Journal of Big Data

SURVEY

Open Access

Blockchain meets machine learning: a survey

Safak Kayikci[1*] and Taghi M. Khoshgoftaar[1]

*Correspondence:
skayikci@fau.edu

[1] Department of Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431, USA

Abstract

Blockchain and machine learning are two rapidly growing technologies that are increasingly being used in various industries. Blockchain technology provides a secure and transparent method for recording transactions, while machine learning enables data-driven decision-making by analyzing large amounts of data. In recent years, researchers and practitioners have been exploring the potential benefits of combining these two technologies. In this study, we cover the fundamentals of blockchain and machine learning and then discuss their integrated use in finance, medicine, supply chain, and security, including a literature review and their contribution to the field such as increased security, privacy, and decentralization. Blockchain technology enables secure and transparent decentralized record-keeping, while machine learning algorithms can analyze vast amounts of data to derive valuable insights. Together, they have the potential to revolutionize industries by enhancing efficiency through automated and trustworthy processes, enabling data-driven decision-making, and strengthening security measures by reducing vulnerabilities and ensuring the integrity of information. However, there are still some important challenges to be handled prior to the common use of blockchain and machine learning such as security issues, strategic planning, information processing, and scalable workflows. Nevertheless, until the difficulties that have been identified are resolved, their full potential will not be achieved.

Keywords: Blockchain, Machine learning, Internet of things, Supply chain, Medicine, Finance, Security

Introduction

Machine learning (ML) imitates the learning mechanism of the human brain and the rapid development in recent years has led to the introduction of many applications that make human life easier. It is one of the most important sub-branches of artificial intelligence and it involves the development of algorithms and models that can analyze and learn from data, making predictions or decisions without being explicitly programmed to do so. This allows for self-improvement and adaptability in various applications, such as image and speech recognition, recommendation systems, and text processing.

The blockchain concept was presented by Satoshi Nakamoto in the year 2008 by using consensus protocol [1]. The blockchain functions as a secure, decentralized digital ledger that stores information about transactions, including time, date, price, and participants

[2]. Data integrity, security, trustworthiness, and decentralization are blockchain's four main qualities, which are intended to increase trust and safety.

The combination of blockchain and machine learning holds the potential for creating a secure, decentralized, smart, and effective network transaction and administration system. Both academia and industry have shown great interest in the benefits that this combination brings, such as improved information and model contribution, enhanced security and confidentiality, and reliable decision-making in machine learning. ML is assumed to have a substantial effect on the advancement of blockchain in communication and networking systems by increasing efficiency, scalability, and security. This combination enables the secure and transparent storage of large datasets, allowing machine learning models to access and train on reliable data. Also, it enhances data privacy and control by providing decentralized ownership and permissioned access. Lastly, it enables the development of decentralized machine learning models, allowing participants to contribute their computational resources while maintaining data privacy, leading to more collaborative and efficient machine learning ecosystems.

The purpose of this study is to highlight the areas where blockchain and machine learning are used together. We completed our literature search on Feb 15th, 2023 with the referenced papers. The rest of the paper is organized as follows. "Blockchain technology" provides an overview of blockchain with explanations of Ethereum, smart contracts and consensus algorithms. "Machine learning" describes machine learning. "Literature review" provides the literature overview on the integration of these two technologies, along with their contributions, gaps, and advantages in the fields of finance, medicine, supply chain, and security. "Real world examples" gives some real-world examples for blockchain and machine learning integration. Finally, "Conclusion" concludes the article with key highlights, comments and future trends.

Blockchain technology

Notebooks have been an integral part of business processes since ancient times. While the concept of a notebook has not changed over time, the technology supporting it has evolved from paper records to digital archives. Computer scientists seek solutions to the issues of how best to process, store and transmit data and they come up with a digital ledger which is a tool used to record transactions. The newest technology obtained at the end of these searches is blockchain technology. In this sense, a blockchain ledger is a decentralized and immutable digital ledger that records transactions in a chronological chain of blocks. It ensures transparency, security, and integrity by distributing the ledger across multiple participants, making it resistant to tampering and providing a trusted source of truth [3]. One of a blockchain's primary goals is to past transactions. The idea of keeping track of transactions on a ledger is fundamental to the blockchain. A general definition of blockchain is a type of digital ledger that allows you to store any data you want and access it later using the hash value you obtain. Blockchain and databases can look very similar. Both technologies actually have the idea of saving data, but while it is possible to add, delete, and change data in a database, it is only possible to add new information to a blockchain.

Blockchain was first proposed by Haber and Stornetta in the early 1990s as a mechanism to digitally date, hash, and link data in a chronological manner [4]. Blockchains have undergone enhancements, modifications, and adjustments as the technology has grown and spread. Now it is viewed as a suitable solution for the identification, registration, distribution, transfer, and tracking of any digital asset since it combines the concepts of “database” and “network” in computer systems. Blockchain safeguards the integrity of data by eliminating centralized control through its decentralized nature. By distributing the ledger across multiple participants and utilizing consensus mechanisms, blockchain ensures that any changes to the data require majority agreement, making it extremely difficult for any single entity to manipulate or tamper with the information stored on the blockchain. It also features a sizable distributed network of independent users. Full nodes are the collective name for all of the network’s computers. The network’s full nodes verify all transactions before they are added to the ledger and recorded. One of the most significant and potent features of blockchains is the elimination of the need for a central authority in the database structure [5].

Most networks today work with decentralized architecture. In a decentralized system, all computer nodes form the larger computer network. Decentralized systems have many advantages. They can share files, peripherals, and other tools. They are more reliable than a centralized system as they are not prone to a single point of failure. When more resources are required, decentralized systems can address this issue by expanding the network with new machines. By coordinating point-to-point transactions, blockchain technology offers solutions to the high cost, inefficiencies, and insecure data storage issues that are present in centralized organizations. On the other hand, decentralization has several potential disadvantages too. When decision-making is distributed, it can be difficult to ensure that all parties are working towards the same goals. This can lead to coordination problems such as duplication of effort or conflicting priorities, higher costs and longer lead times They may be less efficient than centralized ones, as decision-making can be slower and more cumbersome. Also, decentralized systems may not be equitable, as power and decision-making may be concentrated in the hands of a few influential actors. This can lead to disparities in access to resources and opportunities [6].

Peer-to-peer networks (P2P) are the foundation of the blockchain approach. Peer-to-peer systems are distributed software platforms made up of nodes that allow users to directly use one another’s computational resources, such as processing speed, storage space, or information delivery [7]. Users make their computers into nodes of the peer-to-peer network with equal rights and duties when they connect them to the network. All nodes in the system have the same functional capabilities and obligations notwithstanding users’ varying resource contributions. Computers belonging to all users are therefore resource suppliers as well as consumers. Each block in a blockchain created by the blockchain has a series of transactions that take place at a given moment. With a P2P design, this data structure can be expanded and shared by numerous clients. Every network node in a peer-to-peer network is linked to every other network node. These are the nodes that assist with block storage and mine blocks in accordance with the parameters outlined in the blockchain algorithm. Distributed ledger technology, or DLT, is another name for this arrangement, in which blocks and ledgers are dispersed among different network nodes [8]. Peer-to-peer networking enables us to effectively overcome the scalability and single source failure problems that plague client–server design.

Records are kept in a separate journal called a ledger and are frequently referred to as transactions. In order for all transactions to be replicated over a peer-to-peer blockchain network created by a distributed database, the ledger keeps a consistent copy of each network node that is active. A varied and decentralized network of data records called blockchain is evolving. All transactions on the blockchain are validated by other blocks in the network and recorded in each block. Interactions happen peer-to-peer without authorization or intermediary control. Figure 1 shows the P2P blockchain network, but for simplicity the connections between all nodes are not shown.

The fundamental units of the blockchain are blocks. A header and a body make up a block. It includes the block's metadata, including the Title, Version, Prior block (which points to the Previous block), Timestamp, Nonce, Bits, and Merkle-root shown in Fig. 2. The block body is made up of transactions and a transaction counter [9].

Data structures called blocks are systematically added one block at a time to the Blockchain. Blockchain can be characterized as a growing collection of records in which cryptographically secured structures known as blocks are in the form of a linked list.

The Genesis is the first link in the chain (formation block). Only the Genesis block does not make reference to the previous block digest. Depending on the use case for the blockchain, several block designs are used. The blocks on the Ethereum blockchain are distinct from those on the Bitcoin blockchain [10]. The sizes and types of information that blocks can hold will vary depending on how the system is constructed. The cryptographic hash function is used to link the blocks together, creating an impossibly complex mathematical connection as shown in Fig. 3. Each block contains the values shown below:

1. Data: The stored character string.
2. Nonce: A distinct number in the mining.
3. Previous hash: The hash of a block that precedes the current block. This parameter creates the cryptographic link to the following block.
4. Hash: Fingerprint of some block-stored data.

