a decentralized, immutable, and public ledger, in which the transactions are stored in chained blocks without the existence of a trusted central authority. Whenever the user passes the data to block chain, then it will directly inject the data to the privacy preservation layer for observing whether the received data is from the authorized user or not. If the received information is from the authorized user, then blockchain permit the user to access the system in a secured manner otherwise it won't permit the user for transaction. The research will be implemented in python and the efficiency will be proved by using the metrics accuracy, specificity, sensitivity and precision.

4. Results and Discussion

4.1 Encryption:

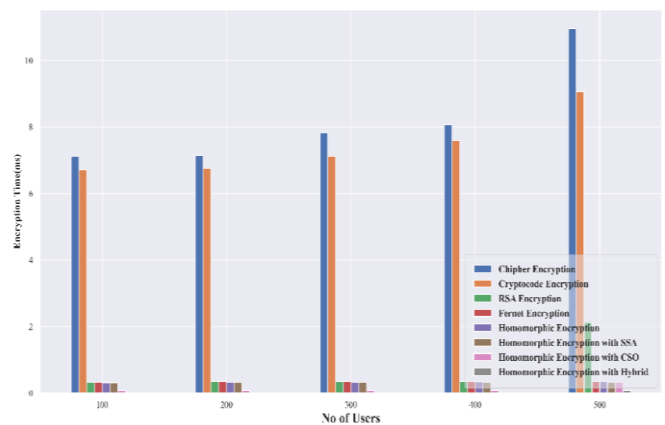


Figure 2.

4.2 Memory Usage:

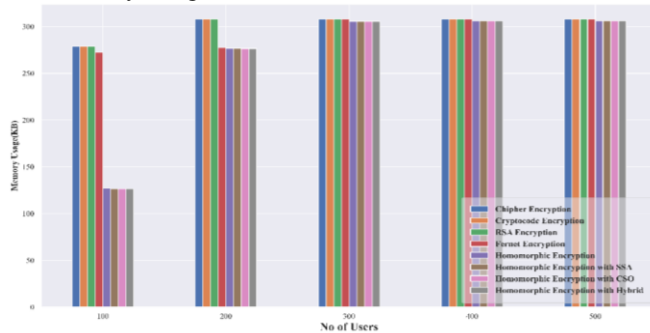


Figure 3.

4.3 Encryption Time:

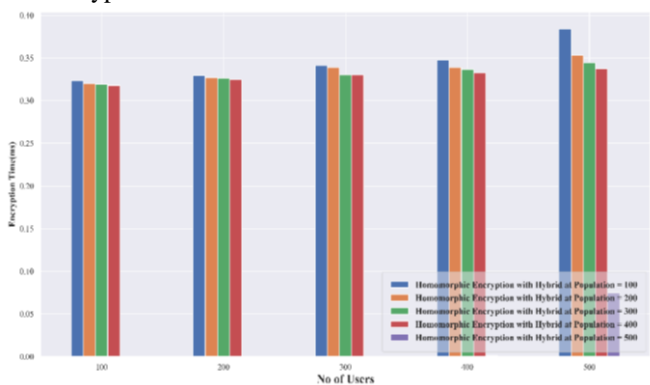


Figure 4.

5. Conclusion

If we disclose data for sharing purposes, privacy protection is a crucial problem. There are several methods for protecting privacy, but they are vulnerable to different kinds of assaults and data loss. In this research, we suggested a practical method for maintaining privacy via homomorphic encryption. Our method safeguards sensitive data with little information loss, improving data usefulness and guarding against many sorts of assault. The outcomes showed that the suggested method retains the usefulness of the data while drastically reducing privacy leaks. In light of this, the new method outperformed the current algorithms.

