PeerJ Computer Science

SA-FLIDS: secure and authenticated federated learning-based intelligent network intrusion detection system for smart healthcare

Radjaa Bensaid[1], Nabila Labraoui[2], Ado Adamou Abba Ari[3], Hafida Saidi[1], Joel Herve Mboussam Emati[4] and Leandros Maglaras[5]

[1] STIC Lab, Abou Bekr Belkaid Tlemcen University, Tlemcen, Algeria
[2] LRI Lab, Abou Bekr Belkaid Tlemcen University, Tlemcen, Algeria
[3] LaRI Lab, University of Maroua, Cameroon, Cameroon
[4] Department of Mathematics and Computer Sciences, University of Dschang, Cameroon
[5] School of Computer Science, Edinburgh Napier University, Edinburgh, United Kingdom

ABSTRACT

Smart healthcare systems are gaining increased practicality and utility, driven by continuous advancements in artificial intelligence technologies, cloud and fog computing, and the Internet of Things (IoT). However, despite these transformative developments, challenges persist within IoT devices, encompassing computational constraints, storage limitations, and attack vulnerability. These attacks target sensitive health information, compromise data integrity, and pose obstacles to the overall resilience of the healthcare sector. To address these vulnerabilities, Network-based Intrusion Detection Systems (NIDSs) are crucial in fortifying smart healthcare networks and ensuring secure use of IoMT-based applications by mitigating security risks. Thus, this article proposes a novel Secure and Authenticated Federated Learning-based NIDS framework using Blockchain (SA-FLIDS) for fog-IoMT-enabled smart healthcare systems. Our research aims to improve data privacy and reduce communication costs. Furthermore, we also address weaknesses in decentralized learning systems, like Sybil and Model Poisoning attacks. We leverage the blockchain-based Self-Sovereign Identity (SSI) model to handle client authentication and secure communication. Additionally, we use the Trimmed Mean method to aggregate data. This helps reduce the effect of unusual or malicious inputs when creating the overall model. Our approach is evaluated on real IoT traffic datasets such as CICIoT2023 and EdgeIIoTset. It demonstrates exceptional robustness against adversarial attacks. These findings underscore the potential of our technique to improve the security of IoMT-based healthcare applications.

Subjects Computer Networks and Communications, Data Mining and Machine Learning, Security and Privacy

Keywords Intrusion detection, Cybersecurity, Smart healthcare

Submitted 28 May 2024
Accepted 23 September 2024
Published 13 December 2024

Corresponding author
Leandros Maglaras,
l.maglaras@napier.ac.uk

Academic editor
Sedat Akleylek

Additional Information and Declarations can be found on page 31

DOI **10.7717/peerj-cs.2414**

INTRODUCTION

The rapid rise of technologies, particularly in artificial intelligence, has significantly influenced various industries including healthcare (*Lee & Yoon, 2021*). Smart healthcare introduces innovative ideas and architectures that leverage the recent advancements in

How to cite this article Bensaid R, Labraoui N, Abba Ari AA, Saidi H, Mboussam Emati JH, Maglaras L. 2024. SA-FLIDS: secure and authenticated federated learning-based intelligent network intrusion detection system for smart healthcare. **PeerJ Comput. Sci. 10:e2414 DOI 10.7717/peerj-cs.2414**

Information and Communication Technology (ICT), including the Internet of Things, Cloud and Fog computing, wearable devices, Electronic Health Records (EHRs), and more (*Kumari et al., 2018*). Thus, smart healthcare has the potential to enhance citizen health, deliver exceptional services, reduce healthcare costs, and empower healthcare practitioners to make more precise diagnoses and treatment decisions (*Mondejar et al., 2021*).