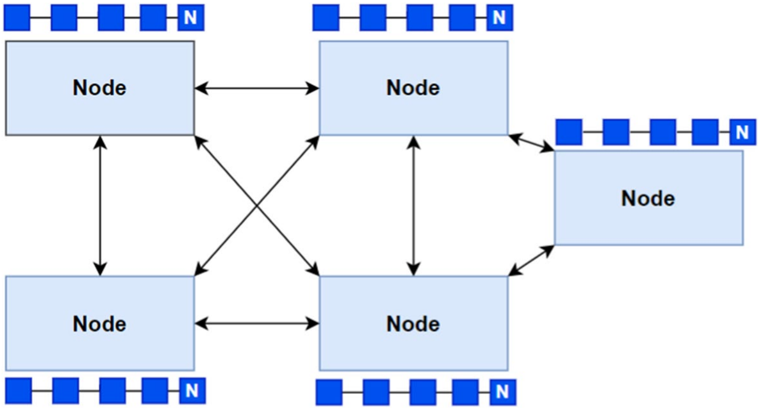

Fig. 1 P2P blockchain network

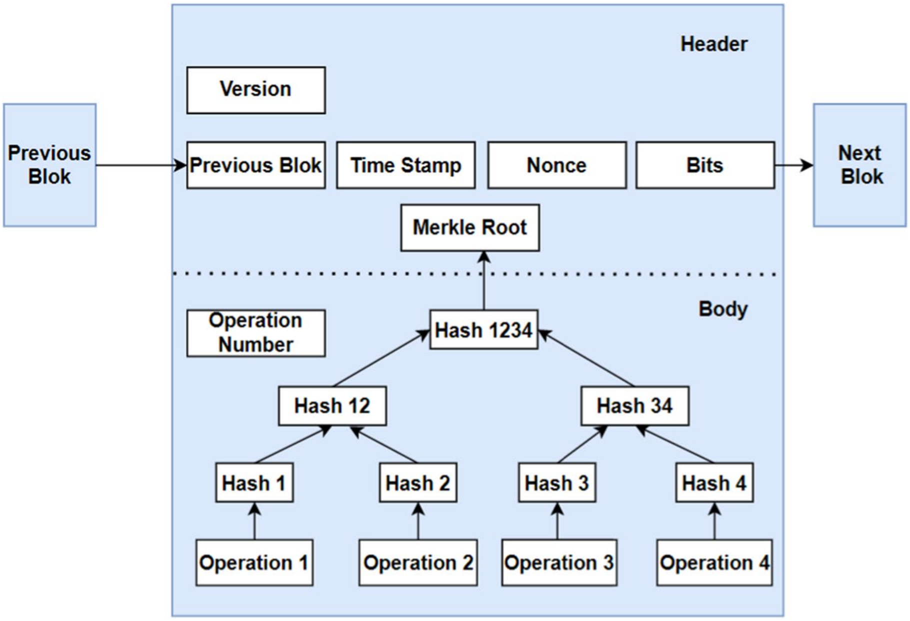

Fig. 2 Internal structure of the block

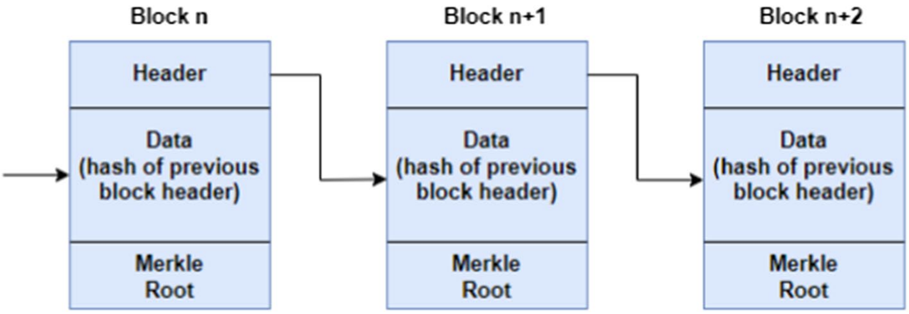

Fig. 3 Representation of a blockchain

A block's hash value is determined by using additional information, a nonce, and the previous hash field. Since every block in a blockchain contains a cryptographic digest of the block before it, updating any block in the system necessitates replacing all succeeding blocks as well. Because of this, information saved on a blockchain is typically regarded as secure and unchangeable. Every node in a blockchain network keeps a copy of the whole blockchain. The amount of storage space needed to hold the complete blockchain grows as more blocks become available.

Blockchains organize branching nodes using a mathematical structure. The structure formed as a result of branching is called a "mixed tree" or a "Merkle tree". Merkle tree refers to the framework of transactions in the particular block for the corresponding blockchain. The Merkle tree, a data structure used to hash and check the consistency of a dataset including cryptographic hashes, is used to hold the summary of all transactions in a block. It also summarizes all transactions included in a block, making it possible to quickly determine whether a transaction is a part of a block.

Merkle trees provide efficient verification of the integrity of big data structures as shown in Fig. 4.

The Merkel root field and Merkle tree function are used by the Blockchain to compute the hash, which displays the hash of the most recent block. The common blockchain transaction process is as follows [11]:

- Blockchain creates a digital signature using public and private keys to assure security.
- With these keys, authentication is carried out and authorization is given.
- Enables members to do network mathematical validations and come to an agreement on any specific value.
- Using the private key, the sender broadcasts the transaction via the network. The recipient's public key to carry out a transaction is included in the block along with a timestamp and digital signature.
- The verification of the procedure starts once the material is published.
- To process a transaction, nodes in the network endeavor to unravel its riddle. Nodes need computing resources to finish the problem.
- Nodes will be rewarded with bitcoins when the problem is solved. Proof-of-work issues are these kinds of projects.
- The timestamp is appended to the existing block when all participating nodes in consensus agree on a solution. The block might include anything, including cash or data.
- Existing nodes in the network are updated whenever a new block is added to the chain. The time it takes to update existing nodes in the network when a new block is added to the blockchain can vary depending on several factors like the specific blockchain protocol being used, the consensus mechanism employed, the network's speed and congestion, and the computational power of the nodes.

A blockchain uses a cryptographic hash method to connect two adjacent blocks. The Header Hash, which is processed by the cryptographic hash function that safeguards the transactions in each block as well as the linked blocks, acts as a pointer between the blocks. The block additionally retains the hash of the previous block in addition to the hash of the current block. The blockchain is more secure because of these characteristics of the block on the chain. By grouping fresh transactions into blocks and cryptographically connecting the blocks in a specified order, the blockchain is updated. After

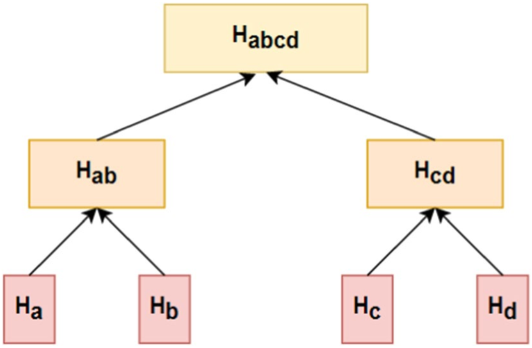

Fig. 4 Calculation of blocks using Merkle tree

all participating nodes have been verified, each block is connected to the one before it. Old blocks are harder to replace when more new ones are installed.

Ethereum

Ethereum is a decentralized, open-source blockchain platform designed for creating and executing smart contracts and decentralized applications (DApps). Smart contracts are self-executing contracts with the terms and conditions directly written into code. These contracts run on the Ethereum network and automatically execute when predetermined conditions are met, providing a trustless and tamper-proof way to conduct transactions and automate various processes. The Ethereum Virtual Machine (EVM) is a crucial component of the Ethereum network [12]. It's a computational environment that allows the execution of smart contracts. It provides a secure and isolated environment where code can run without interference from other contracts or the underlying network. This isolation ensures that computations and state changes are consistent across all nodes in the network. The EVM operates as a decentralized, distributed computing system. When a transaction containing a smart contract is broadcasted to the Ethereum network, all nodes in the network execute the contract independently, ensuring consensus on the outcome. The results of these computations, such as changes in data or token transfers, are recorded on the blockchain.

Currently, Ethereum's main challenge is scalability. The network relies on a consensus mechanism called Proof of Stake (PoS) and processes transactions through a single chain, which limits the number of transactions it can handle at a given time (around 15–45 transactions per second). Sharding is a proposed upgrade intended to address this limitation. It is a technique that involves partitioning the Ethereum network into smaller groups called shards. Each shard operates as its own blockchain with its set of validators, transaction history, and smart contracts. These shards can process transactions and execute smart contracts in parallel, greatly increasing the network's overall throughput. In a sharded Ethereum, a transaction is processed by only a subset of the shards, rather than every node on the network. This means that the overall capacity for processing transactions increases linearly with the number of shards, potentially allowing the network to handle thousands to tens of thousands of transactions per second. However, ensuring consistency and security across shards requires a robust cross-shard communication mechanism, and it introduces new challenges in terms of managing state across multiple chains. Research and development are ongoing to address these complexities and make sharding a reality on the Ethereum network. Once successfully implemented, sharding could significantly enhance Ethereum's capacity, making it more suitable for applications with high transaction volumes, such as financial systems, supply chain management, and decentralized exchanges.

Smart contracts

Smart contracts are a key feature of many blockchain platforms. They are self-executing contracts with the terms of the agreement between buyer and seller being directly written into lines of code [13]. These contracts are stored on a blockchain, and are automatically executed when predetermined conditions are met. They are self-executing electronic contracts that define the conditions of a business partnership's legal and commercial agreements. Such contracts allow transactions and agreements to be conducted without a central authority. Blockchain technology acts as an intermediary in smart contracts to implement all business agreements, protocols, and programmed information exchanges. In a blockchain, smart contracts are used to automate various functions, such as the transfer of digital assets or the verification of identity. Because smart contracts are executed on a blockchain, they are transparent, tamper-proof, and can be verified by all parties involved. In a broad sense, smart contracts can be thought of as a collection of computer-processable functions [14].

With the emergence of approaches for modeling and analysis of business processes in the 1990s, and later the development of workflow management approaches for machine-assisted execution of these models, the foundations for current approaches to smart contracts were laid. The term "smart contracts" was first conceptualized by Szabo in 1996 [15]. In his article "Forming and Securing Relationships in Public Networks", he likens smart contracts to vending machines. The vending machines perform automatic transactions according to the number of coins thrown into them and the product selection made through a simple computer vending software. Based on this approach, blockchain system designers have designed smart contracts. Smart Contracts are electronic transactional processes that automatically carry out the terms of a contract when a specific set of conditions arises. Users can launch a smart contract by sending a transaction to the contract address. When certain events take place, they transfer cryptocurrency automatically.

