Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2023.044529

ARTICLE

Blockchain-Based Cognitive Computing Model for Data Security on a Cloud Platform

Xiangmin Guo[1,2], Guangjun Liang[1,2,*], Jiayin Liu[1,2] and Xianyi Chen[3,*]

[1]Department of Computer Information and Cyber Security, Jiangsu Police Institute, Nanjing, 210031, China

[2]Jiangsu Electronic Data Forensics and Analysis Engineering Research Center, Jiangsu Police Institute, Nanjing, 210031, China

[3]School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, 210044, China

*Corresponding Authors: Guangjun Liang. Email: Liang_Forensics@163.com, guangjun_liang345@outlook.com; Xianyi Chen. Email: 002582@nuist.edu.cn

Received: 01 August 2023 Accepted: 23 October 2023 Published: 26 December 2023

ABSTRACT

Cloud storage is widely used by large companies to store vast amounts of data and files, offering flexibility, financial savings, and security. However, information shoplifting poses significant threats, potentially leading to poor performance and privacy breaches. Blockchain-based cognitive computing can help protect and maintain information security and privacy in cloud platforms, ensuring businesses can focus on business development. To ensure data security in cloud platforms, this research proposed a blockchain-based Hybridized Data Driven Cognitive Computing (HD2C) model. However, the proposed HD2C framework addresses breaches of the privacy information of mixed participants of the Internet of Things (IoT) in the cloud. HD2C is developed by combining Federated Learning (FL) with a Blockchain consensus algorithm to connect smart contracts with Proof of Authority. The "Data Island" problem can be solved by FL's emphasis on privacy and lightning-fast processing, while Blockchain provides a decentralized incentive structure that is impervious to poisoning. FL with Blockchain allows quick consensus through smart member selection and verification. The HD2C paradigm significantly improves the computational processing efficiency of intelligent manufacturing. Extensive analysis results derived from IIoT datasets confirm HD2C superiority. When compared to other consensus algorithms, the Blockchain PoA's foundational cost is significant. The accuracy and memory utilization evaluation results predict the total benefits of the system. In comparison to the η values 0.004 and 0.04, the η value of 0.4 achieves good accuracy. According to the experiment results, the number of transactions per second has minimal impact on memory requirements. The findings of this study resulted in the development of a brand-new IIoT framework based on blockchain technology.

KEYWORDS

Blockchain; Internet of Things (IoT); blockchain based cognitive computing; Hybridized Data Driven Cognitive Computing (HD2C); Federated Learning (FL); Proof of Authority (PoA)

1 Introduction

The term "cloud storage" refers to storing data digitally on servers in different physical locations. Information within the network is organized, administered, and protected by a third-party provider.

Data saved on the provider's servers are always guaranteed to be available, whether accessed through a public or private network. The ability to store, retrieve, and manage data on the cloud allows businesses to switch from a capital expenditure strategy to an operational expenditure model regarding data storage costs. In addition, cloud storage is adaptable, allowing businesses to extend or contract their data presence as needed.

Cloud storage varies from traditional storage in the following ways. Cloud storage refers to storing digital information across a network of remote servers. Archiving data in the form of digital files on physical media such as Hard Disk Drives (HDDs), Solid-State Drives (SSDs), or external storage devices is what is known as "traditional storage." Security, consistency, and efficiency are the three main features of cloud storage. In light of cloud provider's ability to deliver assets upon request through the Internet, the Information Technology sector is moving away from traditional storage and towards online storage [1].

Cloud storage incorporates security event tracking at your fingertips, or you may utilize suggestion tools that employ machine learning to discover over-provisioned accounts and inform you proactively, or you can encrypt all data. It may also combine security visibility across the entire infrastructure into a user-friendly management panel [2]. The Industrial Internet of Things (IIoT) refers to the use of IoT in industrial settings, most frequently for remote monitoring and control of sensors and machines via the cloud [3]. Using a Machine-to-Machine (M2M) connection, industries can utilize wireless automation and control. Cloud computing and related technologies (like data analytics and machine learning) are on the rise, which means that sectors could experience a higher level of automation, leading to new sources of income and corporate structures [4]. Hence, the IIoT is called Industry 4.0 automation or the fourth industrial revolution [5].

To be truly innovative, blockchain technology must provide the following: a distributed, cryptographically-based shared organization; timestamped records of all events; a detailed description of consensus mechanisms; and sufficient transparency and verification of stored information. Data security and economy in the IIoT industry may be improved by combining blockchain and Artificial Intelligence (AI) [6]. In contrast to Artificial Intelligence (AI) systems, Cognitive computing learns by observing patterns and advises human behaviour based on its knowledge. This method is heading toward more informed decision-making and data-driven intelligent manufacturing due to the proliferation of AI and Machine Learning (ML) tools [7]. Poisoning assaults, speed, and a lack of data resources are a few recent problems that still need to be fixed [8]. Researchers can learn more about how cognitive computing works [9] thanks to big data acquired from various sources. On the contrary, privacy issues arise when analyzing data that may be linked to sensitive user information in the cognitive computing system [10]. Industry 4.0 automation's reliability, attack resistance, and incentive systems may be considerably improved by blockchain integration with FL. It provides a cutting-edge framework for adapting the FL concept to big data-driven cognitive computing to enhance the productivity of Industry 4.0 production [11]. The provided architecture has improved the performance and privacy issues associated with cognitive computing.

This research suggests a paradigm for Industry 4.0 models called Hybridized Data Driven Cognitive Computing (HD2C) [12]. HD2C is the fusion of FL and Blockchain. A consensus mechanism is any protocol for establishing and maintaining confidence, agreement, and security in a distributed computer system [13]. The consensus mechanism is vital to every Blockchain network because it ensures the integrity and verifiability of the entire system. To agree on the shared ledger's current state, nodes in a Blockchain network employ a consensus method. Consensus algorithms are one of the most important innovations of blockchain technology. Many different protocols, such as Proof of Stake

(PoS), Proof of Authority (PoA), and Practical Byzantine Fault Tolerant (PBFT), were developed by the blockchain community based on this original idea. These are designed to promote consensus in a decentralized system, making Blockchain crucial [14]. PoA, which uses a system where nodes in the network are assigned the privilege of manufacturing new blocks for the chain via a round-robin or another arbitrary scheme, seems like a good bet [15]. To improve accuracy and resilience against poisoning attacks, an optimization model employs a modified Markovian decision process.

This work's primary contribution may be summed up as follows:

- Blockchain-based HD2C paradigm increases data security and resistance against poisoning assaults in the IIoT sector.
- The Consensus Algorithm Proof-of-Authority (PoA) is fundamental to Blockchain's ability to strengthen the security and integrity of stored data.
- Results reveal that the proposed method can find the best optimization model for Industry 4.0 automation.