Motivation

Despite the transformative potential of smart healthcare systems, integrating various technologies and medical devices also introduces significant vulnerabilities. Cyber-attacks on IoT medical devices can compromise patient safety, data integrity, and the availability of healthcare services (*Djenne & Saïdouni, 2018*). These attacks are becoming increasingly sophisticated, and given the limited resources and insufficient computing capabilities of many IoT devices, there is a substantial risk of these devices being exploited as bots for launching further attacks (*Bensaid et al., 2024*). Therefore, deploying a Network-based Intrusion Detection System (NIDS) is critical to ensure the sustainability and security of smart healthcare-based IoMT applications (*Saidi, Labraoui & Ari, 2022*), secure patient data, and mitigate the evolving threats posed by cyberattacks. However, these systems tend to generate high false positive rates and lack the scalability and efficiency required to combat emerging cyber threats. The adoption of machine learning approaches within IDS highlights the need for intelligent, anomaly-based detection systems that can operate with minimal human intervention (*Radjaa, Nabila & Salameh, 2023*). However, these models have common drawbacks, including reliance on a single entity to manage data from all network users, large-scale medical data storage in cloud servers leading to potential single-point failures, and concerns related to centralized data governance, which raise privacy issues (*Radjaa, Nabila & Salameh, 2023*). To address these challenges, federated learning (FL) offers a promising solution by allowing mobile devices to collaboratively train a shared model while keeping data decentralized (*Lim et al., 2020*). Several studies have employed FL-based IDS (*Iwendi et al., 2021*; *Schneble & Thamilarasu, 2019*) to enhance IoT security while preserving privacy. However, FL is susceptible to adversarial attacks such as poisoning and Sybil attacks, which can distort model accuracy and convergence. Thus, there is a crucial necessity for secure FL mechanisms to guard against manipulated data and models. In this context, the integration of blockchain technology with FL can provide an innovative framework to counter these attacks by ensuring immutability, transparency, and security of the data and model updates (*Qu et al., 2020*; *Ali, Karimipour & Tariq, 2021*).

Contributions

To the best of our knowledge, none of the prior research has specifically emphasized user authentication within the FL process. In our article, we propose the SA-FLIDS framework, a novel Secure and Authenticated FL-based NIDS framework using Blockchain for protecting IoMT-enabled smart healthcare networks. SA-FLIDS employs a secure FL approach for detecting anomalies in the IoMT network, thereby creating an intelligent NIDS. This system can effectively identify and counter cyber-attacks aimed at IoMT

devices, ensuring the security and integrity of the overall system. It is based on blockchain and Self-Sovereign Identity (SSI) technologies for secure FL, employing an identity management and devices authentication scheme to protect FL against adversarial attacks and ensure that only trusted nodes (*Benfriha et al., 2023*) contribute to the training. Hence, participants' privacy is ensured by SA-FLIDS, as the centralized training module does not share users' private data. SA-FLIDS uses gRPC (Remote Procedure Call) for efficient node communication and TLS (Transport Layer Security) for encrypted channels. It also employs the trimmed mean method for model aggregation, reducing the impact of adversarial data and outliers, enhancing resilience against data poisoning, and maintaining the global model's integrity. Our proposed model is assessed using the two latest datasets, CICIoT2023 (*Neto et al., 2023*) and Edge-IIoTset (*Ferrag et al., 2022*), and demonstrates superior performance across key metrics including accuracy, precision, recall, and F1-score, while maintaining a low rate of false positives. Moreover, we assess and evaluate the blockchain-based client authentication framework, employing the decentralized identifiers (DID) and verifiable credentials (VC) model through the Hyperledger Indy blockchain and Aries library. The overall contributions of this article are outlined below:

1. We propose our Secure and Authenticated Federated Learning-based NIDS (SA-FLIDS) framework to identify and prevent cyber-attacks in IoMT-enabled smart healthcare systems.
2. We incorporate blockchain-based Self-Sovereign Identity (SSI) to authenticate participants, ensuring that only trusted nodes participate in the Federated Learning (FL) process.
3. We employ the trimmed mean method aggregation in FL. This approach enhances resilience against data poisoning attacks and preserves the integrity of the global model in a distributed IoMT environment.
4. We implement gRPC with TLS encryption in our work specifically for secure communication in an FL-based NIDS for healthcare IoT. This combination ensures both efficient and secure data exchange between IoMT devices, fog nodes, and servers.
5. We provide a comprehensive evaluation of our proposed system using two recent, real-world IoT security datasets (CICIoT2023 and Edge-IIoTset), demonstrating its effectiveness in detecting a wide range of attacks relevant to smart healthcare environments.