Smart contract code, in contrast to standard computer code, is deposited in a blockchain network, run, and its outcomes are validated by nodes taking part in the blockchain network. By posting a transaction on the blockchain, any user can form a contract. When a message or other contract is received from a user, the smart contract's program code will be performed and cannot be modified after it has been generated. Smart contracts execute a specific piece of their code when triggered by a user via a custom message or an action from another smart contract. These autonomous systems are run on a custom built EVM. A smart contract can be in the form of a crypto asset that can be sent and exchanged. This crypto-asset may have a term built into its software on a given date to send instructions to create another crypto-asset and send it to the wallet it is currently in. While setting up a smart contract, the owner generates the contract and posts it on the blockchain. Companies that agree to the terms of the smart contract engage with it. After the smart contract's conditions are posted on the blockchain, the owner cannot change them. The consensus algorithm of the underlying blockchain system and the code put within serve to enforce the terms of smart contracts. The framework of the smart contract is designed to allow negotiation or comparison of performance, cross-checking or validation. The purpose of using smart contracts is that they allow reliable transactions to be carried out without the involvement of third parties or banks. The smart contract works in three steps as shown in Fig. 5:

1. Smart contracts are written in the form of code. The written code is sent to the blockchain.
2. When an event presented in the contract is triggered, the code causes the event to occur.

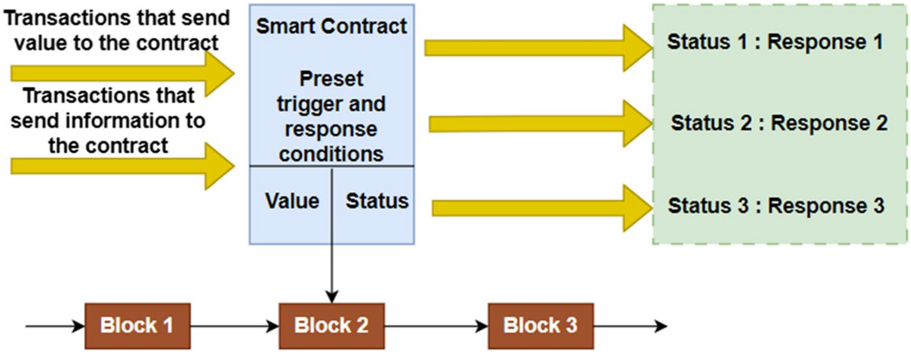
Transactions that send value to the contract
Transactions that send information to the contract
Smart Contract
Preset trigger and response conditions
Value
Status
Status 1 : Response 1
Status 2 : Response 2
Status 3 : Response 3
Block 1
Block 2
Block 3

Fig. 5 Working principle of smart contract

3. Regulators have the power to scrutinize contract activity on the blockchain.

In Bitcoin, the terms and conditions of a successful transaction, double spending verification, and the creation and consumption of new coins are expressed using a straightforward stack language. As a result, it is crucial that the smart contracts that are deployed are precise and behave decisively. With Ethereum, smart contracts are typically short programs that link a transaction to a message. A smart contract is a piece of source code used to represent a computer program. It has the ability to automatically apply a separate agreement's terms that are written in everyday language. Typically, proprietary languages are used to create smart contracts, which are subsequently turned into bytecode. It is contained in independent, self-contained virtual machines or containers that can be installed on any blockchain node.

The developer can choose from a variety of programming languages to write smart contracts and other programs, but Solidity, which is comparable to the C and JavaScript computer languages, is the most often used language for Ethereum. Other programming languages for Ethereum include Serpent, Vyper, LLL, Mutan, and Julia. The smart contract has the ability to send messages to other users or contracts, read and write stored files.

Users can conduct targeted searches or create smart contracts to trigger any action. Every object in the network executes the code when a smart contract method is invoked, and the consensus algorithm compares the results to those of other nodes. Next, as a validation procedure, the smart function call (arguments) is added to the blockchain. A key advantage of code-only smart contracts is that they can be used to automatically process transactions without the need for human intervention during the transaction. Existing procedures can be automated using blockchain and smart contracts, which boosts the blockchain's effectiveness and lowers costs. The processes present in many contracts can be handled automatically by smart contracts, which can also impose financial penalties if certain objective conditions are not met and guarantee the transfer of monies upon specific triggering events. The drawback of the smart contract is that because the hash is used for indexing, it is impossible to alter the code once it has been uploaded to the blockchain. As a result, before deploying to the mainnet, the smart contract code should be tested on the testnet.

Below is a simple example of a smart contract for renting a vacation property:

- Smart Contract Creation: The property owner creates a smart contract using a blockchain platform. The contract includes terms such as rental duration, rental price, and conditions for refundable deposit.
- Contract Deployment: The smart contract is deployed to the blockchain network, becoming a part of the decentralized ledger accessible by all network participants.
- Interaction with the Contract: A potential renter interacts with the smart contract by initiating a rental request and providing necessary information such as desired rental dates and deposit amount.
- Validation and Agreement: The smart contract automatically validates the request against predefined conditions, such as availability of the property during requested dates and sufficiency of the deposit. If the conditions are met, the contract moves to the next step.
- Execution and Payment: The renter submits the rental payment (in cryptocurrency) specified by the contract. The contract verifies the payment and executes the rental agreement, marking the property as reserved for the renter during the specified dates.
- Rental Period: The rental period arrives, and the renter occupies the property as agreed.
- Contract Completion: After the rental period ends, the smart contract automatically checks for any damages and initiates the refund of the deposit to the renter, deducting any agreed-upon fees or charges for damages if applicable.

Throughout this process, the smart contract enforces the agreed-upon terms, eliminates the need for intermediaries such as a rental agency, and ensures transparency and security in the rental transaction.

Consensus algorithms

Blockchain uses consensus techniques to implement rule enforcement. Rules must be developed to provide security and maintain the integrity of the shared ledger when working with untrusted peers to stop double spending and potential hacker assaults. Consensus mechanisms are the names given to these laws and agreements [16]. The network needs to reach consensus via an algorithm in order to update the blockchain. When a consensus is reached, several servers in the distributed network vouch for the system's accuracy at the moment. The mechanisms used by each blockchain to generate agreements on new entries are unique. Consensus-building models come in a wide variety. This is due to the fact that each blockchain has a distinct type of data entry and a varied projected threat level. Based on the anticipated level of threat, the blockchain chooses the consensus algorithm it will employ. For instance, because they anticipate a very high level of threat, Bitcoin and Ethereum adopt a powerful consensus method called proof-of-work (PoW). Moreover, a simpler and quicker consensus method is used by blockchains designed to store financial transactions between well-known parties.

To decide how to validate existing transactions and add new transactions to the blockchain, a consensus mechanism is also utilized. Based on the needs of the use cases, developers and software architects choose which sort of blockchain system (private or public) and consensus method to utilize. Blockchain networks often use the Proof of Work (PoW) and Proof of Stake consensus techniques (PoS) [17].

The original and most extensively used consensus mechanism is proof of work (PoW) shown in Fig. 6. Miners are faced with a mathematical challenge by PoW. The prize for solving the issue is a cryptocurrency given to the miner. The name of the award stems from the fact that it serves as evidence of the "job" completed. Algorithms that power the distributed system in the PoW consensus pay miners for resolving mathematical puzzles. All mining clients on the network receive notifications of new network transactions from the software wallets that carry them out. For the purpose of defending the blockchain against hostile and dishonest network nodes, PoW employs the idea of efficient resource usage by requiring participants in a blockchain network to demonstrate computational work to validate and add new blocks to the blockchain. This computational work serves as a means of securing the network and deterring malicious activities, while also ensuring that participants invest significant computational resources, making it economically impractical to launch attacks or manipulate the blockchain easily.

Proof of Stake (PoS) is a consensus algorithm commonly used by cryptocurrencies to validate blocks. PoW was created as a way to avoid economic and environmental issues such as heavy energy consumption and the cost of mining. The proof-of-stake was created in 2011 and was first implemented by Peercoin in 2012. The Ethereum cryptocurrency had to switch from PoW to PoS in 2018 to keep the number of miners on the network from decreasing with this increase in difficulty, which reduced the difficulty

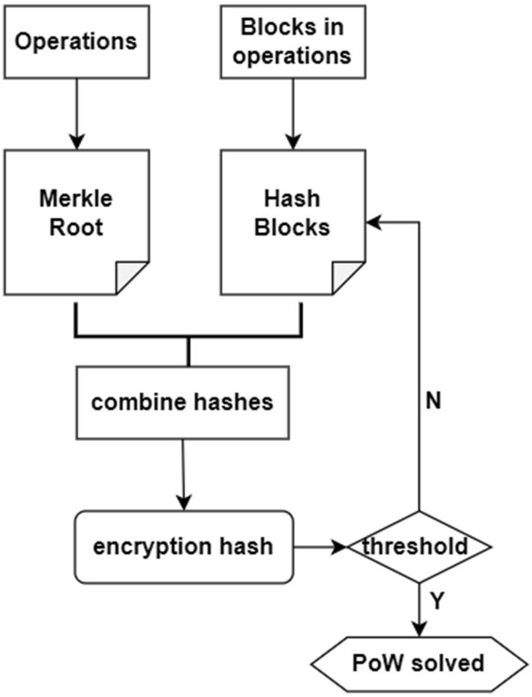

Fig. 6 Schematic view of proof of work

and increased profits and scalability for miner. With POS systems, the decision to create a new block is made based on the network participant's stake or level of commitment. Instead of determining the owner of the business according to the amount of energy consumption that determines the proof of work in PoW, PoS determines the owner of the business according to the size of the share owned by the peers. As a result, there is a distributed consensus that uses less money and energy. Some issues with this consensus technique include:

1. Because large stake holders are more likely to have their blocks added to the blockchain, the consensus is for the block inclusion procedure to be centralized in proportion to the share distribution.
2. Block mining forks, allowing miners to simultaneously mine on all branches. Double spend attacks are therefore simpler to execute in this situation.
3. By accumulating coins for a longer length of time, token age can be utilized to lessen the complexity of the challenge that PoS miners must solve.

