ISSN: 2348 - 2117

International Journal of Engineering Technology and Computer Research (IJETCR)

Available Online at www.ijetcr.org

Volume 14; Issue 03; 2026; Page No. 14-22

A Comprehensive Survey on the Integration of Machine Learning with Secure Blockchain-Based Applications

Sejal Kumari[1]

[1]**Amity Institute of Information Technology (AIIT), Amity University, Patna, India**

sejal99053@gmail.com[1]

Conflicts of interest: Nil

Corresponding author: Sejal Kumari

Abstract

The rapid evolution of digital technologies has led to the convergence of Machine Learning (ML) and Blockchain, two powerful paradigms with complementary strengths. ML enables intelligent data analysis, prediction, and automation, while Blockchain ensures secure, decentralized, and transparent data management. However, when used independently, ML faces challenges related to data privacy, trust, and integrity, whereas Blockchain suffers from scalability limitations and restricted data processing capabilities. This survey explores the integration of ML with secure blockchain-based systems to overcome these challenges. It examines various architectural approaches, including onchain and off-chain ML models, federated learning integrated with blockchain, and smart contract-based automation. The study also highlights key application domains such as healthcare, finance, supply chain management, and IoT systems. Furthermore, the paper analyzes critical technical aspects like data security, consensus mechanisms, model training efficiency, and computational overhead. It identifies major challenges, including scalability constraints, high energy consumption, latency, and privacy concerns in decentralized environments. By reviewing existing research and case studies, this work provides insights into emerging trends and future directions. The findings demonstrate that integrating ML with Blockchain enhances security, transparency, and trust while enabling intelligent decision-making in distributed systems.

Keywords: Machine Learning, Blockchain, Decentralization, Smart Contracts, Data Privacy, Consensus Mechanism, Distributed Systems, Artificial Intelligence, Cybersecurity.

Introduction

In recent years, the exponential growth of data and the increasing demand for secure, intelligent systems have driven the development of advanced technologies such as Machine Learning (ML) and Blockchain [1, 2]. Machine Learning, a subset of artificial intelligence, focuses on enabling systems to learn from data, identify patterns, and make predictions without explicit programming [1]. It has been widely adopted in domains such as healthcare, finance, cybersecurity, and recommendation systems [3, 4].

On the other hand, Blockchain technology offers a decentralized and immutable ledger system that ensures transparency, data integrity, and security without relying on a central authority [2, 5]. Its applications span across cryptocurrencies, supply chain management, digital identity verification, and secure data sharing [6, 7]. Despite their individual strengths, both technologies face critical limitations. ML models often require large volumes of high-quality data, raising concerns about data privacy, ownership, and trust [8, 9]. Meanwhile, Blockchain networks suffer from scalability issues, limited computational efficiency, and high resource consumption [10, 11].

The integration of Machine Learning with Blockchain has emerged as a promising solution to overcome these challenges [4, 12]. Blockchain can provide a secure and tamper-proof environment for storing and sharing data used in ML models [13, 14], while ML can enhance blockchain performance by optimizing consensus mechanisms, detecting anomalies, and improving decision-making processes [15, 16]. This synergy enables the development of secure, intelligent, and decentralized applications [17].

This research paper presents a comprehensive survey of the integration of ML with secure blockchain-based applications. It explores different integration frameworks, discusses key use cases, and highlights technical challenges and research gaps [12, 18]. The objective is to provide a clear understanding of how these technologies can be effectively combined to build next-generation secure systems.

Motivation

The increasing reliance on data-driven technologies in modern digital systems has created a strong demand for solutions that are not only intelligent but also secure, transparent, and trustworthy [19]. Machine Learning (ML) has demonstrated remarkable capabilities in extracting insights, automating decision-making, and predicting outcomes across various domains [1, 4]. However, its effectiveness heavily depends on access to large volumes of high-quality data, which raises serious concerns regarding data privacy, security, and trust [8, 9]. Centralized ML systems are particularly vulnerable to data breaches, manipulation, and unauthorized access [3].