References

- Clifton, C., Kantarcioğlu, M., Doan, A., Schadow, G., Vaidya, J., Elmagarmid, A. and Suci, D., 2004, June. Privacy-preserving data integration and sharing. In *Proceedings of the 9th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, pp.19-26, 2004.
- Fang, W., Wen, X.Z., Zheng, Y. and Zhou, M., 2017. A survey of big data security and privacy preserving. *IETE Technical Review*, Vol.34, Issue.5, pp.544-560, 2017.
- Gosain, A. and Chugh, N., 2014. Privacy preservation in big data. *International Journal of Computer Applications*, Vol.100, Issue.17, 2014.
- Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H.B., Patel, S., Ramage, D., Segal, A. and Seth, K., 2017, October. Practical secure aggregation for privacy-preserving machine learning. In *proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp.1175-1191, 2017.
- Liu, J., Li, X., Ye, L., Zhang, H., Du, X. and Guizani, M., 2018, December. BPDS: A blockchain based privacy-preserving data sharing for electronic medical records. In *2018 IEEE Global Communications Conference (GLOBECOM)*, IEEE, pp.1-6, 2018.
- Zhang, A. and Lin, X., 2018. Towards secure and privacy-preserving data sharing in e-health systems via consortium blockchain. *Journal of medical systems*, Vol.42, Issue.8, p.140, 2018.
- Jin, H., Luo, Y., Li, P. and Mathew, J., 2019. A review of secure and privacy-preserving medical data sharing. *IEEE Access*, 7, pp.61656-61669, 2019.
- Park, J. and Lim, H., 2022. Privacy-preserving federated learning using homomorphic encryption. *Applied Sciences*, Vol.12, Issue.2, p.734, 2022.
- Mendes, R. and Vilela, J.P., 2017. Privacy-preserving data mining: methods, metrics, and applications. *IEEE Access*, 5, pp.10562-10582, 2017.
- Mahanan, W., Chaovalitwongse, W.A. and Natwichai, J., 2021. Data privacy preservation algorithm with k-anonymity. *World Wide Web*, 24, pp.1551-1561, 2021.
- Sharma, M., Chaudhary, A., Mathuria, M., Chaudhary, S. and Kumar, S., 2014, July. An efficient approach for privacy preserving in data mining. In *2014 International Conference on Signal Propagation and Computer Technology (ICSPCT 2014)*, IEEE, pp.244-249, 2014.
- Suneetha, V., Suresh, S. and Jhananie, V., 2020, March. A novel framework using apache spark for privacy preservation of healthcare big data. In *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, IEEE, pp.743-749, 2020.
- Yadav, D., Shinde, A., Nair, A., Patil, Y. and Kanchan, S., 2020, May. Enhancing data security in cloud using blockchain. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, pp.753-757, 2020.
- Aono, Y., Hayashi, T., Wang, L. and Moriai, S., 2017. Privacy-preserving deep learning via additively homomorphic encryption. *IEEE transactions on information forensics and security*, Vol.13, Issue.5, pp.1333-1345, 2017.
- Ni, S., Xie, M. and Qian, Q., 2017. Clustering Based K-anonymity Algorithm for Privacy Preservation. *Int. J. Netw. Secur.*, Vol.19, Issue.6, pp.1062-1071, 2017.
- Mansour, H.O., Siraj, M.M., Ghaleb, F.A., Saeed, F., Alkhamash, E.H. and Maarof, M.A., 2021. Quasi-Identifier recognition algorithm for privacy preservation of cloud data based on risk reidentification. *Wireless Communications and Mobile Computing*, pp.1-13, 2021.
- Chamikara, M.A.P., Bertok, P., Liu, D., Camtepe, S. and Khalil, I., 2020. Efficient privacy preservation of big data for accurate data mining. *Information Sciences*, 527, pp.420-443, 2020.
- Shivashankar, M. and Mary, S.A., 2021. Privacy preservation of data using modified rider optimization algorithm: Optimal data sanitization and restoration model. *Expert Systems*, Vol.38, Issue.3, p.e12663, 2021.
- Osia, S.A., Shamsabadi, A.S., Sajadmanesh, S., Taheri, A., Katevas, K., Rabiee, H.R., Lane, N.D. and Haddadi, H., 2020. A hybrid deep learning architecture for privacy-preserving mobile analytics. *IEEE Internet of Things Journal*, Vol.7, Issue.5, pp.4505-4518, 2020.
- Lee, J.W., Kang, H., Lee, Y., Choi, W., Eom, J., Deryabin, M., Lee, E., Lee, J., Yoo, D., Kim, Y.S. and No, J.S., 2022. Privacy-preserving machine learning with fully homomorphic encryption for deep neural network. *IEEE Access*, 10, pp.30039-30054, 2022.
- Makhdoom, I., Zhou, I., Abolhasan, M., Lipman, J. and Ni, W., 2020. PrivySharing: A blockchain-based framework for privacy-preserving and secure data sharing in smart cities. *Computers & Security*, 88, p.101653, 2020.
- Park, J. and Lim, H., 2022. Privacy-preserving federated learning using homomorphic encryption. *Applied Sciences*, Vol.12, Issue.2, p.734, 2022.