Machine learning

Machine learning is a sub-field of artificial intelligence that includes model identification and computational learning in the artificial construction of processes in the human brain. As stated in the proposition, "Can machines think like humans?" by Alan Turing [18], the starting point of artificial intelligence-based machine learning studies has been whether the learning ability of human beings will be in other objects on earth [19].

The learning ability of human beings is a distinguishing feature from other objects on earth, and it has brought the concept of machine learning with it. The concept of machine learning defined by Arthur Samuel, an American computer scientist in 1959, refers to a computer's ability to learn without explicit programming [20]. With the concept in question, SNARC (Stochastic Neural Analog Reinforcement Calculator), which is the first computer to be developed based on artificial neural networks, and the chess game introduced by Samuel Arthur were the first trials on whether machines could think like humans.

A data analysis technique that deals with the development and evaluation of algorithms, machine learning is a science that uses algorithms to help extract information from the vast amount of data that is currently available. It is the science that gives computers the ability to process without being explicitly programmed and for pattern recognition, classification, and prediction based on models derived from existing data defined as the capacity to choose effective features. It is a data analysis method that generates outputs from algorithms in data-driven modeling. Figure 7 shows the typical machine learning process.

All developments made on the basis of machine learning are based on the ability to perform human behavior by machine without additional human assistance from outside. In this context, the ability to infer certain models and patterns from the data is machine learning. In this whole process, first the data in the input is taken and then the relations within the data are found and it is aimed to output what the model has learned. Machine learning modeling will learn better and optimize the process as new data is added based on various algorithms to improve its performance and improve its intelligence over time.

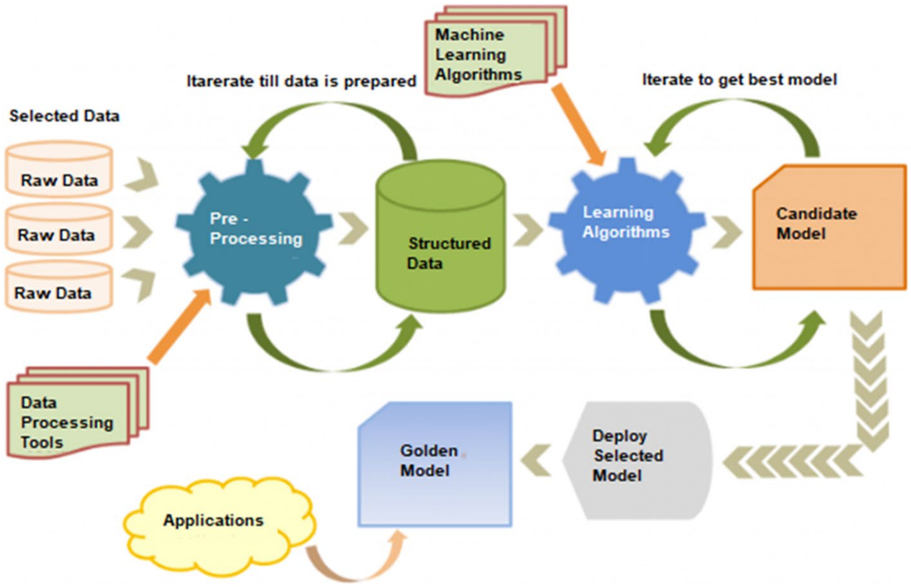

Fig. 7 Machine learning process

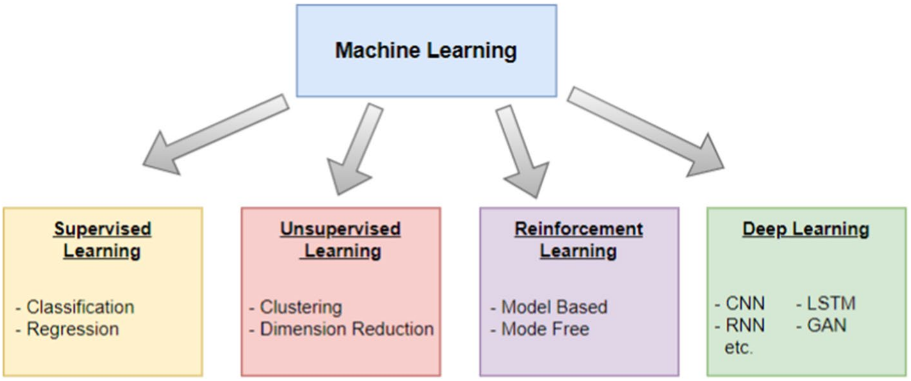

Fig. 8 Taxonomy of machine learning

Given that "learning" is the main focus of the discipline of machine learning, there are numerous sorts that a practitioner can run into. Certain forms of learning explain entire disciplines of research made up of numerous different kinds of algorithms.

Supervised learning is a machine learning paradigm where models are trained on labeled data, meaning the input features are paired with corresponding target labels. The algorithm learns to map input data to output labels by generalizing from the provided examples. In contrast, unsupervised learning involves training models on unlabeled data, aiming to uncover underlying patterns or structures within the dataset without explicit guidance. Algorithms in this category explore the data's inherent relationships, often through techniques like clustering or dimensionality reduction. Reinforcement learning, on the other hand, operates in an interactive environment where an agent learns to make sequential decisions to maximize a reward signal. Through trial and error, the agent receives feedback on its actions and refines its strategy over time, making it particularly well-suited for tasks that involve decision-making and long-term planning, such as game playing or autonomous control systems. Figure 8 shows the most common techniques used in machine learning which are described below.

- Classification is the task of determining which categorical structure a new observation is in, after which learning takes place from structures with categorical data observed in machine learning.
- Regression is modeling that analyzes the numerical values in the data set. By modeling according to the relationship between the independent X variables, the dependent y variable is tried to be estimated. Therefore, unlike classification, regression analysis produces a continuous output.
- Clustering is the separation of data into groups called "clusters" based on various proximity criteria. After the separation, it is expected that the expressions in the same data set will show similarity with each other, while the data in different groups will not show much similarity. It is a method that helps to intuitively divide similar data points into groups, and Euclidean distance measurement is most commonly used in distance measurement.
- Ensemble learning is learning in which more optimum results are obtained by using multiple machine learning algorithms instead of using a single algorithm in machine learning. A more accurate learning is achieved by combining multiple models, and strong predictions can be made with much lower variance (variability) and bias (systemic error) values after training.
- Dimensionality reduction is the process of filtering the desired data from high-dimensional data and reducing them to a smaller size due to the difficulty of storing and analyzing data, especially in parallel with the continuous increase in data in recent years. It is the process of retrieving the dimensions that best represent the existing multiple data structure or the new data combination as a combination of data dimensions to reduce the X dimension of the data to the Y dimension ($Y < X$). Removing unnecessary and meaningless high-dimensional structures in the data is very important for learning performance and optimization.
- Association rule is the process of inferring rules and making associations based on the relationship between the variables in the data. It is actively used especially in shopping and commerce platforms in order to derive propositions such as "those who bought this product also bought these products" or "those who watched this movie also watched these movies", and it is also actively used in determining the criminal profile.
- Deep learning is a subfield of machine learning that focuses on the development and application of artificial neural networks, specifically designed to simulate the complex structure and functioning of the human brain. It involves training these deep neural networks on large amounts of labeled data to automatically learn hierarchical representations of the input data. By iteratively adjusting the network's parameters, deep learning models are capable of extracting and recognizing intricate patterns, features, and relationships in data, enabling them to perform a wide range of tasks such as image and speech recognition, natural language processing, and even decision-making, often surpassing human-level performance in various domains.

- Reinforcement learning is a machine learning paradigm that focuses on training agents to make sequential decisions in an environment to maximize a cumulative reward signal. The agent interacts with the environment and learns through a trial-and-error process, where it receives feedback in the form of rewards or penalties based on its actions. By employing algorithms such as Q-learning or policy gradients, reinforcement learning enables the agent to learn optimal strategies by exploring different actions and observing the corresponding rewards. Through repeated iterations, the agent improves its decision-making abilities, leveraging the learned knowledge to navigate complex environments, solve challenging problems, and achieve long-term goals. Reinforcement learning has found applications in areas such as robotics, game playing, recommendation systems, and autonomous driving.

Machine learning reveals very effective and efficient outputs based on modeling within the learning methods mentioned above. With the increase in big data in parallel with technological developments, machine learning is one of the most popular concepts today. Because it is expected that the work done by humans, especially at the automation level, is started to be done by machines with various algorithms and still taking steps towards machine learning in new sectors optimize business processes, as well as bring along many sociological and even psychological changes in the society. In addition to the use of machine learning in many workplaces engaged in industrial production together with cyber-physical systems, its effectiveness in many areas such as financial services, disease diagnosis, cybersecurity, crime detection and prediction, transportation services and image processing is increasing, and in the near future, the army structure will be created by new robots.

Literature review

In the literature search, the contributions and gaps for studies in the fields of IOT, supply chain, medicine, finance and security were examined. For reference, all mentioned studies are listed in the Table 1 below.

Blockchain and machine learning in IOT

The Internet of Things (IoT) connects devices and enables them to share information. It has become a major advantage for industries such as agriculture, smart homes, and healthcare. However, the centralized architecture of IoT results in major security and privacy concerns. Traditional cryptography methods do not fully protect sensitive information. Therefore, a decentralized solution is necessary. Blockchain technology can provide such a solution by encrypting and digitally signing the data stored in each block, resulting in a high level of authenticity and security. This makes blockchain a suitable option for the healthcare industry, which requires a high level of trust among its many participants. In general, blockchain is ideal for highly distributed applications where tracking activities and maintaining the reliability of data is crucial.