RELATED WORK

This section provides an overview of relevant literature on federated learning (FL), machine learning (ML), and blockchain in the context of Intrusion Detection Systems (IDS) for IoT networks.

Schneble & Thamilarasu (2019) proposed FLIDS, an FL-based IDS for medical cyber-physical systems (MCPS). Their model reduces communication and computation expenses and is effective in identifying various attacks. However, it is vulnerable to poisoning attacks.

Similarly, *Chatterjee & Hanawal (2021)* applied FL with a convolutional neural network for IoT intrusion detection. Their model handles non-IID data and dynamically optimizes through weighted client aggregation. However, it lacks real-world applicability and security considerations, making it vulnerable to network poisoning attacks.

The FL anomaly detection system presented by *Man et al. (2021)* uses GRUs with preprocessing and ensemble learning, outperforming traditional approaches. However, their model lacks security measures when sharing trained models, leading to a risk of data leakage.

Rey et al. (2022) proposed an FL-based malware detection method for IoT devices but overlooked channel security, reducing system resilience against attacks. Adversarial machine learning algorithms could undermine their effectiveness in IoT healthcare systems.

Ruzafa-Alcázar et al. (2021) integrated differential privacy techniques into training an IDS for industrial IoT using FL. While offering privacy guarantees, it is susceptible to inference attacks, impacting overall performance. Similarly, *Zhao et al. (2019)* introduced the MT-DNN-FL (Multi-Task Deep Neural Network in FL), demonstrating high detection rates and reduced training time. However, further optimization is needed to accommodate IoT device limitations. *Friha et al. (2022)* presented FELIDS, a FL-based IDS for securing IoT infrastructures, which uses local learning and the FedAvg algorithm, a widely adopted aggregation method (*McMahan et al., 2017*). The global model, as computed in previous studies (*Schneble & Thamilarasu, 2019*; *Wang et al., 2022*; *Elayan, Aloqaily & Guizani, 2021*; *Wu et al., 2020*) shares model updates between devices and an aggregation server. While offering cost-effective computing, challenges such as communication latency and privacy vulnerabilities require more advanced aggregation techniques.

Ashraf et al. (2022) proposed a blockchain-based FL IDS for IoT healthcare, with sensor monitoring and an artificial neural network (ANN) model for attack detection. Outperforming existing methods. However, privacy concerns and the decentralized nature of patient data need to be addressed.

Preuveneers et al. (2018) integrated FL with a blockchain featuring access control for IDS. The study's model is relatively simple, limiting generalization. Similarly, *Lakhan et al. (2022)* introduced the FL-BETS framework, safeguarding privacy and detecting fraudulent activities using FL and blockchain technology. Computation overhead is a challenge, requiring optimized FL algorithms for IoHT devices. *Baucas, Spachos & Plataniotis (2023)* proposed a fog-based IoT platform using federated learning and blockchain to enhance privacy and security in wearable healthcare devices. Their system maintains patient privacy and provides robust access control, demonstrated through a custom testbed. However, the study's testing with only a single dataset type does not focus specifically on securing the IoMT network. Furthermore, it does not specifically deal with adversarial attacks on the federated learning process.

Sindhusaranya et al. (2023) proposed a privacy-preserving approach using FL-BEPP (Federated Learning with Blockchain-Enabled Privacy Preservation) to address both soft and hard constraints in fraud prevention and security for the Internet of Medical Things (IoMT). Their method aims to enhance data privacy and security in healthcare systems.

However, the implementation of blockchain transactions and federated learning model updates introduces additional computational overhead, potentially resulting in increased latency and reduced performance, especially in large-scale IoMT deployments. This trade-off between enhanced security and system efficiency presents a challenge for widespread adoption in complex healthcare networks.

