JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

LEVERAGING BLOCKCHAIN FOR SECURING DIGITAL FORENSIC EVIDENCE IN IOT SYSTEMS

[1]Vismaya S, [2]Ms Aparna A
[1]MCA Scholar, [2]Associate Professor
Department of MCA
Nehru College of Engineering and Research Centre, Pambady, India

Abstract : Digital evidence is what digital forensics is all about. The study of data gathering, processing, analysis, reporting, and detection is known as digital forensics. promoting digital forensics' application in law enforcement inquiries. You can use digital forensics to determine what information was stolen and how it was distributed or duplicated. Some hackers intentionally destroy data in order to cause harm to their targets. Important data may inadvertently be corrupted in other situations by malicious software or hacker activity. Security and integrity are issues that digital forensics must deal with. Because IoT devices are secure and reliable, they can gather digital forensic evidence in an IoT environment, endangering cybercrime agencies. Although there is a risk to confidentiality, numerous studies have been conducted recently to enhance the integrity and security of IoT-based digital forensics. Digital forensics continues to suffer security and manipulation challenges, according to recent studies. Therefore, a strategic and efficient approach that not only safeguards integrity and security but also foresees dangers is required. Thus, using blockchain technology and hashing algorithms, we offer a clever and practical alternative. Blockchain will be used to store the information gathered by IoT devices. Models enhanced by machine learning will be used to forecast anomalies in the data and transactions. Thus, the ability of the suggested model to anticipate attacks early on makes it effective.

IndexTerms **– Digital Forensics, Blockchain Technology, IOT**

I. INTRODUCTION

In crimes involving digital components, such as stolen data, digital forensics plays a crucial role in assisting law enforcement by identifying, obtaining, evaluating, and reporting electronic evidence. It does, however, confront difficulties like deliberate data deletion by hackers, malware corruption, and security threats related to IoT devices utilized for evidence gathering. The security and integrity of IoT-based digital forensics have improved recently, but confidentiality is still a major worry.Blockchain technology has surfaced as a potentially effective remedy for these problems. According to studies, evidence can be stored on a blockchain by using tamper-resistant techniques to protect the integrity of the evidence and hot and cold blocks to increase security. While Blockchain-enabled chains of custody guarantee that evidence is kept undisturbed during legal proceedings, frameworks such as BLOFF prohibit manipulated records from being used in investigations.The general security of forensic data in social and cloud environments is improved by these developments, as is dependability in court.Even with advancements, tampering and security problems still exist, underscoring the necessity of proactive, intelligent systems to handle new threats. The massive amount of data produced by IoT devices calls for secure, decentralized data control techniques. Blockchain secures sensitive evidence and guarantees data reliability through cryptographic techniques including hashing, digital signatures, encryption, and timestamps. By enabling safe remote storage, analysis, and verification, these techniques increase the effectiveness of forensic investigations. Human mistake, however, is still a risk.Blockchain, hashing algorithms, and machine learning are all combined in this thesis to create an intelligent system. IoT device evidence will be safely saved on Blockchain, and machine learning models will examine transactions, spot irregularities, and anticipate assaults. By solving current shortcomings in data reliability and tamper resistance, this method guarantees the confidentiality, integrity, and security of digital evidence. The system seeks to improve forensic investigations by providing a reliable and effective solution to support legal and law enforcement procedures.