The issue of security is one of the main difficulties with IoT. With the enormous number of interconnected devices, there is an increased risk of cyber attacks and data breaches. IoT devices may connect with one another in a safe, decentralized network thanks to blockchain technology. By using a distributed ledger, which is maintained by

Table 1 List of mentioned papers

Refs.	Domain	Paper title
[21]	Internet of Things	A blockchain-based machine learning framework for edge services in IIoT
[22]	Internet of Things	Detection of Security Attacks in Industrial IoT Networks: A Blockchain and Machine Learning Approach
[23]	Internet of Things	Dynamic access control policy based on blockchain and machine learning for the internet of things
[24]	Supply Chain	A blockchain and machine learning-based drug supply chain management and recommendation system for smart pharmaceutical industry
[25]	Supply Chain	A machine learning based approach for predicting blockchain adoption in supply Chain
[26]	Medicine	A blockchain and machine learning based framework for efficient health insurance management
[27]	Medicine	A novel blockchain-enabled heart disease prediction mechanism using machine learning
[28]	Medicine	Blockchain and machine learning in health care and management
[29]	Medicine	Blockchain-orchestrated machine learning for privacy preserving federated learning in electronic health data
[30]	Medicine	Healthcare Ledger Management: A Blockchain and Machine Learning-Enabled Novel and Secure Architecture for Medical Industry
[31]	Finance	An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning
[32]	Security	A machine learning approach for blockchain-based smart home networks security
[33]	Security	BChainGuard: A New Framework for Cyberthreats Detection in Blockchain Using Machine Learning
[34]	Security	Efficient privacy-preserving machine learning for blockchain network

multiple parties, blockchain can create a tamper-proof and transparent record of all transactions and data exchanges. IoT is facing difficulties in providing secure and private communications due to its large size and widespread deployment. Efforts have been made to use blockchain for decentralized protection and privacy, but these solutions require high levels of computation and time, making them unfeasible for many IoT applications. As IoT networks are integrated into critical industrial infrastructure, it is necessary to find alternative solutions to address potential security risks. To that end, solutions combining Blockchain and Machine Learning techniques have been implemented to address the threats to Industrial Internet of Things (IIoT) networks. By combining machine learning and blockchain techniques, a real-time approach to identifying and countering attackers in an IIoT network can be established, while also reducing the computational burden on network nodes when the network is secure and no extra encryption processes are taking place.

Regarding the use of machine learning in the field of cybersecurity, attack techniques are becoming more and more complex with the rapid development of web and mobile technologies. For this reason, machine learning that adapts to new and unknown conditions with various learning types in all kinds of challenging and complex structures is a potential resource. Today, when the use of machine learning in the field of cybercrime and security is examined; phishing detection, intrusion detection, authentication with keystroke gestures, testing the security of protocol applications, testing human security verifications, cryptology, spam e-mail, and message detection, insults over social networks, cyberbullying and terrorist crimes detection seems to be most recent studies.

In this context, it is considered that the modeling based on machine learning has an undeniable importance in the field of cybercrime and security, and the importance of machine learning in cybersecurity will increase even more on the basis of artificial intelligence, due to the increasing amount of data in our increasingly digitalized world, which does not allow for any other effective analysis. Additionally, blockchain can be used to manage IoT device identity and access control. By creating a decentralized identity system, which is based on blockchain technology, IoT devices can be securely authenticated and authorized to access specific resources or services. This can help to prevent unauthorized access and ensure the integrity of the IoT ecosystem.

Machine learning can also be used to enhance the intelligence of IoT systems. By analyzing the vast amount of data generated by IoT devices, machine learning algorithms can identify patterns, make predictions, and generate insights that can be used to optimize system performance and improve user experience. For example, machine learning can be used to predict equipment failures, optimize energy consumption, or improve supply chain efficiency.

Another potential application of blockchain and machine learning in IoT is in the area of smart cities. By using blockchain technology to create a decentralized and secure system for managing smart city infrastructure, stakeholders can improve efficiency, reduce costs, and increase transparency. Additionally, machine learning can be used to analyze data from multiple sources, such as traffic sensors, air quality sensors, and weather data, to optimize transportation routes, reduce energy consumption, and improve public safety.

Tian et al. [21] has proposed a blockchain and machine learning-based framework for improving edge services in the Industrial Internet of Things (IIoT). This framework, called BML-ES, leverages smart contracts to encourage collaboration among edge services, and includes an aggregation strategy to verify and combine model parameters for more accurate decision tree models. The framework also uses the SM2 public key cryptosystem to maintain the security and privacy of data in edge services. The results of theoretical analysis and simulations show that the BML-ES framework is secure, efficient, and effective in improving the accuracy of edge services in the IIoT. Nonetheless, this work still needs to explore how to lower communication overhead.

Vargas et al. [22] aimed to bring together previous approaches to create a comprehensive security system for IoT device networks. This solution would be able to identify potential threats, activate secure information transfer methods, and accommodate the computational limitations of industrial IoT. The proposed solution was successful in meeting these goals and is presented as a viable method for detecting and countering intrusions in an IoT network. However, this model is not always able to overcome traditional detection mechanisms such as intrusion detection systems. In their research, the processing of the machine learning and blockchain algorithms are executed in the collector node, requiring that this node concentrates all the data and has a higher processing capacity than the sensor nodes.

Outchakoucht et al. [23] proposed a dynamic and completely decentralized security policy for access control in the Internet of Things (IoT). This solution utilizes the blockchain to guarantee the highly distributed aspect required in IoT, while also incorporating machine learning algorithms, particularly those in the reinforcement learning category, to offer a dynamic, optimized, and self-adjusting security policy. But in order to address privacy concerns, their suggested framework needs an integrating notion of collective intelligence, a thorough case study, as well as an implementation as a practical proof of concept.

In addition to the challenges using blockchain and machine learning in IoT mentioned above, there is also need for interoperability and standardization. With an extensive number of different IoT devices and systems, there is a risk of fragmentation and incompatibility. To fully realize the benefits of these technologies, stakeholders will need to agree on common standards and protocols. However, there are initiatives underway, such as the Open Connectivity Foundation, which are working to create standards for IoT interoperability.

Blockchain and machine learning in supply chain

One of the primary benefits of blockchain technology in supply chain management is its ability to create a secure and immutable record of transactions. By using a distributed ledger system, which is maintained by multiple parties, blockchain can create a transparent and tamper-proof record of all supply chain transactions. This can help to reduce fraud, increase transparency, and improve trust between supply chain partners. Additionally, blockchain technology can be used to create a more efficient supply chain by automating certain processes. For example, smart contracts can be used to automatically trigger transactions or payments when certain conditions are met. This can help to reduce the need for intermediaries and reduce costs in the supply chain.

Machine learning can also be used to improve the efficiency and accuracy of supply chain management. By analyzing large amounts of data from multiple sources, including sensors, social media, and transaction records, machine learning algorithms can identify patterns and predict future trends. This can help supply chain stakeholders to make more informed decisions, reduce waste, and improve supply chain visibility.

One of the most promising applications of blockchain and machine learning in supply chain management is in the area of traceability. Traceability is a critical issue in supply chain management, particularly in industries such as food and pharmaceuticals, where product safety is a significant concern. By using blockchain technology to create a secure and transparent record of all supply chain transactions, stakeholders can quickly and accurately trace products back to their source, which can help to prevent foodborne illnesses or counterfeit products.

Machine learning can also be used to improve traceability by analyzing data from multiple sources, including GPS data, sensor data, and transaction records. This can help to identify potential supply chain issues before they become a problem, such as delays or quality issues.

Another potential application of blockchain and machine learning in supply chain management is in the area of sustainability. Sustainability is becoming an increasingly important issue for supply chain stakeholders, as consumers and regulators demand more environmentally friendly practices. By using blockchain technology to create a transparent record of all supply chain transactions, stakeholders can monitor and report on their sustainability practices. Additionally, machine learning can be used to identify opportunities for improvement, such as reducing waste or optimizing transportation routes.

Changes are validated based on the consent of all parties involved in a blockchain, where transactions are continuously recorded and handled in a secure and verifiable manner. Transactions cannot be changed or deleted after being approved by all parties, providing benefits like data integrity and security. Using blockchain in supply chain transactions improves security, openness, traceability, and productivity. The use of blockchain outside of financial services is mostly experimental and focuses on the technology rather than the issues of selection and implementation. Additionally, it enables better supply chain integration, resulting in improved overall performance. For high-value products, deploying the technology might be economically advantageous, but it might be difficult for low-cost ones. By offering real-time product tracking, reduced product transportation costs, highly secure transactions, and protection against counterfeiting, the supply chain powered by blockchain technology increases customer trust. It upgrades the conventional supply chain strategy into a more reliable, automated, secure, auditable, and transparent system and entirely blocks the entry of counterfeit goods.

Abbas et al. [24] developed a cutting-edge blockchain and machine learning-based drug supply chain management and recommendation system (DSCMR). The machine learning-based drug recommendation system for customers and the blockchain-based drug supply chain management make up the system. Hyperledger fabrics are used to set up the medication supply chain management, which keeps track of the drug distribution process in the pharmaceutical sector. Based on trained data from a public drug review dataset from UCI, the recommendation system uses N-gram and LightGBM models to offer the best medications to clients. A REST API is used to connect the blockchain system and the machine learning module. However, to verify the effectiveness and validity of the system, they must increase the network size and implement their machine learning models in real-time pharmaceutical companies. This will improve their machine learning models' accuracy and recommendation outcomes.

Kamble et al. [25] suggested using machine learning to estimate an organization's chances of adopting blockchain technology successfully. The report sees blockchain technology as a dynamic skill that businesses must have to remain competitive and identifies the critical drivers of blockchain adoption, including partner preparedness, competition pressure, perceived usefulness, and perceived user-friendliness. But in this study, to assess the adoption likelihood, the practitioner will need to substitute these probability values (high or low), depending on what is appropriate for their organization. The implementation of a decision support system will assist the decision-makers in determining their current likelihood of adoption and creating adoption plans.

One obstacle of using blockchain and machine learning in supply chain management is the need for standardization. Supply chains can be complex and involve multiple parties, each with their own systems and processes. To fully realize the benefits of blockchain and machine learning in supply chain management, stakeholders will need to agree on common standards and protocols. However, there are initiatives underway, such as the Blockchain in Transport Alliance, that are working to create standards for blockchain-based supply chain management.