The paper is structured as follows: Section 2 discusses the related works of the proposed model; the proposed blockchain-enabled federated learning and HD2C with PoA consensus mechanism are investigated in Section 3 and its sub-sections; the result obtained by the proposed model and its analysis are represented in Section 4; finally, overall summarization of the proposed model is discussed in Section 5.

2 Related Works

The COGNITWIN (Cognitive Digital Twins) project framework, described by [16], is a four-stage Big Data and artificial intelligence pipeline designed to enable digital twins. Hybrid digital twins consist of data-driven digital twins and first-order physical models.

To address the need for de facto standards like fast healthcare interoperable resources (FHIR) to facilitate the meaningful exchange of healthcare records among all the participants, reference [17] presented a blockchain with proof of authorization for sharing sensitive medical records securely. Data is exclusively governed within the nodes, and communication with end users is authorized via smart contracts and PoA.

In [18], Zhang et al. proposed a blockchain-based cognitive cloud computing (BC-CCC) architecture. The Internet of Things (IoT)'s most potent edge computing systems are used by artificial intelligence (AI) to uncover information retrieved from a massive amount of sensory input, such as cyber impacts. BC-CCC devised numerous strategies for incorporating individual traits into cryptographic and security frameworks.

Connected automated cars (CAVs) in the Cognitive Internet of cars (CIoVs) can better plan their routes and decrease congestion by using real-time cognitive perception and prediction of traffic conditions, as stated by [19].

Reference [20] proposed the cognitive systems of Blockchain Technology, Cognitive Computing, and Healthcare Innovations to provide security for carrying out any trade of products or services. This study surveyed Blockchain technology innovation in fields other than finance (i.e., fields unrelated to Bitcoin). In the final section of the study, the researchers weighed the pros and cons of applying blockchain technology in two intra-horizons: healthcare and cognitive computing.

Cloud-IIoT-Based Electronic Health Record Privacy Preserving through a convolutional neural network (CNN) was proposed by [21] to increase data security and Privacy through Blockchain-based Federated learning and avoid Critical concerns in the healthcare sector. This method uses Blockchain and deep learning to protect the privacy of electronic health records.

With potential and issues, Federated learning meets Blockchain in edge computing was proposed by [22]. By coordinating several mobile devices to train a general AI model without releasing their data and considerably improving Privacy, Federated Learning (FL) has been proposed as a solution for shared data training. However, when a slow node improperly uses the ML models of other clients without proper training, the FL chain encounters a plagiarism problem during the block verification procedure.

In [23], a Blockchain-enabled federated learning data protection aggregation approach is introduced for the IIoT that uses differential Privacy and homomorphic encryption. This study employed distributed K-means clustering with differential Privacy and homomorphic encryption, distributed random forest with differential Privacy.

Qu et al. [24] proposed decentralized Privacy in fog computing via blockchain-enabled federated learning, and analyzed the efficacy, security, and scalability of FL-Block (federated learning enabled by Blockchain). The proposed FL-Block strategy is meant to close the gap and solve the abovementioned issues. Unfortunately, attacks on the FL smart contract typically cause significant financial losses.

In engineering education, Problem-Based Learning (PBL) is an effective paradigm for cybersecurity training. It integrates lecture-based instruction with laboratory tasks to improve problem-solving abilities and prepare students for real-world issues. A Knowledge Graph-guided online laboratory environment improves learning results, and cybersecurity awareness, and promotes continuous development [25].

Reference [26] proposed a privacy-preserving Distributed Application (DA) that utilizes blockchain technology for managing secure healthcare certificates. The system interfaces with healthcare centres, verifiers, and authorities through smart contracts, aiming to enhance security and overcome issues like collusion and phishing. Experimental tests using Etherscan reveal improved efficiency compared to existing techniques.

Reference [27] presented a decentralized blockchain-based cab-sharing system that addresses issues like congestion and privacy concerns. It replaces central authorities with a reputation system on Ethereum's platform, ensuring driver and rider information remains private. The system's use of smart contracts maintains efficiency and security, offering a practical cab-sharing solution.

The objective of this study [28] is to develop a virtual educational environment for teaching algorithms, big data processing, and machine learning. It makes use of Unity's Visual IoT/Robotics Programming Language Environment (VIPLE) and a traffic simulator. The study focuses on developing realistic traffic data and implementing dynamic routing algorithms using VIPLE, with actual traffic data from Arizona's Maricopa County being used to improve simulation accuracy.

The proposed suite of security measures addresses current and future security issues, including a path-aware network architecture, source authentication, bandwidth reservation, and privacy-focused forwarding [29].

Data security in a cloud environment requires multilevel cyber security analysis and additional processing mechanisms in the IIoT sector. Cyber security is achieved by implementing a blockchain-enabled FL in this research. This cognitive computing is constructed by HD2C, which is explained in the following section. The comparison between the existing methods is shown in Table 1.

Table 1: Comparison of existing methods

Reference	Methodology used	Results	Limitations
[6]	AI benchmark suite for assessing the performance of DCE platforms in Machine Learning (ML) and cognitive science applications	The suggested method assists those who build computer systems and create AI applications for clouds, edge devices, and mobile devices that are backed by 5G mobile networks and AIoT resources.	The decision-making and data-driven intelligent manufacturing are not informative due to the explosion of AI and ML tools.
[17]	A blockchain with proof of authorization is presented for sharing sensitive medical records securely	The proposed blockchain model produces more security and intelligence because of smart contracts.	Low throughput and scalability, and limited development of blockchain in healthcare data sharing are some of the limitations of the proposed model.
[18]	A Blockchain-Based Cognitive Cloud Computing (BC-CCC) architecture is proposed	The proposed methodology improves the data transmission rate, security ratio, security, throughput and data trading ratio.	The implementation cost of the BC-CCC methods is very high.
[21]	Cloud-IIoT-based electronic health record privacy preserving through CNN	The experimental findings reveal that the model's classification and performance outperform other current strategies.	The proposed model removes the database along with the accessibility of the health records.
[24]	Decentralized privacy in fog computing is proposed through blockchain-enabled federated learning	The proposed FL-Block strategy solves the network congestion, latency, and a lack of local control problems effectively.	Attacks on the FL smart contract typically cause significant financial losses.

(Continued)

Table 1 (continued)

Reference	Methodology used	Results	Limitations
[26]	Privacy-preserving Distributed Application (DA)	Experimental tests using Etherscan reveal improved efficiency compared to existing techniques.	Healthcare schemes suffer from various security attacks like collusion, phishing, and masquerade.

3 Blockchain-Enabled Federated Learning

Blockchain is utilized as a foundational framework to increase a variety of performance indicators, such as speed and precision, and federated learning is used to build cognitive computing. PoA is utilized in Blockchain as a consensus algorithm since it requires less time and energy than PoW or PoS. It can verify transactions at a faster rate. With PoA, you get the most significant features of both PoS and PoW in one convenient package; it is faster than PoW and more secure.