II. LITERATURE SURVEY

The detection and analysis of electronic evidence is greatly aided by digital forensics. Identification, acquisition, processing, and reporting of digital data—often crucial to law enforcement investigations—are all part of this area of forensics. It plays a crucial role in identifying instances of data theft, duplication, or distribution while tackling issues such as malevolent actors' deliberate data erasure or corruption. IoT systems, however, present security and integrity issues because of insufficient security controls, which may compromise forensic evidence that has been gathered.In order to improve the security and integrity of digital forensics, experts have looked into blockchain technology. To protect evidence, for example, Hahn (2010) suggested hot and cold Blockchain systems, while Hamid Lone and Naaz Mir (2017) created Blockchain-based techniques to guarantee evidence integrity. Agbedanu and Jurcut (2021) also presented BLOFF, a Blockchain-based forensics paradigm for IoT, to stop log manipulation. In 2020, Akhter et al. introduced a blockchain-based cloud solution for protecting digital forensics. Furthermore, Gopalan et al. (2019) promoted blockchain-powered chains of custody to stop data manipulation in courtrooms, and Nelson (2020) emphasized blockchain possibilities for gathering social media data from the Internet of Things.Time stamps, digital signatures, cryptography, data concealing, and data digestion are just a few of the strong techniques needed to preserve digital evidence. The dependability and integrity of the evidence are guaranteed by these methods, which are essential for court cases. To safeguard sensitive data, for instance, hashing and data encryption are frequently employed. However, human error—intentional or not—continues to be a major risk when handling data. Decentralized approaches are required for effective administration due to the massive volumes of data generated by the proliferation of IoT devices. Secure Internet of Things services are becoming more practical with the development of Blockchain technology. One way to improve data security is to preserve evidence at a safe distance from crime scenes and to do assessments remotely. Throughout the investigation, legal and technological experts stress the importance of preserving data originality and integrity. Blockchain technology and IoT-based digital forensics promise to improve security, guarantee the integrity of the evidence, and foresee possible threats, notwithstanding certain obstacles. Intelligent and economical approaches will advance digital forensics in IoT systems as the area develops.

III. METHODOLOGY

As the significance of data gathering and processing services for AI learning data increases, research into the intersection of blockchain and AI has just begun. This section, which presents pertinent research based on the proposed AI learning data environment model, links blockchain with AI data. Due to their ability to collect and store private data, IoT nodes are turning into a treasure trove for malicious actors. For an IoT network to be implemented successfully, identifying compromised nodes and gathering and preserving proof of an attack or malicious behavior have become critical tasks. The purpose of this work is to guarantee the integrity and security of the blockchain-based IoT forensics dataset by employing machine learning models to forecast harmful attacks.

3.1 Proposed Framework

In many forms of AI technology, including deep learning specifically, algorithms, computer systems, and data learning are all linked. To build an AI model with a certain characteristic, there needs to be a sufficient dataset available for AI learning. The AI machine learning procedure based on IoT integrity using blockchains is shown in Figure 3.1.

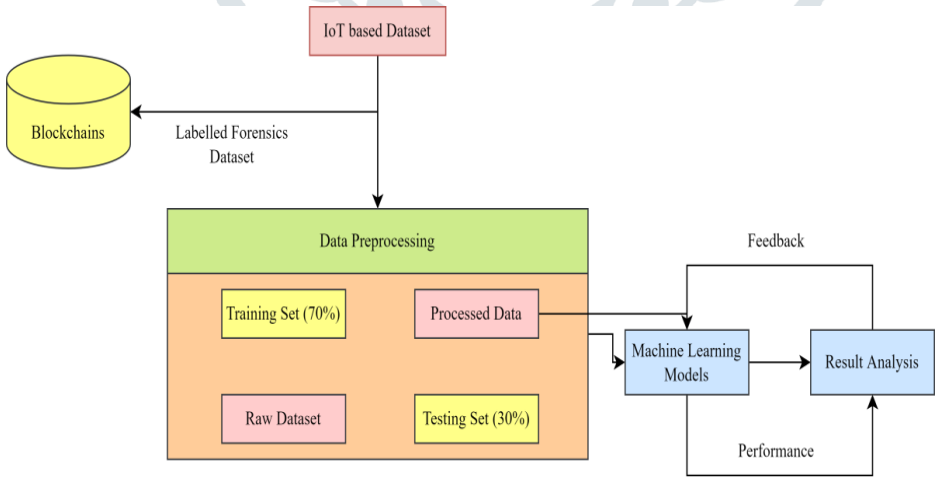


Figure 3.1 Proposed Framework

3.2 Data Collection

This stage allows for the collection of numerical data that is not organized. Preprocessing begins with the extraction of data appropriate for the goals and features of the AI being developed. Because of their extensive use, malicious third parties have been attacking Internet of Things (IoT) technology more frequently. For this threat to be effectively addressed, countermeasures such network intrusion detection systems must be created. A representative and well- structured dataset is necessary for both training and confirming the algorithms' dependability. Nothing is known about the Botnet situations that were used, even though there are a lot of network datasets available. This study proposes a novel dataset called Bot-IoT, which