Blockchain and machine learning in medicine

Blockchain and machine learning technologies are revolutionizing the healthcare industry, particularly in the field of medicine. These technologies have the potential to improve patient outcomes, enhance data security, and increase efficiency in medical research and clinical trials.

One of the primary benefits of using blockchain technology in medicine is the potential to create a secure and decentralized system for storing and sharing patient data. By using a distributed ledger, which is maintained by multiple parties, blockchain can create a tamper-proof and transparent record of all medical transactions and data exchanges. This can help to improve patient privacy and data security, which are critical considerations in the healthcare industry.

Additionally, blockchain can be used to create a decentralized identity system, which is based on blockchain technology, to securely authenticate and authorize patient access to medical records. By doing so, patients can control who has access to their data, thereby enhancing their privacy and security.

Machine learning can also be used to analyze medical data and improve patient outcomes. By analyzing large datasets of medical records, machine learning algorithms can identify patterns, predict outcomes, and generate insights that can help doctors diagnose diseases, develop treatment plans, and improve patient outcomes.

One promising application of blockchain and machine learning in medicine is in the field of clinical trials. By using blockchain technology to create a secure and transparent system for managing clinical trial data, stakeholders can improve the efficiency and accuracy of the trial process. Furthermore, machine learning can be used to analyze data from multiple sources, such as patient medical records, genetic data, and clinical trial data, to identify potential treatment options and improve patient outcomes.

Another potential application of blockchain and machine learning in medicine is in the area of drug supply chain management. By using blockchain technology to create a transparent and secure record of all drug supply chain transactions, stakeholders can quickly and accurately trace drugs back to their source, which can help to prevent counterfeiting and ensure drug safety. Additionally, machine learning can be used to analyze data from multiple sources, such as clinical trial data and drug efficacy data, to optimize drug development and improve patient outcomes.

Blockchain technology may be applied to a wide range of devices and is used in healthcare to guarantee the privacy of countless medical records. Electronic health records and remote patient monitoring are now possible in the healthcare sector thanks to the Internet of Things. The enormous volume of healthcare data produced by numerous sources can be problematic for data quality. Blockchain technology synchronizes information across healthcare providers, provides a solution to these issues. Each block contains private health information that can only be accessed by those with permission. The advantages of using blockchain in the healthcare sector include decentralization, consent management, immutability, and enhanced capacity. By preventing unauthorized alterations, the immutability of blockchain data enables the creation of disease prediction models utilizing machine learning algorithms.

Goyal et al. [26] aimed to create a framework that leverages blockchain technology and machine learning for the health insurance sector, which is both quick and economical.

The current health insurance system has two major issues, slowness and high cost, but the proposed blockchain-based health insurance model has resolved these issues. The results indicate that the proposed model is trustworthy, affordable, and swift. The downside to this model is that it was tested using only random forest classifier. It needs to be tested in comparison with more algorithms.

Using data saved on a blockchain, Hasanova et al. [27] suggested a heart disease prediction system based on machine learning and the Sine Cosine Weighted K-Nearest Neighbor (SCA_WKNN) approach. A safe, impenetrable source of information for learning and storing patient data is the blockchain. When the SCA_WKNN algorithm's performance was compared to that of other algorithms in terms of accuracy, precision, recall, F-score, and root mean square error, it revealed improvements in accuracy of 4.59% and 15.61% over W K-NN and K-NN, respectively. Peer-to-peer storage and the decentralized storage offered by the blockchain were also evaluated for latency and throughput; the blockchain-based decentralized storage outperformed the peer-to-peer storage by 25.03%. One of the drawbacks of the proposed system is the high cost of the operation depending on how many transactions are made through the system. Also, this technique is not recommended for low latency applications because of the modest delay in transaction times caused by the system's decentralized structure.

Jain et al. [28] utilized a supervised learning approach to train a machine learning algorithm on datasets obtained from sources like MedLine. To simplify the data, they applied the "bag of words" algorithm to reduce dimensionality. A blockchain network was used to safeguard patient healthcare data, enabling secure interactions between patients and licensed physicians. The trained model was given a fresh batch of medical data, which it sorted by disease after removing any personal information. Their article suggests a novel healthcare model that, while yet in its infancy, can undoubtedly serve as a foundation for numerous further healthcare models in the future.

Passerat-Palmbach et al. [29] investigated the combination of blockchain and machine learning in more depth, focusing on the decentralization and federation of the learning process, as well as the audibility and incentivization it enables. They evaluated the cost-benefit of prior work and established a framework for a sophisticated blockchain-powered machine learning system for privacy-preserving federated learning in healthcare, offering new value in the field of health. Their method has limitations, including the discoverability of data and analytical processes on the safe public blockchain while maintaining the privacy of the analytical processes and the value created by producing data/compute matches that were previously forbidden, immoral, and impractical.

Khan et al. [30] concentrated on two key goals. Initially, they suggested a stochastic gradient descent method based on machine learning for managing medical records and streamlining routine operations of e-Healthcare systems. This technique assesses the loss of medical data during computation and guarantees effective data transmission. Second, to safeguard transactions and guarantee immutable storage, they suggested a cutting-edge, secure, and serverless blockchain-based architecture for the medical sector. In order to preserve health-related information utilizing blockchain Technology, this architecture combines ledger optimization, secure management, protection, integrity, anti-forgery, and controlled access.

In addition to challenges described for each study above, there is a risk of fragmentation and incompatibility with a high number of different healthcare systems and technologies. To fully realize the benefits of these technologies, stakeholders will need to agree on common standards and protocols. However, there are initiatives underway, such as the Global Consortium for Healthcare Blockchain, which are working to create standards for healthcare interoperability.

Blockchain and machine learning in finance

One of the most promising applications of blockchain and machine learning in finance is fraud detection. Fraud is a significant problem in the financial industry, and it can be difficult to detect and prevent. However, by using machine learning algorithms to analyze financial data and blockchain technology to create a secure and transparent ledger, financial institutions can quickly identify fraudulent activities and prevent them from causing significant damage. By combining these two technologies, banks can create a secure and efficient system for detecting and preventing fraud in real-time.

Another potential application of blockchain and machine learning in finance is in the area of loan underwriting. Traditionally, the loan underwriting process involves a significant amount of manual labor and paperwork, which can be time-consuming and error-prone. However, by using machine learning algorithms to analyze data from multiple sources, including social media, financial institutions can quickly and accurately determine a borrower's creditworthiness. Additionally, by leveraging blockchain technology, lenders can create a secure and immutable record of loan transactions, which can help to reduce the risk of fraud and increase the efficiency of the loan underwriting process.

Decentralized financial systems, known as decentralized finance (DeFi) is another application of blockchain technology in finance. DeFi platforms use blockchain technology to create a transparent and secure system for financial transactions, without the need for intermediaries such as banks or other financial institutions. By using smart contracts, which are self-executing contracts that automatically enforce the terms of the agreement, DeFi platforms can create a more efficient and secure financial system. By using machine learning algorithms to analyze financial data, DeFi platforms can provide personalized financial services to users, such as investment advice or automated trading strategies.

Cryptocurrencies are a type of virtual currency that utilize cryptography for protection. They are decentralized and have an open-source nature, operating on a peer-to-peer network. Cryptocurrencies primarily use complex cryptographic algorithms that require a network of computers to perform complex mathematical computations.

Chowdhurry et al. [31] utilized machine learning methods on the indices and components of cryptocurrencies in order to predict and forecast their prices. The objective was to use machine learning algorithms and models to predict the close (closing) price of the cryptocurrency index 30 and 9 components, making it simpler for users to trade in these currencies. Various machine learning techniques and algorithms were employed and the models were compared to determine the most accurate results. Using an ensemble learning method, they achieved an accuracy of 92.4%. As a drawback in their study, K-NN model has not performed well when used for forecasting, which has been caused by the existence of noisy random characteristics and high volatility.

The main drawback of using blockchain and machine learning in finance is the need for large amounts of data. Machine learning algorithms require large datasets to train and improve, which can be challenging in the financial industry, where data is often sensitive and difficult to obtain. However, advances in data privacy and security, as well as the increased adoption of blockchain technology, are making it easier for financial institutions to collect and analyze large amounts of data by providing a transparent and immutable ledger that securely records financial transactions. This enables financial institutions to access a comprehensive and reliable dataset, eliminating the need for reconciling multiple disparate systems, reducing data discrepancies, and facilitating efficient data analysis for various purposes such as risk assessment, auditing, compliance, and financial reporting. Blockchain's decentralized nature allows for enhanced data sharing and collaboration among multiple parties, further streamlining the data collection and analysis process in the financial industry.

Blockchain and machine learning in security

One of the key benefits of using blockchain in security is the ability to create a tamper-proof and transparent record of all security-related transactions. By using a distributed ledger, which is maintained by multiple parties, blockchain can create a system that is resistant to tampering or alteration. This can be used to create a secure record of all security-related transactions, such as network access attempts, software updates, and system changes. This can help to improve security by creating a clear and transparent record of all security-related activities, which can be audited and verified by multiple parties.

Machine learning can also be used in security to improve threat detection and response. By analyzing large datasets of security-related data, machine learning algorithms can identify patterns, detect anomalies, and generate insights that can help security professionals to identify and respond to threats more quickly and effectively. This can help to improve security by reducing the time between detection and response, which can be critical in preventing cyber-attacks.

Another potential application of blockchain and machine learning in security is in the area of identity and access management. By using blockchain technology to create a decentralized identity system, which is based on blockchain technology, stakeholders can create a secure and transparent system for authenticating and authorizing user access to digital resources. This can help to improve security by reducing the risk of identity theft, unauthorized access, and other security threats. Also, machine learning can be used to analyze user behavior data, such as access logs and usage patterns, to identify potential security risks and anomalies. This can help security professionals to detect and respond to threats more quickly and effectively, thereby improving overall security.