Begum et al. (2024) proposed BFLIDS, a system combining blockchain and federated learning for intrusion detection in IoMT networks. The approach uses decentralized model training to preserve privacy, blockchain for secure record-keeping, and smart contracts for system management. They modified the FedAvg algorithm to improve accuracy and resilience against attacks. While BFLIDS showed competitive performance, the study didn't address the resource constraints of IoMT devices or the potential computational overhead from smart contract integration.

The primary goal of the articles mentioned is to establish an IDS-based federated learning framework to maintain security and preserve privacy in IoT applications. However, many existing solutions encounter challenges in addressing specific potential attacks, such as poisoning, Sybil, Data Tampering, and Eavesdropping attacks within FL-based IDS, particularly in healthcare systems. Ensuring security and privacy in efficient FL-based IDS remains a crucial concern, which is the focal point of our investigation in this study.

To address these challenges, we introduce a new Secure and Authenticated FL-based NIDS framework using Blockchain (SA-FLIDS) for fog-IoMT-enabled smart healthcare systems. The SA-FLIDS framework leverages FL to train a shared prediction model while maintaining decentralized data on the devices themselves and incorporates a blockchain-based SSI model for a privacy-preserving authentication scheme. This combination ensures that only trusted IoMT devices contribute to the FL process. Moreover, our framework distinguishes itself from existing literature by incorporating a robust aggregation method, specifically the trimmed mean, to reduce the influence of outliers or malicious participants while computing the global model. Furthermore, we use TLS encryption combined with secure communication protocols like gRPC which helps ensure data integrity during transmission between devices, fog nodes, and servers.

BACKGROUND

This section provides essential contextual background for our proposed model. Initially, we introduce FL-based NIDS. Subsequently, we delve into Blockchain-based SSI techniques that are pertinent to and integrated within our proposed framework.

Federated learning for IoT intrusion detection

Federated learning offers enhanced privacy and security in IoT networks by minimizing data transmission (*Radjaa, Nabila & Salameh, 2023*). In IDS, this approach enables the development of more intelligent machine learning models exposed to diverse data sources while ensuring user privacy (*Sarhan et al., 2023*). In this process, models are downloaded and updated locally on IoT devices using their data, then transmitted to a central server for aggregation, resulting in an improved global model. However, effective

data distribution presents practical and technical challenges for successful federated learning deployment.

Blockchain-based SSI

Now shifting our focus to Blockchain-based SSI, this section explores the utilization of Blockchain technology in establishing secure and decentralized identity management systems.

Blockchain

Is a distributed ledger technology for data transmission and storage that records the history of all transactions. Blockchain has been involved due to its decentralization, immutability, and persistence properties in the distributed peer-to-peer (P2P) network. It utilizes asymmetric cryptography to ensure transactions are done safely (*Saidi et al., 2022*). Blockchain is based on several elements used to create a secure, transparent, and decentralized system. Each element plays a crucial role in maintaining the integrity and functionality of the blockchain (*Mboussam Emati & Mboussam, 2023*), including block, transaction, consensus mechanism, smart contract, mining, immutable ledger, cryptographic keys, and hash function.

Self-sovereign identity

is a decentralized approach and a new model for digital identity. Self-sovereign identity (SSI) aims to empower individuals to possess and control digital proof of their credentials. Thus, it helps to prove who we are by establishing trusted relationships to access information (*Saidi et al., 2022*). SSI is based on two main standards: Decentralized identifier (DID) and verifiable credential (VC):

- **Decentralized identifier** is a new type of identifier, defined by the W3C (*Emati, Mboussam & Tchendji, 2023*). DIDs are designed to enable Self-Sovereign Identity on the internet, providing a way for individuals to have control over their own digital identity without the need for a central authority. Thus, users can selectively disclose only the necessary information for a particular transaction or interaction (*Saidi et al., 2022*).
- **Verifiable credential** is a standard method for digitally expressing credentials in a cryptographically secure way. It can include metadata, claims, and proofs used to verify a credential (*Thomas, Ramaguru & Sethumadhavan, 2022*). Credentials are created and signed by the issuer using his private key and then issued to the holder, enabling the verifier to confirm the VC. A holder keeps and shares the received credentials with a verifier. The verifier accepts and approves these credentials (*Figueroa-Lorenzo, Benito & Arrizabalaga, 2021*). based on the public key associated with a DID.