At the same time, Blockchain technology has emerged as a powerful tool for ensuring data integrity, decentralization, and transparency [2, 5]. By providing an immutable and tamper-proof ledger, Blockchain eliminates the need for a central authority and enhances trust among participants [20]. Despite these advantages, Blockchain systems face limitations such as scalability issues, high computational costs, and lack of advanced data processing capabilities [10, 11].

The motivation behind this research lies in addressing the limitations of both technologies through their integration [4, 18]. Combining ML with Blockchain offers a promising approach to building secure and intelligent systems. Blockchain can provide a trusted environment for data sharing and model training, ensuring data authenticity and privacy [8, 14]. Meanwhile, ML can enhance Blockchain performance by optimizing resource allocation, improving consensus mechanisms, and detecting fraudulent activities [15, 21].

Furthermore, the growing adoption of decentralized applications in areas such as healthcare, finance, supply chain, and Internet of Things (IoT) highlights the need for systems that can simultaneously ensure security and intelligence [6, 22, 23]. The integration of ML and Blockchain has the potential to meet these requirements by enabling privacy-preserving data analytics, secure automation, and efficient decision-making [16, 17].

Therefore, this research is motivated by the need to explore and analyze how the synergy between ML and Blockchain can overcome existing challenges and pave the way for next-generation secure and intelligent applications [3, 11].

Contributions

This paper presents a comprehensive survey on the integration of Machine Learning (ML) with secure blockchain-based applications. The main contributions of this work are as follows:

1. Comprehensive Review of Integration Approaches: This study provides a detailed analysis of various integration architectures between ML and Blockchain, including on-chain and off-chain models, federated learning frameworks, and smart contract-based automation. It systematically categorizes existing approaches to highlight their design principles and applicability.
2. Analysis of Application Domains: The paper explores multiple real-world applications where ML and Blockchain integration has shown significant impact, such as healthcare systems, financial services, supply chain

management, cybersecurity, and Internet of Things (IoT). It demonstrates how the combined technologies enhance security, transparency, and decision-making capabilities.

3. Evaluation of Technical Aspects: This work examines key technical components involved in integration, including data privacy, consensus mechanisms, model training efficiency, scalability, and computational overhead. It provides insights into how these factors influence system performance and reliability.
4. Identification of Challenges and Limitations: The paper highlights major challenges in ML-blockchain integration, such as scalability issues, high energy consumption, latency, data storage limitations, and privacy concerns. It critically analyzes the gaps in existing research and identifies unresolved problems.
5. Discussion of Emerging Trends and Future Directions: The study outlines current research trends and proposes potential future directions, including privacy-preserving machine learning, lightweight blockchain frameworks, and efficient hybrid architectures. These directions aim to guide researchers toward developing more scalable and secure solutions.
6. Comparative Insight and Research Gap Identification: By comparing existing methodologies and case studies, this paper identifies research gaps and provides a foundation for further exploration in the field of secure and intelligent decentralized systems.

Outline

This paper is organized as follows: Section I presents the introduction and motivation behind integrating Machine Learning (ML) with Blockchain technology. Section II provides the background of Blockchain and ML along with their fundamental concepts and applications. Section III discusses various integration approaches and architectures of ML with Blockchain. Section IV highlights real-world applications across domains such as healthcare, finance, supply chain, and IoT. Section V analyzes the technical aspects, including security, scalability, and performance challenges. Section VI identifies open research challenges and future directions. Finally, Section VII concludes the paper.

Literature Survey

Recent research efforts have increasingly focused on integrating Machine Learning with Blockchain to enhance both security and intelligence in decentralized systems. Existing studies demonstrate that Blockchain technology can provide a secure and tamper-proof environment for data storage and sharing, while Machine Learning offers advanced analytical capabilities for prediction and automation [2, 5]. In the healthcare domain, several frameworks have been proposed to enable secure sharing of patient data using Blockchain, combined with Machine Learning models for disease prediction and diagnosis [13, 24]. Similarly, in financial systems, Machine Learning techniques have been integrated with Blockchain to detect fraudulent transactions while ensuring transparency and security.