Recently, blockchain technology has emerged as a robust decentralized solution for securing data integrity. The integration of smart contracts in blockchain provides a secure environment for building peer-to-peer applications. While blockchain has been widely adopted by the research community as a means of protecting against cyberattacks, the technology itself may also be the target of such threats.

Blockchain uses decentralized consensus algorithms for verifying and validating transactions, which are intended to become an integral part of the blockchain network, as opposed to conventional centralized security and privacy techniques. Many of the ML algorithms now in use, nevertheless, rely on centralized frameworks, which can result in security lapses and single points of failure. The trustworthiness of data is essential for ML algorithms to deliver correct results because centralized authority poses concerns about maintaining privacy, false authentication, and data tampering. For some situations, even a minor security flaw in the ML algorithm can lead to large false-positive rates. Additionally, the computation of ML models often relies on a trusted third party (e.g. a cloud service provider), which raises privacy concerns. As a result, there is a need for decentralized ML frameworks, and blockchain could be a potential solution. Moreover, ML integration into blockchain aids in problem analysis and enhances the network's overall security and privacy.

The privacy and security aspects must be taken into account from the design stage onward to produce a machine learning model for a blockchain system that can be trusted. The model should prevent any privacy breaches from the data during the learning process. This is because databases often contain sensitive information about individuals, such as medical records. Even though the learning process may only produce summarized information, partial sensitive information can still be derived. Differential privacy (DP), which introduces noise via a random technique, addresses privacy breaches while protecting personal information. On the subject of system security, computation in a distributed network involving numerous entities can occasionally result in system failure because of computational error brought on by unreliable workers or malicious activity on the part of individuals working together to reduce the accuracy of the ML model by providing false local gradients.

Khan et al. [32] proposed a blockchain-based solution for secure and private IoT that utilizes computational resources found in typical IoT environments, such as smart homes, and a Deep Extreme Learning Machine (DELM) instance. The proposed solution, Smart Home Architecture based on blockchain, prioritizes privacy, integrity, accessibility and has been tested to ensure its reliability. The simulation results show that the overhead created by the method is minimal compared to its security and privacy benefits. The proposed DELM blockchain-based architecture was evaluated using statistical methods, which showed that it was much more reliable than other algorithms, achieving 93.91% accuracy. However, this model requires further expansion by the use of additional datasets and different architectural designs.

Aladhadh et al. [33] introduced a framework named BChainGuard for detecting cyber threats in blockchain. The objective of the framework is to identify normal and abnormal behavior in the traffic related to the blockchain network. The classification technique in BChainGuard will be executed locally and the decision function will be embedded as a smart contract. The results of the experiments are promising, with detection accuracy of approximately 95% using SVM and 98.02% using MLP, and a low runtime with minimal gas consumption overhead. The weakness of this approach is that it needs to use federated learning in place of machine learning when the dataset is unavailable to help maintain privacy.

Kim et al. [34] addressed the privacy, security, and performance issues by introducing a privacy-centric machine learning model for a permissioned blockchain. The model comprises of an error-based aggregation mechanism and a differentially private stochastic gradient descent algorithm. Any differentially private learning procedure that calls for the definition of non-deterministic functions can be handled by their model. Attacks by adversarial nodes that try to make the DML model less accurate are repelled by the error-based aggregation rule. The outcomes of their trials demonstrated that, in a differentially private environment, the suggested architecture is more resistant to adversarial attacks than other aggregation rules. The suggested model also offers a high degree of usability due to its minimal processing complexity and transaction latency. Applying the modularized model to the current Hyperledger Fabric open-source considered to be a future task in this study.

Real world examples

Blackbox AI is an example from the real world that use Blockchain and machine learning to streamline and automate the workflow, management, and verification procedures in software development [35]. It is an artificial intelligence coding assistant that offers real-time code completion, documentation, and debugging advice to developers.

An illustration of a supply chain application with AI and blockchain technology is the DHL Global Trade Barometer [36]. It is a brand-new and distinctive early indicator for the current situation and potential growth of global trade. Its foundation is a sizable amount of logistics data that has been examined with artificial intelligence.

The Agr-Food supply chain management solution from AgrBlockIoT is an excellent example of how AI and blockchain can be used in the agricultural sector to provide transparency and traceability [37]. It supports intelligent farming, making it possible to track logistics effectively and improve operational procedures.

The world's first global patent register driven by AI and blockchain, IPwe, addresses issues with erroneous data, out-of-date ownership information, and a lack of transparency in the IP ecosystem [38]. IPwe can quickly examine patent data by fusing natural language processing (NLP), predictive analytics, and machine learning from IBM Watson. It can then make use of the data to provide summaries and analyses that will assist users in spotting profitable opportunities while avoiding potential business dangers.

Conclusion

Blockchain can improve the application of ML by supplying security, anonymity, decentralized intelligence, and reliable decision-making for data and model sharing. Through the use of cryptographic techniques, blockchain systems may safely store massive amounts of data and guarantee the privacy and accountability of the learning process and the final ML model. Secure access control without relying on centralized entities is made possible by decentralized blockchain architecture. Using smart contracts and DApps in blockchain systems, decentralized machine learning applications can also be made possible. Decentralized ML applications can benefit from easy audits and improved collaboration thanks to the usage of blockchain methods, which also makes it possible for transparent records of the data and variables used by ML algorithms in their decision-making processes. Furthermore, ML may enhance the functionality of blockchain by boosting energy and resource efficiency, scalability, security, and privacy, as well as by delivering intelligent smart contracts. The energy sector can manage tasks more intelligently by using ML algorithms, which also increases resource and energy efficiency. ML approaches can optimize data upkeep and storage, identify harmful activity on the blockchain to stop theft, fraud, and illegal transactions, and address scalability difficulties. For example, Liao et al. decreased data latency by 23.5% and increased convergence rate by 15% with use of Q-Learning based optimization in real-world data [42]. Mao et al. saved an average of 2MB of storage for each image by using Attention U-Net framework [43]. NLP approaches can also be used to more efficiently construct and run sophisticated smart contracts. By using self-writing smart contracts, this enables a secure and affordable method for exchanging cash, assets, shares, or anything else of value. Gogineni et al. used a variant of LSTM and reduced the class imbalance by considering only distinct opcode combinations for normal contracts and achieved a weighted average F1 score of 90.0% [44]. Choudhury et al. developed a framework that automatically generates smart contracts from domain-specific business rules in regulatory documents and achieved a precision of 0.95 across 20 training and test protocols [45].

Although the integration of both blockchain and machine learning technologies is seen as potentially promising solutions, their use in network and communication systems currently faces many unresolved problems and hurdles. According to the trilemma, blockchain systems can only have a maximum of two of the three characteristics-scalability, decentralization, and security [39]. The trilemma suggests that there is often a trade-off between these three fundamental aspects of blockchain technology, meaning that improving one aspect may come at the expense of the others. Security guarantees the system's immutability and resistance to assaults, while scalability takes care of the system's capacity to handle transactions. Decentralization enables the system to be fault-tolerant and attack-resistant.

Traditional blockchain networks have inherent limitations on the number of transactions they can process per second. For instance, Bitcoin handles around 7 transactions per second, and Ethereum around 15–45 TPS (Transactions Per Second) [40]. Scalability issues can lead to longer confirmation times for transactions. In scenarios where real-time processing is crucial, this latency can be a significant hurdle. Storing large-scale machine learning models and datasets on a blockchain can be impractical due to storage constraints. This limits the types of applications that can effectively utilize blockchain-based machine learning. The combination of limited computing power and scalability issues can result in high costs for executing machine learning algorithms on a blockchain. This can be a major barrier, especially for resource-constrained applications. Applications that require high transaction throughput or low latency may find it challenging to operate on existing public blockchains. This could lead to slower adoption of blockchain-based machine learning in critical domains. Also, managing complicated communication and networking systems with numerous users that have different quality of service (QoS) requirements remains a challenge. Massive amounts of training data are often needed for ML systems, and this data is frequently implemented at a central network controller with ample storage and processing power. Nevertheless, with the present communication systems, it might not be possible to retrieve such massive amounts of data for ML training. It is also difficult to aggregate data in heterogeneous networks for ML training.

Actionable recommendations to address the potential mitigation strategies for overcoming barriers to successfully implementing blockchain with machine learning include emerging hybrid architectures where critical operations are performed off-chain or on specialized servers, while the blockchain is used for verification and auditing. Choosing machine learning algorithms that are computationally efficient and suitable for the specific task at hand can reduce the demand for high computational power. Implementing layer-2 solutions like Lightning Network for Bitcoin or Layer-2 scaling solutions for Ethereum can increase transaction throughput and reduce confirmation times [41]. Designing systems that leverage a combination of off-chain and on-chain computations allows for flexibility in resource allocation based on the specific requirements of each task.

The lack of clear regulatory and legal frameworks is a significant barrier to the implementation of blockchain and machine learning for several reasons. Without clear regulations, businesses and developers may struggle to understand what compliance standards they need to meet. This uncertainty can lead to hesitation or reluctance to invest in blockchain and machine learning projects. Also, there is legal ambiguity surrounding smart contracts which can lead to confusion about the legal standing of automated agreements. There may be ambiguity about who owns and controls data on a blockchain. Furthermore, blockchain is a global technology, and transactions can occur across borders seamlessly. This creates challenges in terms of determining the jurisdiction that governs transactions, which can lead to legal complexities. Intellectual property issues, including patents and copyrights related to blockchain and machine learning technologies, can be unclear. Establishing ownership and licensing agreements in this context can be challenging. In cases where consumers are involved, clear mechanisms for dispute resolution and consumer protection may be lacking, potentially leaving users vulnerable to fraud or disputes.

A secure and decentralized database, anonymous transactions, utilization of smart contracts, lower transaction fees, and traceability of products are some of the factors driving blockchain's adoption. Despite the benefits offered by these drivers, the adoption of blockchain with machine learning is still in its early stages. Barriers to its successful implementation include the absence of successful implementations, difficulties in integrating with existing systems, scalability issues, limited computing power, and lack of clear regulatory and legal frameworks.