comprises both simulated and actual Internet of Things (IoT) network traffic together with a range of assaults on that traffic. The limitations of the current dataset, including its inability to capture complete network information, accurate labeling, and complex attacks, are also demonstrated to be addressed by a realistic testbed environment.Lastly, we use a number of statistical and machine learning techniques to compare the BoT-IoT dataset's reliability to the benchmark datasets. Blockchain technology makes it simpler and more secure to keep data current and fresh while recording its history. No one can change or erase the data, and you get a current record that is always up to date in addition to a history of data.

.

3.3 Data Storage In Blockchain

Putting the data's hash on the blockchain is the most intelligent way to keep data. A hash code is created using our data as input. Because the data hash is low, the cost is low. Raw data can be stored on a file system as well. The solution to outdated centralized organizations, like banking institutions, is the decentralized technology of blockchain. On a blockchain network, a digital, central "authority" is created by fusing machine learning skills. Data may be distributed to everyone with an app that can access it thanks to the blockchain.Reading and writing access to this ledger can be either restricted ('permitted') or unrestricted ('permission less').A key need for an accurate machine learning model is that the data it utilizes be free of noise, duplication, and missing values; this is why the data is kept on a blockchain network. A hash function that uses cryptography is used to create the digital signature for every independent block. The range of hash functions is extensive.

.

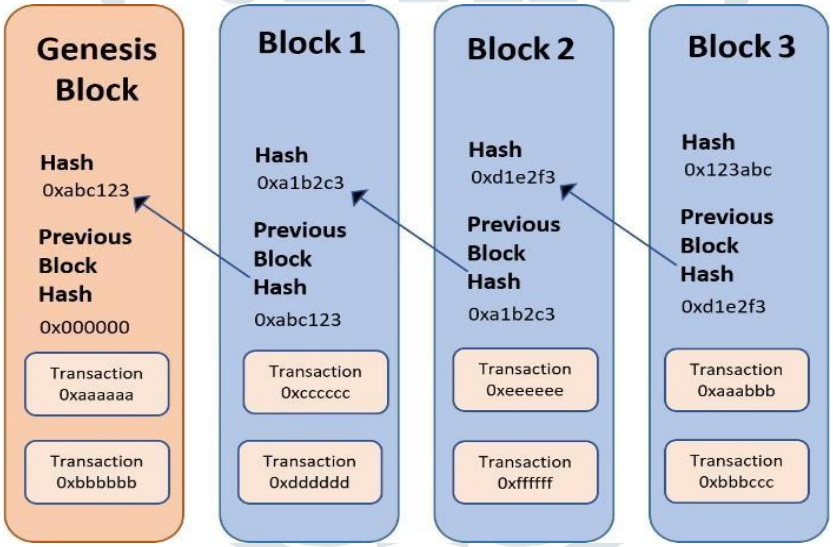


Figure 3.2 Blockchain and Hashing Architecture

3.4 Secured And Privacy Preserving

The proposed model uses an identity-based cryptosystem based on the elliptic curve cryptosystem, a lightweight public-key cryptosystem, to protect patient data. It is not necessary for the IBC to validate the recipient's public key. ECC-based arithmetic is about 20 times faster than modular exponentiation in terms of processing overhead. In addition, the bit length of the 128-bit ECC key makes it just as secure as a 1024-bit RSA key. IBC and ECC have certain qualities that can be advantageous for IoT applications.

3.5 Proposed Blockchain Model

It is the responsibility of P2P networks to ensure that communication between blockchain nodes is unhindered, even when nodes may be spread out across the globe and have equal access to the application. There is no central server in the P2P network, and every node serves as both an information source and an informed consumer. As part of the network's routing process, each node propagates and validates transactions, synchronizes data blocks, and creates and maintains connections with other nodes. Every node (as explained below, the blockchain's data structures include both transactions and blocks).Peer-to-peer networks' flat topology and decentralization are best illustrated by this. In a number of situations, blockchain apps offer APIs (application programming interfaces). Without having to worry about the underlying technical issues, users can interface withem directly using these APIs.