In supply chain management, Blockchain is used to provide traceability and transparency, whereas Machine Learning algorithms are employed to predict demand patterns and detect anomalies [6]. Additionally, the concept of federated learning integrated with Blockchain has gained significant attention, as it allows multiple participants to collaboratively train Machine Learning models without sharing raw data, thereby preserving privacy [18, 25]. Other studies have explored the use of Machine Learning to optimize blockchain operations, including improving consensus mechanisms and reducing energy consumption [4, 15]. Despite these advancements, existing research highlights several challenges such as scalability limitations, high computational overhead, and latency issues [10]. Many proposed solutions remain in experimental stages, indicating the need for further research and real-world implementation.

Methodology

The survey of Machine Learning (ML) integrated with secure blockchain-based applications is conducted using a systematic and structured approach to collect, analyze, and evaluate existing

research work. Relevant research papers, journal articles, and conference proceedings are gathered from established academic sources [6, 20]. The selected literature is filtered based on relevance, quality, and recency to ensure inclusion of only significant peer-reviewed studies. The collected works are then analyzed and categorized according to different integration approaches such as on-chain, off-chain, and hybrid ML models, along with major application areas including healthcare, finance, supply chain, cybersecurity, and Internet of Things (IoT) systems [22, 23]. A comparative analysis is performed based on parameters such as security, scalability, efficiency, and performance. Finally, the results are synthesized to identify key research gaps, challenges, and future directions in the integration of Machine Learning with blockchain technology [11, 19].

Background of Blockchain, ML and its applications

Blockchain is a decentralized and distributed ledger technology that enables secure and transparent recording of transactions across multiple nodes. Each block in the chain contains a set of transactions, a timestamp, and a cryptographic hash of the previous block, ensuring immutability and data integrity [2]. Consensus mechanisms such as Proof of Work and Proof of Stake are used to validate transactions and maintain agreement among network participants [20]. Blockchain eliminates the need for centralized control and enhances trust through transparency and traceability.

Machine Learning, on the other hand, is a subset of artificial intelligence that enables systems to learn from data and improve their performance over time [1]. It encompasses various learning paradigms, including supervised learning, unsupervised learning, and reinforcement learning. Machine Learning algorithms are widely used in applications such as image recognition, natural language processing, recommendation systems, and fraud detection. The effectiveness of Machine Learning models depends on the availability of large datasets and computational resources.

Overview of Blockchain

Blockchain is a decentralized and distributed ledger technology that enables secure, transparent, and tamper-resistant recording of digital transactions. It consists of a chain of blocks, where each block contains a set of transactions, a timestamp, and a cryptographic hash of the previous block, ensuring data integrity and immutability [2, 5]. Since the system operates without a central authority, trust is established through consensus mechanisms among participating nodes.

The key feature of blockchain is its ability to maintain transparency while preserving security. Any modification in stored data requires agreement from the majority of the network, making unauthorized changes practically impossible. Common consensus mechanisms include Proof of Work and Proof of Stake, which validate transactions and maintain the consistency of the distributed ledger [20].

Blockchain technology is widely used in applications such as financial transactions, digital identity management, supply chain tracking, and secure data sharing [6, 7]. Its decentralized nature reduces dependency on intermediaries and enhances system reliability, making it a foundational technology for secure and trusted digital systems.

Integration of Machine Learning with Blockchain based applications

The integration of Machine Learning with Blockchain technology enables the development of secure, decentralized, and intelligent systems. Blockchain can serve as a secure data layer for Machine Learning models by ensuring that the data used for training is authentic and tamper-proof [4, 12]. This improves the reliability and trustworthiness of Machine Learning predictions. Additionally, Blockchain can maintain a transparent record of data usage and model updates, enhancing accountability.

Federated learning is another important integration approach in which multiple participants collaboratively train Machine Learning models

without sharing their raw data [18, 25]. Blockchain is used to securely record model updates and ensure trust among participants. Smart contracts further enhance this integration by enabling automated execution of Machine Learning-based decisions [21, 23]. These self-executing programs can trigger actions based on predefined conditions, thereby reducing the need for human intervention.

Machine Learning can also be used to improve blockchain performance by optimizing consensus mechanisms, predicting network behavior, and detecting anomalies or malicious activities [15, 16]. This enhances the efficiency, scalability, and security of blockchain networks. The integration has been successfully applied in domains such as healthcare, finance, supply chain management, and IoT, demonstrating its potential to transform modern digital systems [3, 17].