Nevertheless, blockchain-enabled FL has captured the interest of academics and the industry to hurry up the spread of FL. Researchers present the paradigm of Hybridized Data-Driven Cognitive Computing (D2C) in Industry IoT using blockchain-enabled federated learning (Fig. 1). Though privacy concerns have emerged from consumers or device edges, cognitive computing necessitates data collecting from several IIoT. To avoid sharing unprocessed information, federated learning is preferred, in which only the model is shared. Decentralized federated learning can be achieved using FL and the PoA consensus mechanism provided by blockchain technology.

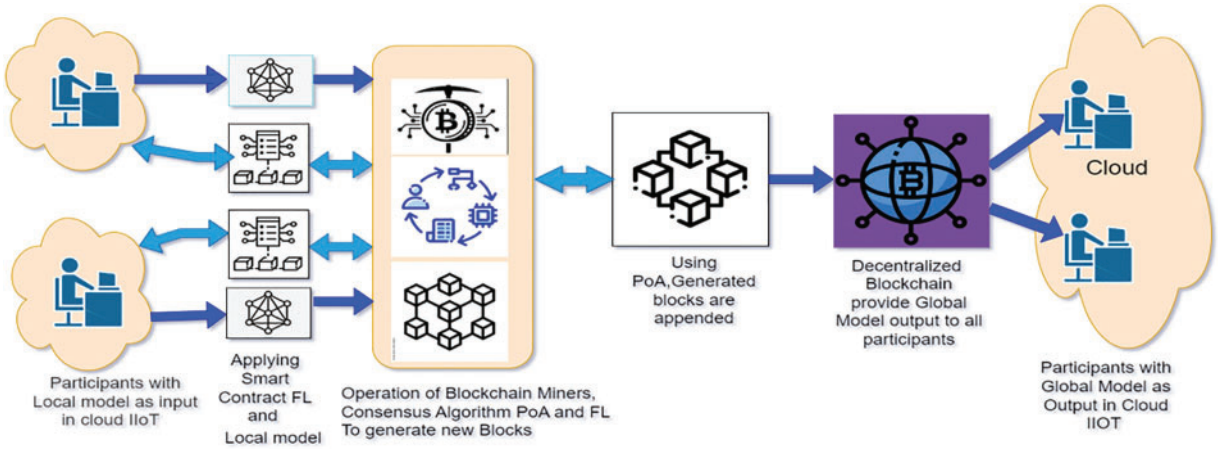

Figure 1: Proposed block diagram of hybridized data-driven cognitive computing

Poisoning-attack-proofing is made possible by decentralization, and it additionally offers incentives and an additional accuracy guarantee through a member choosing the network. As shown in Fig. 1, participants with a local model in cloud IIoT is the collection of input data from the different IIoT sector. When specific criteria are met, agreements can be automatically carried out in a decentralized setting using the smart contract FL. In its simplest form, a smart contract is a piece of computer code designed to automatically carry out the conditions of a contract or agreement without

human interaction. Once a transaction is finalized, it cannot be undone by anybody else. This is done to maintain the network's security, dependability, and immutability. The output of the appended blocks from the blockchain miner is the global model, which has all information with a global identity. The secured and verified output is transferred to all the IIoT participants in the cloud platform, as shown in Fig. 1.

3.1 Data Security in the Cloud

Three fundamental pillars must be present for a cloud-based data security strategy to be successful:

• Identity: A key component of a successful cloud-centric data security policy recognizes the identities of individuals, devices, and apps during the creation, modification, shopping, application, exchange, and ultimately destruction of data.

• Access Boundaries: Limiting who may access the data is the second essential premise for safeguarding data in the cloud. Your data security plan with a cloud focus should use identity to regulate access via policy using several cloud-based services and tools that are already accessible.

• Visibility: If data gates are in place, use the cloud's robust visibility capabilities to audit usage and deliver compliance reports showing how data is managed and accessed by ad hoc administrators. These visibility technologies permit focused response activities by quickly spotting risks and irregularities. On top of these pillars, the program is developed, and other controls can be implemented to achieve the best comprehensiveness for your company.

Fig. 2 shows how cloud storage is set up to protect user information. Data integrity is a crucial part of any computerized system. Data integrity protects data from illegal destruction, modification, or fabrication. In addition, the management entity has access to and rights over some of the company's resources, which helps to safeguard such resources from misuse, misappropriation, and theft. Ensuring data integrity is simple in a standalone system with a single database.

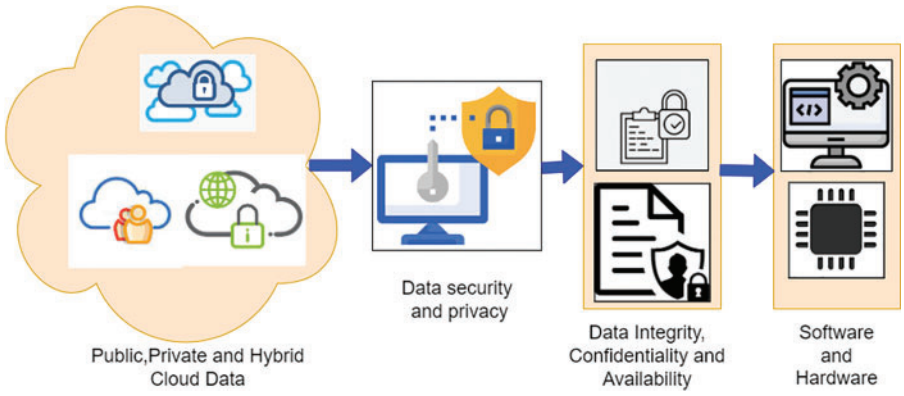

Figure 2: Cloud data privacy and security structure

Consumers must know that their sensitive information will be safe in the cloud. Authentication and other forms of access control ensure that only authorized parties can view sensitive information. Improving cloud dependability and trustworthiness could help put fears about exposed data, compromised authentication, and unauthorized access to rest. Encryption is frequently used to safeguard the confidentiality of data.

In their proposal, Rivest et al. offered homomorphic encryption as a security means. It is unnecessary to decrypt the data during the procedure if the results of the algebraic operation on the cipher text are compatible with the clear operation. This plan may end the need for privacy protections around cloud-based information and actions. Data privacy and security can be guaranteed using a proposed hybrid method that utilizes key-sharing and authentication procedures. A more secure connection between the user and the cloud service provider is possible by implementing strong key-sharing and authentication methods. Using the Rivest–Shamir–Adleman (RSA) public key algorithm [30], the user and cloud service providers can safely exchange keys.

The term "data availability" refers to how easily and quickly a user's data can be used or retrieved in the event of a disaster, such as a damaged hard drive, an Internet Data Center (IDC) fire, or a network outage, and how the user can verify the integrity of their data without having to rely solely on the cloud service provider's credit guarantee. Finally, to have Privacy is to conceal one's identity and other personal information from others, with limited exceptions. Here are the several aspects of personal space:

∗ Under what conditions may someone worry more about disclosing information about the here and now or the future than about information about the past?