3.6 Cloud Blockchain Model

To securely store data, central databases are employed extensively. Hackers, however, are more concentrated. One of the most common methods hackers obtain vast amounts of data is through a script attack against a central database. However, cracking is considerably more challenging with blockchain and distributed ledger technologies. The goal of many blockchain initiatives is to improve the security of data storage. For users, this might be a game changer. People may access their data without any restrictions thanks to the blockchain effort, which may also result in safer data storing techniques. A number of blockchain initiatives mark up the initial cryptocurrency. This not only enables users to profit from third-party data, but also guards against identity theft and other problems brought up by recent widespread data breaches. The integrity and nonrepudiation of messages in blockchain transactions are guaranteed by digital signatures.

3.7 Data Preprocessing

Examples of data preprocessing at this stage of the machine learning model construction process include choosing or removing data properties, integrating existing data properties, and adding or removing missing data values.The data analysis at this level enables artificial intelligence (AI) to leverage knowledge obtained through exploration and inference as well as data patterns discovered in traditional datasets.Raw data has noise, lacks a consistent structure, and is regularly reexamined, making it unsuitable for use by AI systems. Professional data organization and analysis must be done at a stage where errors in data are fixed, overlapping data is removed, and inconsistent data is erased in order to guarantee quality, dependability, correctness, and performance. Preprocessing the data takes about 80% of the process. AI development requires a substantial amount of high-quality data, and quality assurance is crucial.

IV. RESULTS AND DISCUSSION

With the use of machine learning techniques, such as the XGBoost Algorithm and KMEANS Clustering, this part demonstrates how Blockchain technology is used to digital forensics and guarantees the security of every transactional database.

4.1 Community Vs Security Level

Communications are significantly more expensive when signcryption is used. The transmission overhead is mostly determined by the size of the signed message.A traditional network requires only two bytes per user.The figure below, however, displays the security levels and communication overhead. As security standards raise, the overhead of communication increases.

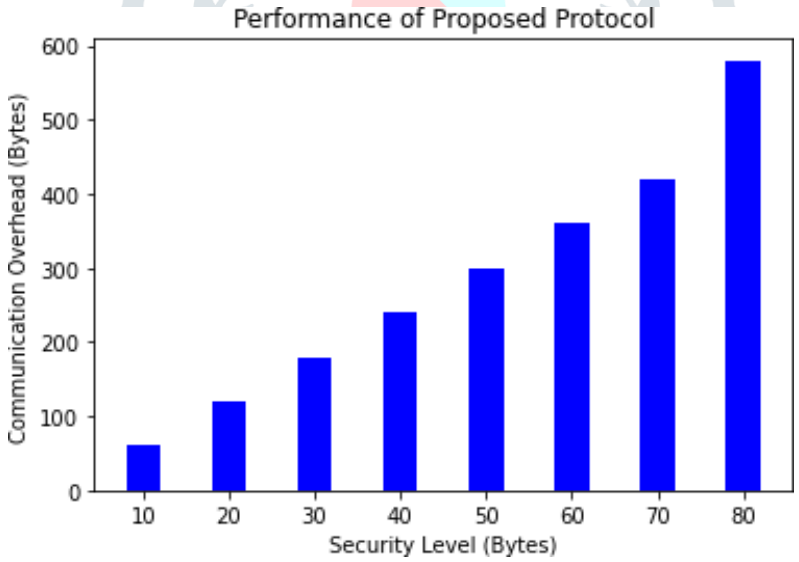


Figure 4.1 Performance of Proposed Protocol

4.2 Blockchaim Performance

The performance of the suggested blockchain-enabled platform was verified in this section by testing it in terms of block size, read throughput, transaction throughput, read latency, and transacton latency. In order to assess the performance of the blockchain network, four peer nodes and one ordered node were chosen as experimental parameters. Throughput was determined by varying the TPS send rate in the suggested blockchain-enabled platform.Two examples of how throughput can be separated are transactional throughput and read throughput. The transaction throughput was defined as the quantity of transactions completed within the designated time frame on the blockchain network. The blockchain network's read operations during the designated time window were measured using read-through. Transaction read-throughput was measured by varying the configuration of random machine usage and TPS transmit.