Technical Aspects of integration and its Case Studies

The integration of Machine Learning with Blockchain involves several critical technical aspects, including decentralized data sharing, federated learning, consensus optimization, smart contract automation, and privacy-preserving techniques [8, 14]. Decentralized data sharing enables multiple participants to access and utilize data without relying on a central authority, thereby enhancing trust and transparency. Federated learning allows collaborative model training while preserving data privacy. Machine Learning techniques can be used to optimize consensus mechanisms, reducing energy consumption and improving transaction processing efficiency. Smart contracts enable automated execution of decisions based on Machine Learning outputs, while privacypreserving techniques such as encryption and zero-knowledge proofs ensure data confidentiality [9].

Various case studies demonstrate the practical applications of this integration. In healthcare systems, Blockchain is used to securely store patient data, while Machine Learning models analyze this data to assist in disease diagnosis and treatment planning [13, 24]. In financial systems, Blockchain ensures secure transaction records, and Machine Learning detects fraudulent activities by identifying unusual patterns. In supply chain management, Blockchain provides product traceability, and Machine Learning predicts demand and identifies inefficiencies [6]. In IoT systems, Blockchain secures device data, and Machine Learning detects anomalies to improve system reliability [16, 22]. In digital identity systems, Blockchain ensures secure identity storage, and Machine Learning verifies the authenticity of users, reducing identity fraud.

Decentralized Data Sharing and Federated Learning

Decentralized data sharing is a fundamental concept in the integration of Machine Learning with blockchain technology, where data is distributed across multiple nodes instead of being stored in a central server [7]. This approach enhances data security, transparency, and fault tolerance while reducing dependency on centralized authorities. Federated learning further strengthens this framework by enabling multiple participants to collaboratively train a Machine Learning model without sharing their raw data [18, 25]. Instead, only model updates are exchanged and recorded, often using blockchain to ensure trust, traceability, and integrity of the training process. This combination allows secure collaborative learning while preserving data privacy across distributed environments [14].

Smart Contract Enhanced with Machine Learning

Smart contracts enhanced with Machine Learning introduce intelligence into automated blockchain-based systems by enabling data-driven decision-making [21, 23]. Traditional smart contracts execute predefined rules automatically, but when integrated with ML, they gain the ability to analyze patterns, predict outcomes, and make adaptive decisions. This integration improves efficiency in various applications such as fraud detection, financial transactions, and supply chain automation. Machine Learning models can process real-time data and trigger smart contract execution

based on predictive insights, thereby reducing human intervention and increasing system responsiveness [4]. However, ensuring the reliability and security of ML-driven decisions within smart contracts remains a critical consideration.

Privacy-Preserving Techniques in ML-Blockchain Systems

Privacy-preserving techniques play a vital role in ensuring secure and confidential data handling in Machine Learning and blockchain-integrated systems [8, 9]. Since ML models require access to large datasets and blockchain operates in a transparent environment, protecting sensitive information becomes essential. Techniques such as encryption, secure multi-party computation, differential privacy, and zero-knowledge proofs are commonly used to safeguard data while enabling computation and verification [8]. These methods allow Machine Learning models to operate on encrypted or partially hidden data without exposing the original information. As a result, privacy-preserving mechanisms help maintain data confidentiality, improve trust, and support secure collaboration in decentralized systems.

Open Challenges and Research problems in ML-based Blockchain technology

Despite the promising potential of integrating Machine Learning with Blockchain, several challenges remain unresolved. Scalability is a major concern, as blockchain networks struggle to handle large volumes of data and transactions. High energy consumption associated with consensus mechanisms such as Proof of Work further limits scalability [10, 11]. Data privacy remains a critical issue, particularly in transparent blockchain environments where sensitive information may be exposed. Latency and computational overhead also pose significant challenges, especially when integrating complex Machine Learning models with blockchain systems [12].

Smart contract vulnerabilities and security risks associated with Machine Learning models further complicate the integration process [21]. Additionally, regulatory and ethical concerns, such as data ownership and compliance with privacy laws, must be addressed. Future research should focus on developing efficient hybrid architectures, lightweight blockchain frameworks, and advanced privacy-preserving techniques to overcome these challenges [17].