∗ Users may feel more comfortable if their friends can proactively seek them out for information, but they may not enjoy receiving constant, unprompted updates.

∗ Users may prefer to have their data reported as an amorphous region (or "extent") rather than a pinpointed place. The trade community is obligated to protect its customers' Privacy and context. To protect individuals' Privacy, businesses must have policies, safeguards, and practices for managing sensitive data. Users risk having their data misused by other users. Users commonly share the same information saved just once due to the widespread use of deduplication technology in cloud storage. While this helps cloud providers save money and storage space, it does not protect users' data from attackers who know the file's hash code. As a result, the cloud storage service may leak confidential data. Therefore, it has been proposed that cloud users' identities be verified via a proof of ownership mechanism.

Hackers may cause cloud services to become more expensive. Consuming resources without paying for them is a form of cloud service payment fraud. Threat actors can use this information to increase the price of cloud services. Participants' data is kept safe in this case, and attacks are prevented using the suggested HD2C.

3.2 HD2C-Federated Learning and Blockchain to Achieve Data Confidentiality and Privacy

A machine learning method called federated learning enables data models to learn from several data sets hosted in various locations (such as regional data centres or a central server) without sharing the training information. However, in today's world of numerous applications, collecting all the data in one place is practically impossible. Obtaining data at a single location also adds the added cost of disseminating the data across secure channels during such precarious periods. However modern artificial intelligence is moving in the direction of a decentralized strategy. The most recent AI models are being trained collaboratively on the edge or at the source through data from smartphones, laptops, private servers, and other devices. As a result, Federated learning, an advanced type of Artificial Intelligence (AI) training, is quickly becoming the norm for complying with new rules for collecting and preserving personal data.

FL has modified many applications by delivering distributed AI solutions at the network's edge utilizing a wide range of IIoT devices. FL makes Distributed learning possible because it allows Artificial Intelligence (AI) functions like AI model training to be sent directly to the devices rather than to a centralized server in the cloud. This allows the devices to contribute to the creation of a shared global model at an aggregator, such as a MEC (Mobile Edge Computing) server at a Base Station (BS) or an Access Point (AP). Several intelligent edge services, such as those in the transportation, healthcare, and automation sectors, find FL appealing because of recent advancements in mobile technology and growing user privacy concerns. Due to the distributed nature of IoT devices, a BS cannot collect all of the data required for AI/ML training, making FL critical for the next generation of cognitive edge networks to acquire full intelligence. With FL, devices and the BS can train a shared model utilizing users' local raw datasets. In the FL operation, M represents the total number of IoT devices. Each m device contributes to the M-device AI model through its unique dataset Dm $\in M$. Each node in the network trains its model during each communication cycle, arriving at an update w_m by minimizing a loss function F (w_m), as shown in Eq. (1).

$$w_m^* = arg\ minF(w_m), m \in M \tag{1}$$

In this case, the loss function used by various FL algorithms may vary. In the case of linear regression FL modelling, the loss function F can be computed as $F(W_m) = \frac{1}{2}(ai^T W_m - b_i)$ 2 where ai and bi are input and output pairs, M i = 1 is the number of inputs, and m is the number of outputs. The MEC server then takes the aggregated and computed updates w_m from each device m and generates a new global model as

$$W_G = 1 \Big/ \sum_{m \in M} |D_m| \sum_{m=1}^{M} |D_m|\ W_m \tag{2}$$

Once the next training cycle has begun, this global model is downloaded to all devices to finish the global learning process. It should be noted that the MEC server computers and updates the global model W_G as given in Eq. (2) in the traditional FL. On the other hand, the FL chain uses blockchain technology to enable decentralized computation of the global model locally on the devices themselves.

3.3 Blockchain to Achieve Reliability of Data

Blockchain is employed as a core framework to improve a range of outcome measures, including speed and precision, and federated learning is used to construct cognitive computing. Blockchain primarily functions as a peer-to-peer (P2P) networked public, trustworthy, and shared ledger. Decentralization, or the idea that a single institution does not control data on a blockchain, is the main tenet of the blockchain concept. However, blockchains are maintained using consensus mechanisms, which means that each node in a blockchain, such as IIoT devices and MEC servers located on edge networks, has equal authority to control and monitor the information contained in blockchains. Due to its decentralized design, blockchain technology is invulnerable to hacking and other data abuse. In addition, blockchain solutions are more reliable and resistant since there is no single point of failure, the central server. The proposed Blockchain-based transaction mechanism with FL is broken down into three phases, as shown in Fig. 3.

∗ In the initial stage, Industrial users use the private and public keys associated with their account to create an action with metadata (such as the user ID), a user signature, and a timestamp. Complex computational tasks, such as mining and the execution of smart contracts FL, may be planned and conducted in the cloud due to the limited computer capability of IIoT devices. The user then

issues a command to the cloud to perform some action, such as handling Internet of Things data, authenticating users, or providing a service.

∗ Second, the cloud storage service fulfils the request by making its available resources accessible to the user. Trade (i.e., the exchange of assets between users and service providers), payment authentication, user verification, and purchasing can all be carried out automatically using smart contracts. In addition, cloud-based virtual machines powered by mining blocks validate user transactions. All miners can now choose this one from the list of pending transactions. A network of digital miners might utilize a PoA-like consensus method.

∗ The third stage involves the miner who validated the data block quickly sending the block's signature to the other miners so they may check it. When a consensus is reached among miners, a newly validated block and its associated signature are added to the Blockchain. Once all nodes in the network have received this block, the Blockchain's global copy has been synchronized with all industrial users.