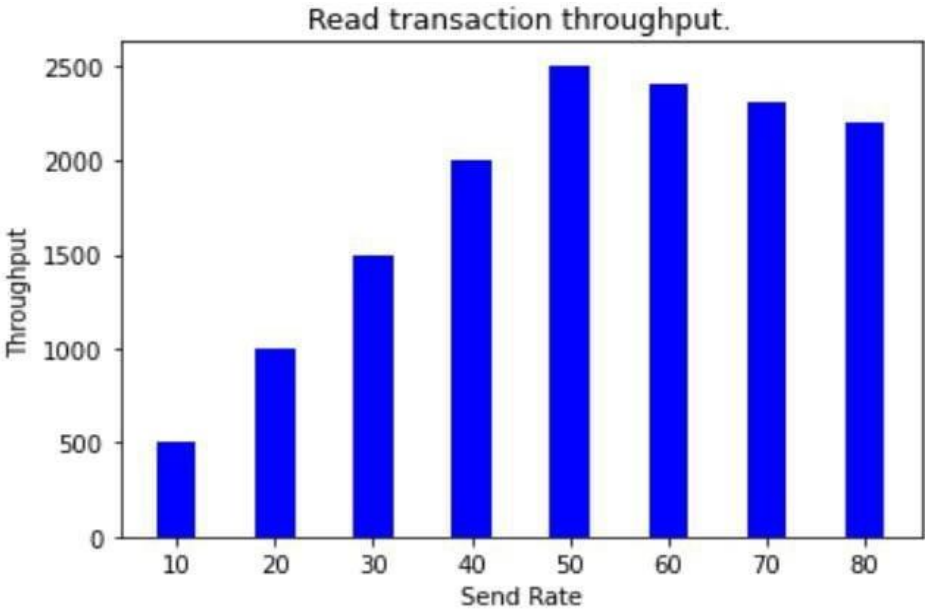


Figure 4.2 Read Transaction Throughout

Effectiveness of Machine Learning Models in Predicting SecurityWe had to imagine the group of damaging strikes in the first step.Two distinct attack types are clustered together in the figure below.