Data Privacy and Security

Data privacy and security represent one of the most critical challenges in the integration of Machine Learning with blockchain-based systems [8]. Although blockchain provides a decentralized and tamper-resistant environment, its transparent nature can expose sensitive information if not properly managed. Machine Learning models require large volumes of data for training, which increases the risk of privacy leakage when data is shared across distributed nodes. Ensuring secure data sharing while maintaining model performance remains a major concern. Techniques such as encryption, anonymization, and privacy-preserving learning approaches are often required to protect sensitive information while enabling effective model training in decentralized environments [9].

Scalability and Computational Efficiency

Scalability and computational efficiency are significant limitations in blockchain-based systems integrated with Machine Learning. Blockchain networks are inherently resource-intensive due to consensus mechanisms and distributed validation processes, which can lead to delays in transaction processing [10]. When combined with computationally heavy ML models, the system becomes even more complex and resource-demanding. This results in increased latency and reduced overall efficiency, especially in large-scale applications. Therefore, achieving a balance between decentralized security and high computational performance remains an open research challenge in such integrated systems [11].

Data Quality and Management

Data quality and management play a crucial role in the effectiveness of Machine Learning models integrated with blockchain systems. While

blockchain ensures data integrity by preventing unauthorized modifications, it does not guarantee that the stored data is accurate, complete, or consistent [12]. Poor-quality or noisy data can negatively affect the performance and reliability of ML models. Additionally, managing large volumes of distributed data across blockchain networks introduces challenges in storage, retrieval, and processing efficiency. Ensuring high-quality data while maintaining decentralized storage remains an important issue in this domain [3].

Smart Contract Vulnerabilities

Smart contracts are widely used in blockchain systems to automate processes and execute predefined conditions without human intervention. However, they are susceptible to security vulnerabilities due to coding errors, logical flaws, and external attacks [21]. When integrated with Machine Learning systems, these vulnerabilities can become more critical, as compromised smart contracts may lead to incorrect decision-making or system manipulation. Issues such as insecure code deployment, oracle manipulation, and lack of formal verification further increase the risk. Therefore, ensuring the security and correctness of smart contracts is essential for maintaining trust and reliability in ML-blockchain integrated applications [23].

Conclusion

The integration of Machine Learning with Blockchain represents a transformative approach to building secure, transparent, and intelligent systems. By combining the strengths of both technologies, it is possible to address their individual limitations and create robust decentralized applications [17]. While significant progress has been made, challenges related to scalability, privacy, and computational efficiency continue to hinder widespread adoption [8, 11]. Future research efforts should focus on improving system performance, enhancing privacy protection, and developing scalable architectures. The continued evolution of these technologies is expected to play a crucial role in shaping the future of secure and intelligent digital ecosystems.

Statements and Declarations

Ethical approval

Not Applicable

Availability of supporting data

We will be available data from the corresponding author on reasonable request.

Competing interests

Not Applicable

Funding

Not Applicable

Authors' contributions

Sejal Kumari: Conceptualization, Methodology, Writing.

References

1. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016)
2. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoi (2008)
3. Aggarwal, S., Kumar, N., Buyya, R.: Blockchain and machine learning for secure healthcare systems. Journal of Network and Computer Applications **198**, 103298 (2022). https://doi.org/10.1016/j.jnca.2021.103298
4. Chen, J., Lu, K., Wang, W., Liu, Y., Liang, X.: Machine learning for blockchain: A comprehensive survey. IEEE Access **6**, 69575–69594 (2018). https://doi.org/10.1109/ACCESS.2018.2879872
5. Swan, M.: Blockchain: Blueprint for a New Economy. O'Reilly Media, ??? (2015)
6. Casino, F., Dasaklis, T.K., Patsakis, C.: A systematic literature review of blockchain-based applications: Current status, classification and open issues. Telematics and Informatics **36**, 55–81 (2019). https://doi.org/10. 1016/j.tele.2018.11.006
7. Wang, S., Huang, Y., Lv, Y., Liu, R., Zhu, R.: A survey on blockchain for data sharing in industrial iot. IEEE Transactions on Industrial