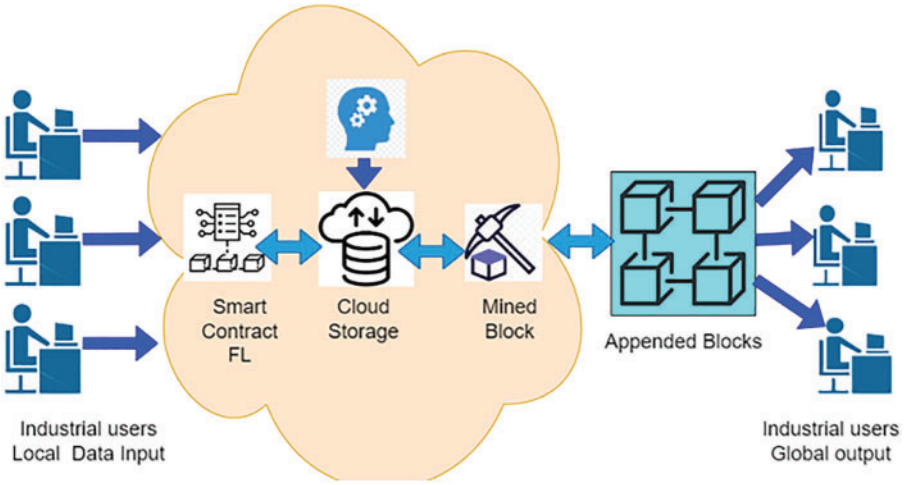

Figure 3: Proposed blockchain-based transaction method using FL

3.4 Cognitive Learning on FL to Achieve the Data Privacy

IIoT devices, mobile devices, and gaming produce enormous amounts of data. Models may be created based on cognitive computing that stores the data initially and then does local training. A machine learning technique known as FL trains the parameters of local models. Then it aggregates those models to create a global model independent of the original data. In many cases, local and global models have the same architecture. However, there could be a lot of circumstances when various local models with various computational difficulties exist. A unique combined learning system called FL is employed to equip the very varied computations and their communication capabilities to satisfy the needs of heterogeneous clients of IIoT devices. Local data from heterogeneous IIoT devices, such as $\{a_1,\ldots,$ and $a_n\}$ are used to train global models. Local model parameters are denoted by the letters $\{y_1,\ldots, y_n\}$. To determine the global parameter yg, the local parameters are averaged in the model. Iterations of this process are performed, and the local parameters of the $(i+1)^{th}$ iteration get the y_g calculation from the previous iteration's procedure. The network size may be controlled by altering the network's breadth for effective cognitive learning to occur. As a result, this can help reduce local parameters while preserving the same model class for the global parameters' underlying design. In addition, this

improves the consistency of the global model as a whole. When determining the global parameters for FL, this paper considers the input channel size (i_g), the output channel size (o_g), and the computational complexity (c). For the submerged layer, the shrinking ratio is crucial.

Equations for the output channel shrinkage ratio are written as shown in the upcoming Eq. (3):

$$s1 = \left(ol^{c+1}|og\right)^{1/c} \tag{3}$$

The following Eq. (4) illustrates the formula for the input channel shrinkage ratio:

$$s2 = \left(il^{c+1}|ig\right)^{1/c} \tag{4}$$

Let's simplify and say that s1 = s2 = s. Then, the following Eq. (5) mentions the local model parameter shrinkage ratio:

$$SR = yl^{c} = yg * s^{2(c-1)} \tag{5}$$

In the foundation of allotted subsets, global model parameters may be built by the local model parameter's computed capability. The computation of the global parameter frequently uses the set difference concept. All participants' parameters are a part of the combined parameter matrix. As a result, models of intermediate complexity contain parameters that are fully aggregated with larger models while being partially aggregated with smaller models. Additionally, smaller local models that can help the global model are more likely to be aggregated. The data in the global model is used to generate the learning aspect of cognitive learning. The primary goals of Algorithms 1 and 2 are to broadcast participant updates and use FL for cognitive learning applications.

3.5 Combining Cognitive Learned Data with Blockchain

Effective cognitive learning data storage relies heavily on blockchain technology. Since these records will be used to make decisions based on them, protecting their anonymity is paramount. The contributing elements of security and Privacy include traits like blockchain-like immutability, tamper-resistance, decentralization, pseudonymous identification, etc. This section discusses blockchain technology in the context of a global information model. Here, Blockchain protects the confidentiality of the globally significant model parameters by embedding the acquired knowledge into the blocks. Here, Blockchain protects the confidentiality of the globally significant model parameters by embedding the acquired knowledge into the blocks. Our proposed blockchain model is structurally indistinguishable from the original Blockchain. The function of Intelligence data secured by the Blockchain is described in Fig. 4.

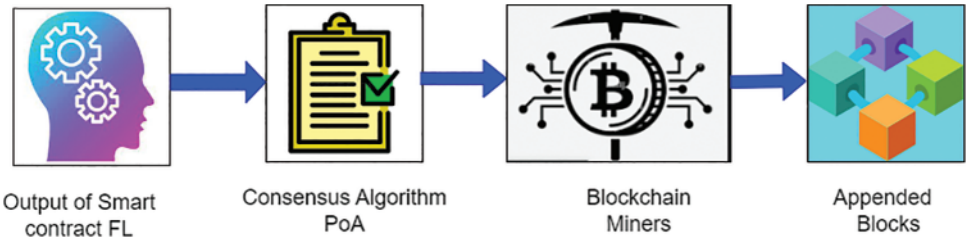

Figure 4: Intelligence data secured by the blockchain

Each block's output from the smart contract includes a data field with newly learned information, a previous block hash, a current block, a timestamp, and a parameter. The prior hash field in the

genesis block (the first block) is all zeros. Each block contains a cryptographic hash of the one before it, creating an unbreakable chain. A small change in any field can result in a different hash for the entire block.

Our approach uses the PoA consensus methodology to ensure that each block is adequately checked and validated before being added to the distributed Blockchain. When they meet the restrictions of the target nonce, the miners keep trying to generate the random parameter.

Once a miner has the required parameter, they can announce the block to the distributed Blockchain as a new block. Each miner's Blockchain will be updated to include the new block, creating a shared ledger.

Due to its append-only nature, Blockchain can only have data added to it in chronological order. This feature shows that information stored on a blockchain is nearly impossible to alter after it has been recorded.

Algorithm 1: Cognitive Learning FL Algorithm

Variables: y1, yg, s, p, m
Data generated by IoT devices and Local participants are used as procedure input.
Start by initializing the global model parameters and local participants with the data and ***wg***.
T=0, 1, 2, ... perform After each data production cycle.
S–a random collection of active customers
$k \in S$ any participants
Calculate complexity level **"C"** using local data.
Calculate the hidden shrinkage ratio (SR), the input channel linkage ratio (s2), and the output channel shrinkage ratio (s1).
End C
do for all complexity levels
Utilizing the aggregation of local model parameters, compute the global model parameter
End
End
End

Algorithm 2: Blockchain-based security for stored cognitive learning data.

T, l, and ***n*** are all observable variables.
yk, Xk as input
Start with: ***Bk***-Local data ***Xk*** is divided up into batches of varying sizes, ***T***
After the allotted time has passed
For various batch $bk \in T$ do,
For all participants, $k \in S$
$\mathbf{L_k} \leftarrow \mathbf{n}\Delta l\ (yl, bk)$
Yk ← yk-lk
channel linkage ratio(S2) and the SR
End
End

4 Result and Analysis

This section used the performance of the proposed HD2C framework with cloud platform service provider, which supports many types of 1000 nodes, for evaluation. Some nodes are special-purpose vehicles, while others are complete nodes. A Linux Virtual Machine is installed on each node. Learned data privacy is assessed in a testing setting. There were five sets of tests, with the number of participants in each set corresponding to the difficulty level. To communicate with other IoT devices and offer secure signatures, every IoT device is issued a public and private key.