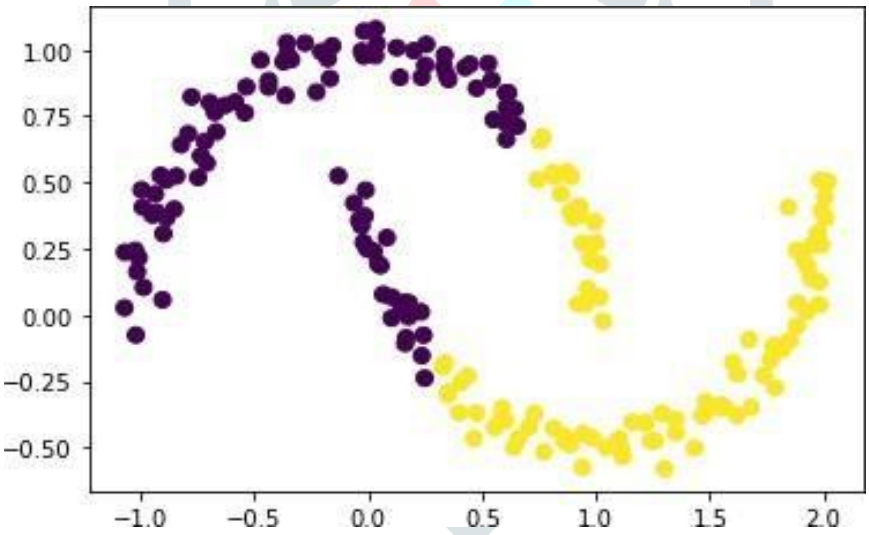
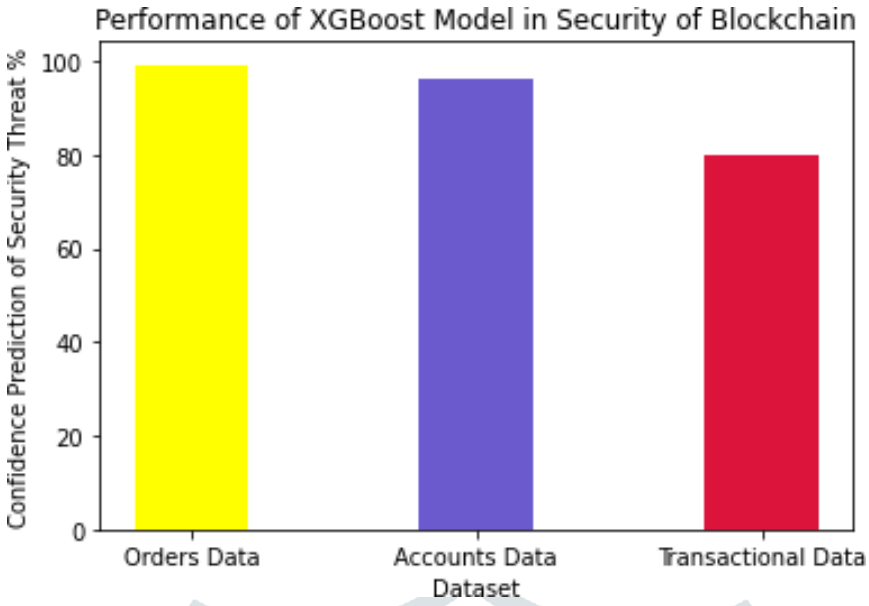


Figure 4.4 Visualization of clusters

To forecast early time attacks, we have employed the XGBoost algorithm. In order to guarantee system security and integrity, XGBoost's performance is displayed in the figure below for early attack prediction. For requests data, accounts data, and data related to transactions, respectively, XGBoost has shown accuracy of 99.8%, 95%, and 79%. However, KMEANS has demonstrated confidence clustering with accuracy levels of 58%, 59%, and 98% for each set of data, respectively.