CYBER ATTACKS AND RISKS IN IOMT-ENABLED SMART HEALTHCARE SYSTEMS

Due to the critical nature of patient well-being, reliable and secure communication is vital in smart healthcare (*Lee & Yoon, 2021*). Medical IoT devices, with resource constraints like poor battery life and limited memory, are vulnerable to hacking attempts,

potentially integrating them into botnets. Common cyberattacks on compromised IoMT devices include:

- *DoS and DDoS attacks* aim to undermine IoMT availability, with DoS using a single botnet and DDoS utilizing multiple botnets (*Bensaid et al., 2024*).
- *Information gathering attacks* collect comprehensive data about IoMT, often using reconnaissance-like scanning attacks (*Jensen, Gruschka & Herkenhöner, 2009*).
- *Exploiting web-based vulnerabilities* targets web services on IoMT devices, employing methods like injection, hijacking, and DoS (*Jensen, Gruschka & Herkenhöner, 2009*).
- *Communication spoofing attacks* enable unauthorized access to network traffic, facilitating data theft and malware dissemination (*van der Merwe et al., 2018*).
- *Brute-force attacks* attempt to discover passwords or passphrases by iteratively trying words from predefined lists (*Stiawan et al., 2019*).
- *The Mirai attack*, a widespread DDoS assault, specifically targets IoMT devices (*Gamblin, 2017*).

Therefore, if a malicious IoMT device compromises a fog server, unauthorized access to sensitive patient data and EHRs will be possible, affecting the data privacy and security of patients. This compromised data often includes private and sensitive information such as credit card details, health conditions, and other confidential data, thereby exposing patients to significant risks. Additionally, the unavailability of fog servers disrupts essential healthcare services such as the monitoring of patient's vital signs. As a consequence, the ability to track and monitor essential health indicators in real-time is compromised, posing potential risks and presenting challenges in delivering timely and suitable care, as illustrated in Fig. 1. Additionally, in critical situations, such as emergencies, the heightened risk to patient's lives, underscores the severity of potential consequences. Therefore, it is crucial to establish robust cybersecurity measures and implement comprehensive security protocols to protect patient data which are essential for the sustainability of critical healthcare services and the effective mitigation of risks.

SA-FLIDS SYSTEM ARCHITECTURE AND DESIGN GOAL

In this section, we discuss the SA-FILDS architecture and the threat model, followed by the design goals of the SA-FILD system.

SA-FLIDS system architecture

SA-FLIDS system brings a novel Secure and Authenticated Federated Learning-based NIDS framework for Smart Healthcare using Blockchain Technology, secure communication protocols, DID, and VC. SA-FLIDS system aims to examine network traffic, identify and mitigate cyber-attacks against IoMT devices, and enhance the security of smart healthcare systems. The architecture of our proposed model comprises three layers: the Cloud layer, the Fog layer, and the IoMT devices layer, as depicted in Fig. 2.

1. **Cloud layer:** This layer provides an underlying infrastructure and resources that enable the provision of on-demand and adaptable services accessible from any location