This study contributes to the Edge-IIoT set, a novel, sizable, and realistic cyber security dataset of Internet of Things (IoT) and Industrial Internet of Things (IIoT) applications, which machine learning-based intrusion detection systems can use in centralized or federated learning modes. This data set's summary lists 52 files with 1638 distinct columns. Edge computing, cloud computing, software-defined networking, a blockchain network, fog computing, and an IoT/IIoT perception layer are the seven components. Following the preparation and analysis of the realistic cyber security dataset [31], this paper provides an introductory exploratory data analysis and assesses machine learning algorithms' performance in centralized and federated learning settings. A 16-byte ID is used to distinguish each framework-connected IoT device. The FL method is utilized for effective cognitive learning. The FL method streamlines the process by cutting down on the amount of computation and data transfer required. As a result, fewer local models are trained compared to the global model. Blockchain computations are performed on the private Ethereum platform. This vers00ion employs a Core i7-8565U CPU (1.8 GHz, 1992 MHz, 4 Cores, 8 Logical Processors). Our proposed system's performance was analyzed for the following characteristics [32].

4.1 Memory Utilization

Memory utilization is mainly determined by the amount of the fundamental information of the block, omitting transaction data, as well as the size of the transactions. To calculate Memory Utilization here, the inputs are taken as appended output blocks from the blockchain miners with different numbers of transactions per block considered. Memory usage is analyzed for three distinct block sizes (10 transactions, 20 transactions, and 30 transactions). The blocks store the results of the global models. This number is then used to calculate how many transactions should be included in the block. Fig. 5 shows that when the number of transactions per block increases, HD2C's memory usage decreases. According to the experiment results, the memory requirements are mostly unaffected by the number of transactions per second.

Accuracy: To fine-tune the magnitude of the steps taken during each iteration toward minimizing the loss function, the learning rate is a tuning parameter in optimization methods used in machine learning and statistics. To calculate accuracy, input is the data sample size to the account number of appended blocks. Additionally, the performance is assessed at three different learning rates (η), namely 0.004, 0.04, and 0.4. The learning rate is a tuning parameter in an optimization method that controls the size of each iterative step taken toward the minimum of a loss function in cognitive learning and statistics. As shown in Fig. 6, with the increased value of the learning rate, the accuracy has been increased. Compared to the η value of 0.004 and 0.04, the η value of 0.4 attain good accuracy as shown Fig. 6. The HD2C accuracy score is a metric that evaluates models based on how many predictions they get right out of the total number of predictions they make. As illustrated in Fig. 6, good accuracy is also seen for high data sample sizes. Rapid learning results in maximum accuracy, yet the graph shows that linear learning occurs first and then becomes constant in all three instances. A similar

graph is anticipated in the event of bigger data sample sizes (in the thousands). In comparison to current solutions, this also ensures excellent scalability.

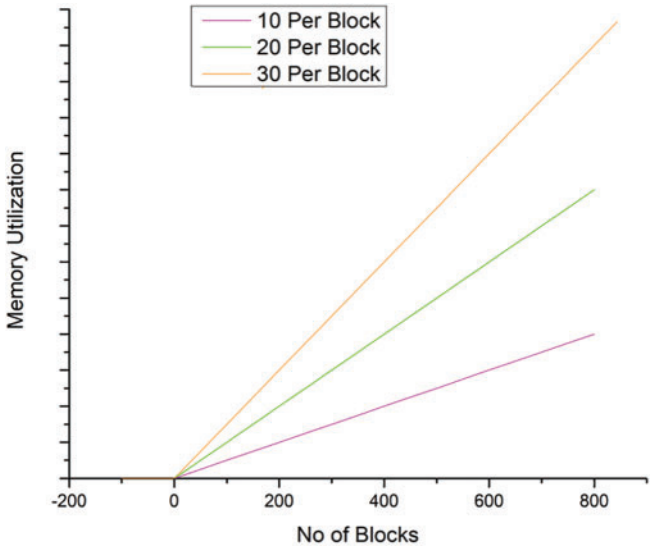

Figure 5: Memory utilization with number of appended blocks in HD2C

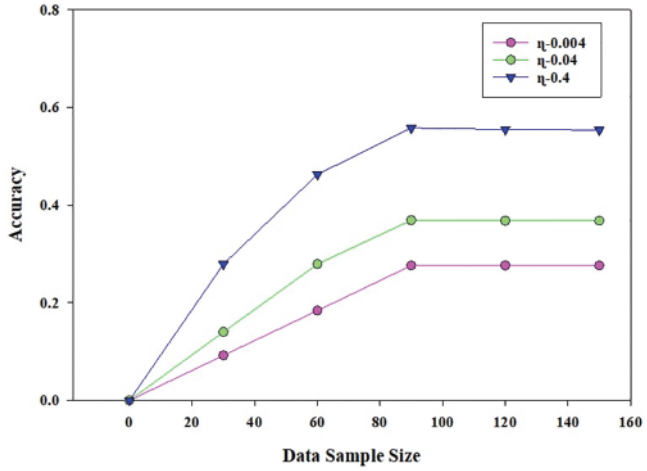

Figure 6: Data sample size's effect on HD2C accuracy over a range of learning rates

Latency: The time elapsed between a stimulus and a response or reaction is known as latency. Shorter response times can result in faster brain processing or improved memory in cognitive methods.

The network calculates it.

Duration = Serialization Delay + Propagation Delay

Delay in Propagation = Distance/Speed

Packet Size (Bits)/Transmission Rate (bps) = Serialization Delay

A thread, usually referred to as the execution or control thread, is a single sequential flow of operations that is carried out during a process. Now, every Operating System process can execute

a thread. Here proposed system consists of four threads to access output blocks immediately. In addition, a process may include several threads. The time to produce blocks with 10, 20, and 30 transactions per block utilizing two threads is shown in Fig. 7. Fig. 8 shows the processing time for four threads for generating blocks with 10, 20, and 30 transactions per block. Fig. 9 shows the amount of time it takes to run eight threads, each of which generates a block with a different number of transactions in it: 10, 20, and 30 more blocks are added and more transactions are processed in each block, execution time increases, but execution time per thread decreases. Fig. 8 shows little execution time, which is to be expected given that our system has four cores and four threads. Using this metric, the time it takes to execute a block varies depending on its environment. Less time spent in execution thanks to instantaneous block updates in a blockchain will lead to a faster network. By doing so, the system's ledger will be more in sync with its most recent learnings.