4.5 Performance of XGBoost in securing the blockchain

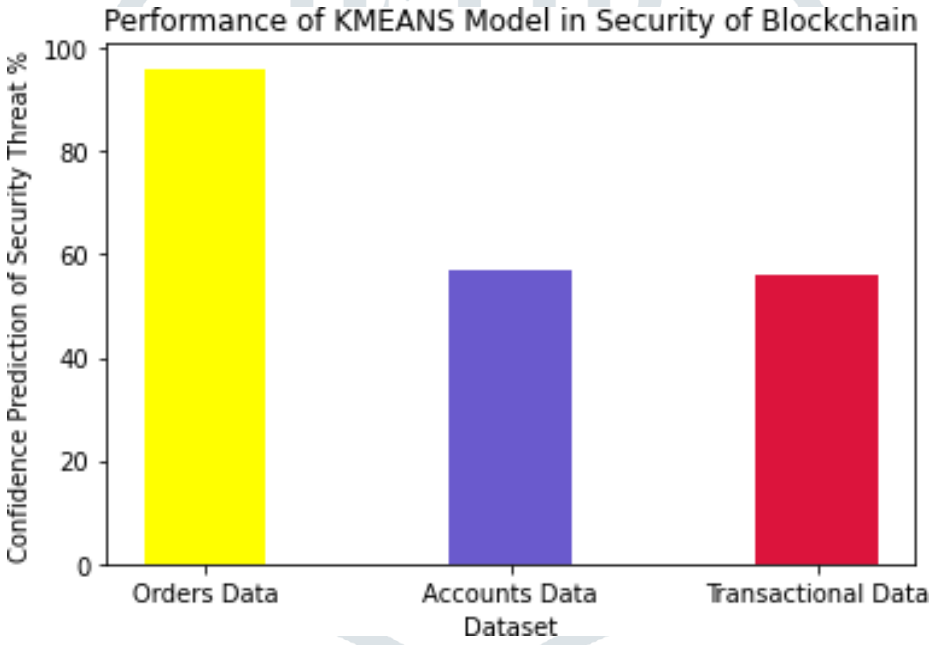


Figure 4.6 Performance of KMEANS model in securing the Blockchain

V .CHALLENGES AND FUTUREWORK

- Ensuring Data Privacy and Confidentiality: maintaining sensitive information's privacy and confidentiality while striking a balance with the requirement for access to evidence, particularly in IoT environments
- Resource and Scalability Limitations: Handling the processing and storage requirements brought on by growing data quantities, as well as the scalability of blockchain systems.
- Managing the diversity and complexity of evidence gathered from several IoT devices, which sometimes entail varying formats, protocols, and security levels, is known as the "complexity of multi-source evidence."

Cutting-edge technologies like blockchain, machine learning, and the Internet of Things are being integrated to determine the future of digital forensics. Blockchain offers forensic evidence storage that is impenetrable, guaranteeing data integrity and openness throughout investigations. Massive data generation from IoT devices will increase their use in gathering evidence in real time, requiring safe frameworks to manage confidentiality concerns. By improving the capacity to identify, anticipate, and stop cyberthreats, machine learning models will allow for early action and reduce harm.

Quantum computing and advanced artificial intelligence (AI) are predicted to further transform digital forensics by speeding up data analysis and increasing accuracy as cyber threats get more complex. Investigations will be quicker and more effective thanks to automation and cloud-based forensic technologies. Since every industry is becoming more and more dependent on technology,

digital forensics will continue to be essential in the years to come for preventing cybercrime, safeguarding data integrity, and assisting law enforcement.

VI . CONCLUSION

A subfield of forensics called "digital forensics" is focused on looking into digital evidence. Digital forensics is the study of finding, obtaining, processing, analyzing, and reporting information that has been stored in an electronic format. Digital forensics support is crucial for law enforcement investigations since almost every incident of illegal activity contains electronic evidence. It is possible to use digital forensics tools to find out what data has been stolen and how it has been copied or shared. It is plausible that certain hackers may intentionally destroy data in order to harm their targets. Important data may be corrupted in various situations without the user's knowledge due to harmful software or hacker activity.Threats to security and integrity are among the challenges faced by digital forensics. Because of worries about the security and integrity of the data gathered, IoT devices have the potential to gather digital forensic evidence in an IoT environment, which might be extremely dangerous for cybercrime organizations. Many studies have recently focused on the security and integrity of digital forensics based on the Internet of Things (IoT), but the biggest problem facing researchers is confidentiality. Despite technological advancements, recent studies and related investigations have shown that tampering and security-related problems in digital forensics still exist. As a result, a clever and efficient model is required that helps the system function by anticipating risks and ensuring security and integrity. Using the Hashing algorithm in conjunction with Blockchain technology, we are introducing an intelligent and efficient system. After crime evidence is gathered by Internet of Things devices, the information is kept on a blockchain. To anticipate irregularities in the evidence and transactions, we will use machine learning- enhanced models during that time. The capacity of the suggested approach to identify and predict attacks early enough makes it effective. The early time attack that was predicted using the XGBoost algorithm proved successful. To preserve system security and integrity, XGBoost performs well in terms of early attack detection, as shown in Figure 1. When it comes to early attack detection in order data, accounts data, and transactional data, XGBoost has shown accuracy of 99.8%, 95%, and 79%, respectively, for early attack prediction in the three distinct data types. According to the results, KMEANS has shown that confidence clustering is accurate with 98 percent, 58 percent, and 59 percent for each data set..

VII . REFERENCES

[1] P. Agbedanu and A. D. Jurcut “BLOFF: A Blockchain-Based Forensic Model in IoT,” Revolutionary Applications of Blockchain-Enabled Privacy and Access Control, pp. 59–73, 2021.