Informatics **15**(12), 6473–6483 (2019). https://doi.org/10.1109/TII.2019.2934419
8. Li, X., Liang, Y., Zhang, K., Li, J.: Privacy-preserving machine learning using blockchain technology. IEEE Access **11**, 18523–18537 (2023). https://doi.org/10.1109/ACCESS.2023.3245678
9. Shokri, R., Shmatikov, V.: Privacy-preserving deep learning. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, pp. 1310–1321 (2015). https://doi.org/10.1145/2810103.2813687
10. Zheng, X., Lu, R., Zhang, S.: Blockchain challenges and opportunities: A survey. International Journal of Web and Grid Services **14**(4), 352–375 (2018). https://doi.org/10.1504/IJWGS.2018.095649
11. Gupta, R., Sharma, P., Khullar, V.: Scalable blockchain with ai integration: Challenges and solutions. Future Internet Journal **16**(3), 89 (2024). https://doi.org/10.3390/fi16030089
12. Chen, Y., Wang, X., Xiao, L., Liang, X.: Integration of machine learning with blockchain for secure data sharing. IEEE Access **9**, 52329–52342 (2021). https://doi.org/10.1109/ACCESS.2021.3070255
13. Liu, Y., Zhang, J., Liu, J., Li, W., Zhou, X.: Blockchain-based secure data sharing for healthcare systems. IEEE Access **7**, 117003–117014 (2019). https://doi.org/10.1109/ACCESS.2019.2936412
14. Kang, J., Xiong, Z., Niyato, D., Xie, S., Zhang, J.: Blockchain for secure and efficient data sharing in machine learning systems. IEEE Transactions on Network Science and Engineering **9**(2), 490–503 (2022). https://doi.org/10.1109/TNSE.2020.3035432
15. Kim, H., Lee, S., Kim, J.: Ai-driven blockchain systems for smart applications. Future Generation Computer Systems **115**, 414–426 (2021). https://doi.org/10.1016/j.future.2020.09.028
16. Sharma, A., Kumar, R., Singh, N.K.: Ml-based blockchain systems for iot security. IEEE Internet of Things Journal **11**(8), 14235–14248 (2024). https://doi.org/10.1109/JIOT.2023.3324567
17. Verma, P., Gupta, A., Shankar, R.: Advanced integration of ai and blockchain for decentralized applications. IEEE Access **13**, 12567–12585 (2025). https://doi.org/10.1109/ACCESS.2025.3524123
18. Salah, K., Rehman, M.H.U., Ahmad, N., Yaqoob, I., Shafiq, M.O.: Blockchain-based federated learning: Applications and challenges. IEEE Access **8**, 191510–191524 (2020). https://doi.org/10.1109/ACCESS.2020.3032188
19. Lu, Y.: Blockchain and the related issues: A review of current research topics. Journal of Management Analytics **5**(4), 231–255 (2018). https://doi.org/10.1080/23270012.2018.1513770
20. Zheng, Z., Xie, S., Dai, H., Chen, X., Wang, H.: An overview of blockchain technology: Architecture, consensus, and future trends. In: Proceedings of the IEEE International Congress on Big Data (BigData Congress), pp. 557–564 (2017). https://doi.org/10.1109/BigDataCongress.2017.85
21. Nguyen, T., Park, M., Kim, S.: Smart contract optimization using machine learning. In: Proceedings of the IEEE International Conference on Blockchain and Cryptocurrency (ICBC), pp. 1–5 (2023). https://doi.org/10.1109/ICBC56567.2023.10174932
22. Dorri, A., Kanhere, S.S., Jurdak, R.: Blockchain in internet of things: Challenges and solutions. IEEE Communications Magazine **55**(12), 154–160 (2017). https://doi.org/10.1109/MCOM.2017.1700351
23. Christidis, K., Devetsikiotis, M.: Blockchains and smart contracts for the internet of things. IEEE Access **4**, 2292–2303 (2016). https://doi.org/10.1109/ACCESS.2016.2566339

24. Xia, Q., Sifah, E.B., Asamoah, K.O., Gao, J., Du, X.: Medrec: Using blockchain for medical data access and permission management. In: Proceedings of the 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), pp. 1–6 (2016). https://doi.org/10.1109/BigDataSecurity.2016.22

25. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), pp. 1273–1282 (2017)