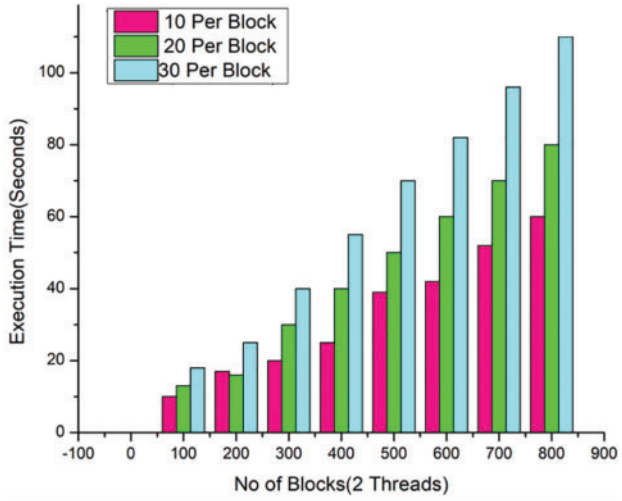

Figure 7: Latency with 2 threads is compared with HD2C

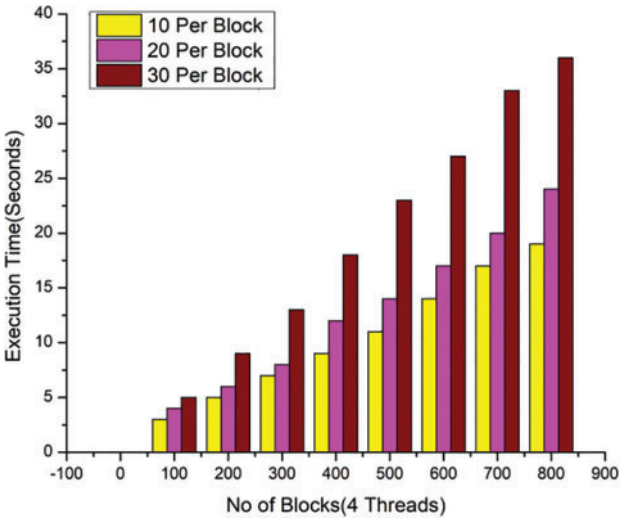

Figure 8: Latency with 4 threads is used in HD2C

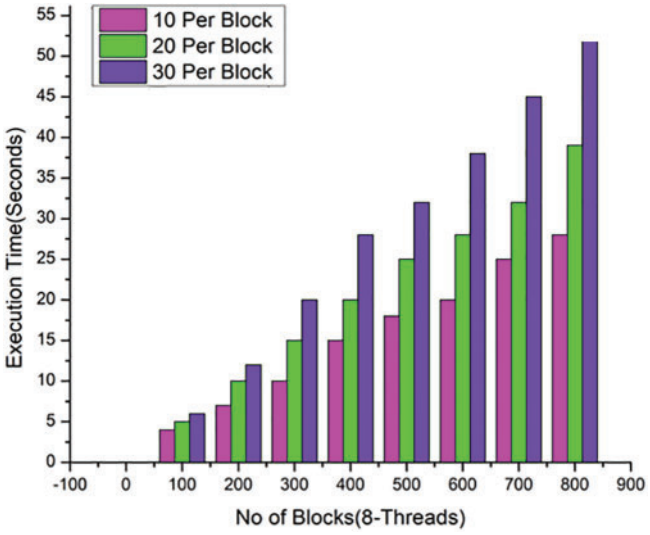

Figure 9: Latency with 8 threads is compared with HD2C

4.2 Evaluation of Block Rate and Signal-to-Noise Ratio

The time miners or other network validators take to verify the transactions in a block and generate a new block on a blockchain is measured in units called "blocks per second" or "block rate." Figs. 9 and 10 display the effect of varying the Signal-to-Noise ratio on the time it takes for HD2C to converge. The signal-to-noise ratio (SNR) is frequently used; a higher SNR value implies a better-quality signal. The signal-to-noise ratio is often expressed in decibels and can be computed with a logarithm of base 10. However, how Signal and noise levels are measured will determine the exact formula.

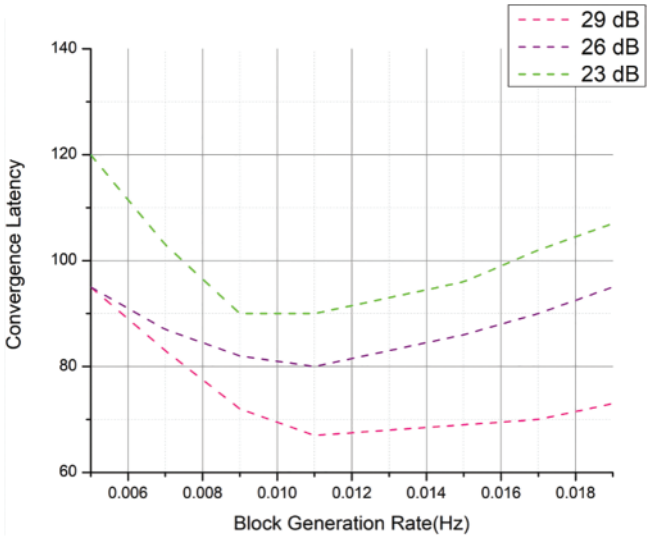

Figure 10: Block generation rate with the convergence latency in HD2C

For instance, if microvolts are used as the unit of measurement, the formula would be:

$$20\ log_{10}(\text{Ps/Pn}) = \text{S/N}$$

The Signal is denoted by Ps, where Ps is the microvolt level and the noise by Pn. The miner's output takes the signal strength as appended block strength with noise affected in output. As seen in Fig. 10, the latency decreases as the signal-to-noise ratio rises, demonstrating that HD2C is more effective than previous techniques.

5 Conclusion

The blockchain-based HD2C technique has been employed to build and implement the suggested framework. This study provides a valuable foundation for enhancing the IIoT infrastructure's performance in a complex environment. This framework is suitable for guaranteeing communication security in the future when large amounts of data are transferred in a diverse environment. FL Idea solves issues with efficiency and Privacy. FL requires less iteration to achieve the best results, whereas Blockchain safeguards the Privacy of learned information. When checking the legitimacy of a block, the PoA consensus protocol is applied. The accuracy and memory utilization evaluation results predict the system's total benefits. The findings of this study created a brand-new IIoT framework using blockchain technology. The Blockchain PoA's foundational cost is significant compared to other consensus algorithms.

Acknowledgement: The author wishes to express heartfelt thanks to the methodologies that have made a substantial contribution to this research.

Funding Statement: No funds were received for this research.

Author Contributions: The authors confirm their contribution to the paper as follows: study conception and design; data collection: Xiangmin Guo; analysis and interpretation of results: Guangjun Liang, Xianyi Chen; draft manuscript preparation: Jiayin Liu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] A. Ganne, "Cloud data security methods: Kubernetes *vs*. Docker Swarm," *International Research Journal of Modernization in Engineering Technology*, vol. 4, no. 11, pp. 1–6, 2022.

[2] S. Achar, H. Patel and S. Hussain, "Data security in cloud: A review," *Asian Journal of Advances in Research*, vol. 17, no. 4, pp. 76–83, 2022.

[3] M. A. Rahman, M. M. Rashid, M. S. Hossain, E. Hassanain, M. F. Alhamid *et al.,* "Blockchain and IoT-based cognitive edge framework for sharing economy services in a smart city," *IEEE Access*, vol. 7, pp. 18611–18621, 2019.