[2] Khan, M. Uddin, A. A. Shaikh, A. A. Laghari, and A. E. Rajput “MF-Ledger: Blockchain Hyperledger Sawtooth-Enabled Novel and Secure Multimedia Chain of Custody Forensic Investigation Architecture, IEEE Access, vol. 9, pp. 103637–103650, 2021, doi: 10.1109/ACCESS.2021.3099037.

[3] Chen, X. Huang, F. Liu, and H. Yin “A Location-Based Blockchain Evidence Preservation Wireless Communication Scheme for HuaTaiYiMei v Tongdao Technology Development Case,” Proceedings of the 2021 4th International Conference on Electronic Device and Mechanical Engineering (ICEDME 2021), pp. 38–41, 2021, doi: 10.1109/ICEDME52809.2021.00016.

[4] S. Patil, S. Kadam, and J. Katti “Security enhancement of forensic evidences using blockchain,” Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV 2021), pp. 263–268, 2021, doi: 10.1109/ICICV50876.2021.9388486.

[5] M. Stoyanova, Y. Nikoloudakis, S. Panagiotakis, E. Pallis, and E. K. Markakis “A Survey on the Internet of Things (IoT) Forensics: Challenges, Approaches, and Open Issues,” IEEE Communications Surveys & Tutorials, vol. 22, no. 2, pp. 1191–1221, 2020, doi: 10.1109/COMST.2019.2962586.

[6] Kim, M. Park, and D. H. Lee “AI-IDS: Application of Deep Learning to Real-Time Web Intrusion Detection,” IEEE Access, vol. 8, pp. 70245–70261, 2020, doi: 10.1109/ACCESS.2020.2986882.

[7] M. Stoyanova, S. Panagiotakis, E. Pallis, and E. K. Markakis “A Comprehensive Review of IoT Forensics: Emerging Challenges and Blockchain Solutions,” IoT-Based Solutions for Cloud Computing, 2020.

[8] S. Hahn “Evidence Management in Forensics with Blockchain Integration,” Forensic Dentistry, Second Edition, pp. 395–404, 2020.

[9] Hamid Lone and R. Naaz Mir “Forensic-Chain: Blockchain Digital Evidence Chain for IoT Systems,” Scientific Practices in Cybersecurity Journal, *2021*.

[10] M. Conti, S. K. Das, C. Lal, and S. Ruj**,** "A Survey on Security and Privacy Issues of Blockchain Technology," IEEE Communications Surveys & Tutorials, vol. 21, no. 2, pp. 116-135, 2019. This survey provides insights into the security and privacy challenges of blockchain technology, which are relevant for its application in digital forensics.

[11] M. S. Akhtar and T. Feng, "Using Blockchain to Ensure the Integrity of Digital Forensic Evidence in an IoT Environment," EAI Endorsed Transactions on Creative Technologies, vol. 9, no. 31, pp. e2, 2022. A blockchain and ML-based solution for IoT forensic data integrity.

[12] T. Nayerifard et al**.**, "Machine Learning in Digital Forensics: A Systematic Literature Review," arXiv:2306.04965, 2023. ML applications in digital forensics with research gaps.

[13] A. J. Akbarfam et al**.,** "ForensiBlock: A Provenance-Driven Blockchain Framework," arXiv:2308.03927, 2023. Blockchain framework for forensic data integrity and auditability.

[14] R. Kumar, W. Wang, J. Kumar, Z. Zakria, T. Yang, and W. Ali, "Collective Intelligence: Decentralized Learning for Android Malware Detection in IoT with Blockchain," arXiv preprint arXiv:2102.13376, 2021. This paper proposes a framework that combines decentralized learning and blockchain for malware detection in IoT devices.

[15] M. A. Ferrag and L. Maglaras**,** "DeepCoin: A Novel Deep Learning and Blockchain-Based Energy Exchange Framework for Smart Grids," *IEEE Transactions on Engineering Management*, vol. 67, no. 4, pp. 1285-1297, 2020. Although focused on smart grids, this paper presents a framework that combines deep learning and blockchain, which can be adapted for securing IoT forensic data.