There are several potential future developments and trends for the integration of blockchain and machine learning. Federated learning, a technique where a model is trained across multiple devices or servers holding local data samples, could be integrated with blockchains to ensure the integrity of updates and consensus on model updates. Enhanced techniques for privacy-preserving machine learning, such as secure multiparty computation and homomorphic encryption, may be further integrated with blockchain technology to enable secure and private data processing. Efforts still continue to develop and adopt more energy-efficient consensus mechanisms for public blockchains, reducing the environmental impact and making them more suitable for computationally intensive tasks like machine learning.

Abbreviations

AI	Artificial intelligence
BML-ES	Blockchain and machine learning-based framework for improving edge services
DAPP	Decentralized application

DeFi	Decentralized finance
DELM	Deep extreme learning machine
DLT	Distributed ledger technology
DP	Differential privacy
DSCMR	Drug supply chain management and recommendation system
EVM	Ethereum virtual machine
GPS	Global positioning system
IIoT	Industrial internet of things
IoT	Internet of things
ML	Machine learning
MLP	Multi layer perceptron
NLP	Natural language processing
P2P	Peer to peer
PoW	Proof of work
PoS	Proof of stake
QoS	Quality of service
SCA WKNN	Sine cosine weighted K-nearest neighbor
SNARC	Stochastic neural analog reinforcement calculator
SVM	Support vector machine
UCI	University of California, Irvine

Acknowledgements
Not applicable.

Author contributions
SK performed the literature search, analyzed results, and wrote the manuscript. TMK guided the direction of the search, suggested the techniques and helped to finalize the work. Both authors read and approved the final manuscript.

Funding
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Availability of data and materials
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 6 June 2023 Accepted: 28 October 2023
Published online: 06 January 2024

References

1. Nakamoto S. Bitcoin whitepaper. https://bitcoin.org/bitcoin.pdf- 17. 07. 2019; 2008.
2. Zheng Z, Xie S, Dai H-N, Chen X, Wang H. Blockchain challenges and opportunities: a survey. Int J Web Grid Serv. 2018;14(4):352–75.
3. Natarajan H, Krause S, Gradstein H. Distributed ledger technology and blockchain 2017.
4. Haber S, Stornetta WS. How to time-stamp a digital document. Berlin: Springer; 1991.
5. Niranjanamurthy M, Nithya B, Jagannatha S. Analysis of blockchain technology: pros, cons and swot. Clust Comput. 2019;22:14743–57.
6. Chu S, Wang S. The curses of blockchain decentralization. arXiv preprint arXiv:1810.02937 2018.
7. Fox G. Peer-to-peer networks. Comput Sci Eng. 2001;3(3):75–7.
8. Romero Ugarte JL. Distributed ledger technology (dlt): introduction. Banco de Espana Article. 2018;19:18.
9. Sheth H, Dattani J. Overview of blockchain technology. Asian J Convergence Technol (AJCT) ISSN-2350-1146, 2019.
10. Vujicic D, Jagodic D, Randic S. Blockchain technology, bitcoin, and ethereum: a brief overview. In: 2018 17th International Symposium Infoteh-jahorina (infoteh), pp. 1–6, 2018. IEEE.
11. Kiayias A, Panagiotakos G. Speed-security tradeoffs in blockchain protocols. Cryptology ePrint Archive 2015.
12. Hirai Y. Defining the ethereum virtual machine for interactive theorem provers. In: Financial Cryptography and Data Security: FC 2017 International Workshops, WAHC, BITCOIN, VOTING, WTSC, and TA, Sliema, Malta, April 7, 2017, Revised Selected Papers 21, 2017;520–535. Springer.
13. Mohanta BK, Panda SS, Jena D. An overview of smart contract and use cases in blockchain technology. In: 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2018;1–4. IEEE.

14. Zheng Z, Xie S, Dai H-N, Chen W, Chen X, Weng J, Imran M. An overview on smart contracts: challenges, advances and platforms. Futur Gener Comput Syst. 2020;105:475–91.
15. Szabo N. Formalizing and securing relationships on public networks. First monday 1997.
16. Mingxiao D, Xiaofeng M, Zhe Z, Xiangwei W, Qijun C. A review on consensus algorithm of blockchain. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2567–2572; 2017. IEEE.
17. Sriman B, Ganesh Kumar S, Shamili P. Blockchain technology: Consensus protocol proof of work and proof of stake. In: Intelligent Computing and Applications: Proceedings of ICICA 2019, pp. 395–406 2021. Springer.
18. Turing AM. Computing machinery and intelligence. Netherlands: Springer; 2009.
19. Kayikci S. A deep learning method for passing completely automated public turing test. In: 2018 3rd International Conference on Computer Science and Engineering (UBMK), 2018;41–44. IEEE.
20. Samuel AL. Machine learning. Technol Rev. 1959;62(1):42–5.
21. Tian Y, Li T, Xiong J, Bhuiyan MZA, Ma J, Peng C. A blockchain-based machine learning framework for edge services in iiot. IEEE Trans Industr Inf. 2021;18(3):1918–29.
22. Vargas H, Lozano-Garzon C, Montoya GA, Donoso Y. Detection of security attacks in industrial iot networks: a blockchain and machine learning approach. Electronics. 2021;10(21):2662.
23. Outchakoucht A, Hamza E-S, Leroy JP. Dynamic access control policy based on blockchain and machine learning for the internet of things. Int J Adv Comput Sci Appl. 2017;8(7).
24. Abbas K, Afaq M, Ahmed Khan T, Song W-C. A blockchain and machine learning-based drug supply chain management and recommendation system for smart pharmaceutical industry. Electronics. 2020;9(5):852.
25. Kamble SS, Gunasekaran A, Kumar V, Belhadi A, Foropon C. A machine learning based approach for predicting blockchain adoption in supply chain. Technol Forecast Soc Chang. 2021;163: 120465.
26. Goyal A, Elhence A, Chamola V, Sikdar B. A blockchain and machine learning based framework for efficient health insurance management. In: Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems, pp. 511–515; 2021.
27. Hasanova H, Tufail M, Baek U-J, Park J-T, Kim M-S. A novel blockchain-enabled heart disease prediction mechanism using machine learning. Comput Electr Eng. 2022;101: 108086.
28. Jain S, Anand A, Gupta A, Awasthi K, Gujrati S, Channegowda J. Blockchain and machine learning in health care and management. In: 2020 International Conference on Mainstreaming Block Chain Implementation (ICOMBI), 2020;1–5. IEEE.
29. Passerat-Palmbach J, Farnan T, McCoy M, Harris JD, Manion ST, Flannery HL, Gleim B. Blockchain-orchestrated machine learning for privacy preserving federated learning in electronic health data. In: 2020 IEEE International Conference on Blockchain (Blockchain), 2020;550–555. IEEE.
30. Khan AA, Laghari AA, Shafiq M, Cheikhrouhou O, Alhakami W, Hamam H, Shaikh ZA. Healthcare ledger management: A blockchain and machine learning-enabled novel and secure architecture for medical industry. Human-Centric Comput Informat Sci. 2022;12.
31. Chowdhury R, Rahman MA, Rahman MS, Mahdy M. An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning. Physica A. 2020;551: 124569.
32. Khan MA, Abbas S, Rehman A, Saeed Y, Zeb A, Uddin MI, Nasser N, Ali A. A machine learning approach for blockchain-based smart home networks security. IEEE Network. 2020;35(3):223–9.
33. Aladhadh S, Alwabli H, Moulahi T, Al Asqah M. Bchainguard: a new framework for cyberthreats detection in blockchain using machine learning. Appl Sci. 2022;12(23):12026.
34. Kim H, Kim S-H, Hwang JY, Seo C. Efficient privacy-preserving machine learning for blockchain network. IEEE Access. 2019;7:136481–95.
35. BlackBox AI. https://www.useblackbox.io/. Accessed: 19 Sept 2023.
36. DHL Global Trade Barometer. https://lot.dhl.com/global-trade-barometer-gtb/. Accessed 19 Sept 2023.
37. Agr-Food supply chain management. 3. https://www.hindawi.com/journals/jfq/2022/4228448/. Accessed 19 Sept 2023.
38. IP transaction platform IPwe. https://www.ibm.com/case-studies/ipwe/. Accessed 19 Sept 2023.
39. Altarawneh A, Herschberg T, Medury S, Kandah F, Skjellum A. Buterin's scalability trilemma viewed through a state-change-based classification for common consensus algorithms. In: 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), 2020;0727–0736. https://doi.org/10.1109/CCWC47524.2020.9031204.
40. Sarode RP, Singh DG, Watanobe Y, Bhalla S. High-volume transaction processing in bitcoin lightning network on blockchains. Int J Comput Sci Eng. 2023;26(4):445–58.
41. Poon J, Dryja T. The bitcoin lightning network. Scalable o-chain instant payments, 2015;20–46.
42. Liao Z, Peng J, Chen Y, Zhang J, Wang J. A fast q-learning based data storage optimization for low latency in data center networks. IEEE Access. 2020;8:90630–9.
43. Mao D, Li Z, Chen Z, Rao H, Zhang J, Liu Z. A semantic segmentation algorithm for distributed energy data storage optimization based on neural networks. In: 2022 IEEE 7th International Conference on Smart Cloud (SmartCloud), 2022;115–120. IEEE.
44. Gogineni AK, Swayamjyoti S, Sahoo D, Sahu KK, Kishore R. Multi-class classification of vulnerabilities in smart contracts using awd-lstm, with pre-trained encoder inspired from natural language processing. IOP SciNotes. 2020;1(3): 035002.
45. Choudhury O, Dhuliawala M, Fay N, Rudolph N, Sylla I, Fairoza N, Gruen D, Das A. Auto-translation of regulatory documents into smart contracts. IEEE Blockchain Initiative (September), 2018;1–5.

Publisher's Note