[4] S. R. Pokhrel, Y. Qu and L. Gao, "QoS-aware personalized privacy with multipath TCP for industrial IoT: Analysis and design," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 4849–4861, 2020.

[5] J. Wan, J. Yang, Z. Wang and Q. Hua, "Artificial intelligence for cloud-assisted smart factory," *IEEE Access*, vol. 6, pp. 55419–55430, 2018.

[6] T. Hao, K. Hwang, J. Zhan, Y. Li and Y. Cao, "Scenario-based AI Benchmark evaluation of distributed Cloud/Edge computing systems," *IEEE Transactions on Computers*, vol. 72, no. 3, pp. 719–731, 2023.

[7] S. Liao, J. Wu, J. Li, A. K. Bashir, S. Mumtaz *et al.,* "Cognitive popularity-based AI service sharing for software-defined information-centric networks," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 2126–2136, 2020.
[8] J. Wang, J. Luo, X. Liu, Y. Li, S. Liu *et al.,* "Improved Kalman filter based differentially private streaming data release in cognitive computing," *Future Generation Computer Systems*, vol. 98, pp. 541–549, 2019.
[9] M. Grissa, A. A. Yavuz and B. Hamdaoui, "Location privacy preservation in database-driven wireless cognitive networks through encrypted probabilistic data structures," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 2, pp. 255–266, 2017.
[10] Y. Qu, S. Yu, W. Zhou, S. Peng, G. Wang *et al.,* "Privacy of things: Emerging challenges and opportunities in wireless Internet of Things," *IEEE Wireless Communications*, vol. 25, no. 6, pp. 91–97, 2018.
[11] S. Gupta, A. K. Kar, A. Baabdullah and W. A. Al-Khowaiter, "Big data with cognitive computing: A review for the future," *International Journal of Information Management*, vol. 42, pp. 78–89, 2018.
[12] X. Huang, "A data-driven WSN security threat analysis model based on cognitive computing," *Journal of Sensors*, vol. 2022, pp. 1–10, 2022.
[13] M. Chen, W. Li, G. Fortino, Y. Hao, L. Hu *et al.,* "A dynamic service migration mechanism in edge cognitive computing," *ACM Transactions on Internet Technology*, vol. 19, no. 2, pp. 1–15, 2019.
[14] M. Coccoli, P. Maresca and L. Stanganelli, "The role of big data and cognitive computing in the learning process," *Journal of Visual Languages & Computing*, vol. 38, pp. 97–103, 2017.
[15] S. Wadhwa, S. Rani, G. Kaur, D. Koundal, A. Zaguia *et al.,* "HeteroFL blockchain approach-based security for cognitive Internet of Things," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–8, 2022.
[16] P. Unal, O. Albayrak, M. Jomâa and A. J. Berre, "Data-driven artificial intelligence and predictive analytics for the maintenance of industrial machinery with hybrid and cognitive digital twins," in *Technologies and Applications for Big Data Value*, Cham: Springer, pp. 299–319, 2022.
[17] N. A. Asad, M. T. Elahi, A. A. Hasan and M. A. Yousuf, "Permission-based blockchain with proof of authority for secured healthcare data sharing," in *2nd Int. Conf. on Advanced Information and Communication Technology (ICAICT)*, Dhaka, Bangladesh, pp. 35–40, 2020.
[18] H. Zhang, Z. Zang and B. Muthu, "Knowledge-based systems for blockchain-based cognitive cloud computing model for security purposes," *International Journal of Modeling, Simulation, and Scientific Computing*, vol. 13, no. 4, pp. 2241002, 2022.
[19] H. Chang, Y. Liu and Z. Sheng, "Blockchain-enabled online traffic congestion duration prediction in cognitive Internet of Vehicles," *IEEE Internet of Things Journal*, vol. 9, no. 24, pp. 25612–25625, 2022.
[20] J. Daniel, A. Sargolzaei, M. Abdelghani, S. Sargolzaei and B. Amaba, "Blockchain technology, cognitive computing, and healthcare innovations," *Journal of Advances in Information Technology*, vol. 8, no. 3, pp. 130–139, 2017.
[21] J. A. Alzubi, O. A. Alzubi, A. Singh and M. Ramachandran, "Cloud-IIoT-based electronic health record privacy-preserving by CNN and blockchain-enabled federated learning," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 1, pp. 1080–1087, 2022.
[22] D. C. Nguyen, M. Ding, Q. V. Pham, P. N. Pathirana, L. B. Le *et al.,* "Federated learning meets Blockchain in edge computing: Opportunities and challenges," *IEEE Internet of Things Journal*, vol. 8, no. 16, pp. 12806–12825, 2021.
[23] B. Jia, X. Zhang, J. Liu, Y. Zhang, K. Huang *et al.,* "Blockchain-enabled federated learning data protection aggregation scheme with differential privacy and homomorphic encryption in IIoT," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 6, pp. 4049–4058, 2021.
[24] Y. Qu, L. Gao, T. H. Luan, Y. Xiang, S. Yu *et al.,* "Decentralized privacy using blockchain-enabled federated learning in fog computing," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5171–5183, 2022.
[25] Y. Deng, Z. Zeng, K. Jha and D. Huang, "Problem-based cybersecurity lab with knowledge graph as guidance," *Journal of Artificial Intelligence and Technology*, vol. 2, no. 2, pp. 55–61, 2021.
[26] P. Sharma, S. Namasudra, N. Chilamkurti, B. G. Kim and R. Gonzalez Crespo, "Blockchain-based privacy preservation for IoT-enabled healthcare system," *ACM Transactions on Sensor Networks*, vol. 19, no. 3, pp. 1–17, 2023.

[27] S. Namasudra and P. Sharma, "Achieving a decentralized and secure cab sharing system using blockchain technology," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2022.
[28] Z. Zhang, G. De Luca, B. Archambault, J. Chavez and B. Rice, "Traffic dataset and dynamic routing algorithm in traffic simulation," *Journal of Artificial Intelligence and Technology*, vol. 2, no. 3, pp. 111–122, 2022.
[29] S. Namasudra, D. Devi, S. Choudhary, R. Patan and S. Kallam, "Security, privacy, trust, and anonymity," *Advances of DNA Computing in Cryptography*, vol. 1, pp. 138–150, 2018.
[30] S. A. Shawkat, B. A. Tuama and I. Al_Barazanchi, "Proposed system for data security in distributed computing in using triple data encryption standard and Rivest Shamir Adelman," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 6, pp. 6496–6505, 2022.
[31] A. K. Shrestha and J. Vassileva, "User acceptance of usable blockchain-based research data sharing system: An extended TAM-based study," in *2019 First IEEE Int. Conf. on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*, Los Angeles, CA, USA, pp. 203–208, 2019.
[32] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras and H. Janicke, "Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning," *IEEE Access*, vol. 10, pp. 40281–40306, 2022.