- [23]. Mahanan, W., Chaovalitwongse, W.A. and Natwichai, J., 2021. Data privacy preservation algorithm with k-anonymity. *World Wide Web*, 24, pp.1551-1561, 2021.
- [24]. Suneetha, V., Suresh, S. and Jhananie, V., 2020, March. A novel framework using apache spark for privacy preservation of healthcare big data. In 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), IEEE, pp.743-749, 2020.
- [25]. Cui, Z., Fei, X.U.E., Zhang, S., Cai, X., Cao, Y., Zhang, W. and Chen, J., 2020. A hybrid blockchain-based identity authentication scheme for multi-WSN. *IEEE Transactions on Services Computing*, Vol.13, Issue.2, pp.241-251, 2020.
- [26]. Allouch, A., Cheikhrouhou, O., Koubâa, A., Toumi, K., Khalgui, M. and Nguyen Gia, T., 2021. Utm-chain: blockchain-based secure unmanned traffic management for internet of drones. *Sensors*, Vol.21, Issue.9, p.3049, 2021.
- [27]. Hasan, H.R., Salah, K., Jayaraman, R., Arshad, J., Yaqoob, I., Omar, M. and Ellahham, S., 2020. Blockchain-based solution for COVID-19 digital medical passports and immunity certificates. *Ieee Access*, 8, pp.222093-222108, 2020.
- [28]. Lin, C., He, D., Huang, X., Kumar, N. and Choo, K.K.R., 2020. BCPPA: A blockchain-based conditional privacy-preserving authentication protocol for vehicular ad hoc networks. *IEEE Transactions on Intelligent Transportation Systems*, Vol.22, Issue.12, pp.7408-7420, 2020.
- [29]. Wan, J., Li, J., Imran, M. and Li, D., 2019. A blockchain-based solution for enhancing security and privacy in smart factory. *IEEE Transactions on Industrial Informatics*, Vol.15, Issue.6, pp.3652-3660, 2019.
- [30]. Lin, C., He, D., Kumar, N., Huang, X., Vijayakumar, P. and Choo, K.K.R., 2019. HomeChain: A blockchain-based secure mutual authentication system for smart homes. *IEEE Internet of Things Journal*, Vol.7, Issue.2, pp.818-829, 2019.
- [31]. Yazdinejad, A., Parizi, R.M., Dehghantanha, A., Zhang, Q. and Choo, K.K.R., 2020. An energy-efficient SDN controller architecture for IoT networks with blockchain-based security. *IEEE Transactions on Services Computing*, Vol.13, Issue.4, pp.625-638, 2020.
- [32]. Shahnaz, A., Qamar, U. and Khalid, A., 2019. Using blockchain for electronic health records. *IEEE access*, 7, pp.147782-147795, 2019.