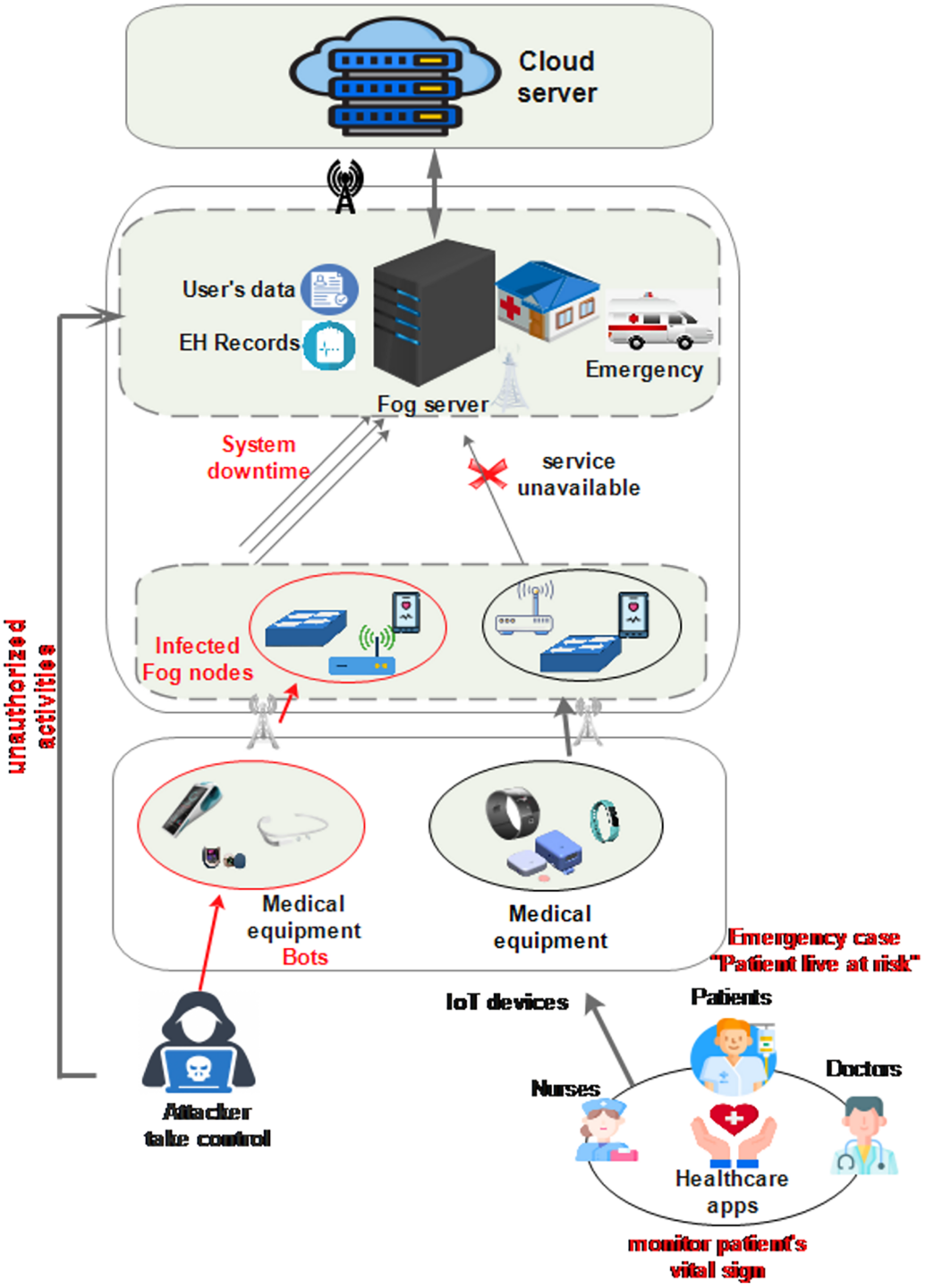


Figure 1 **Cyber-attacks in smart healthcare.** Icon credit: Electronic Records icon (mcmurryjulie, https://pixabay.com/fr/vectors/pronostic-ic%C3%B4ne-le-dossier-du-patient-2803190/, Pixabay license); Doctor icon (Freepik, https://www.flaticon.com/free-icon/doctor_5065189, Flaticon license); Nurse icon (Freepik, https://www.flaticon.com/free-icon/nurse_9133509, Flaticon license); Patient icon (Freepik, https://www.flaticon.com/free-icon/patient_4228704, Flaticon License).
Full-size DOI: 10.7717/peerj-cs.2414/fig-1

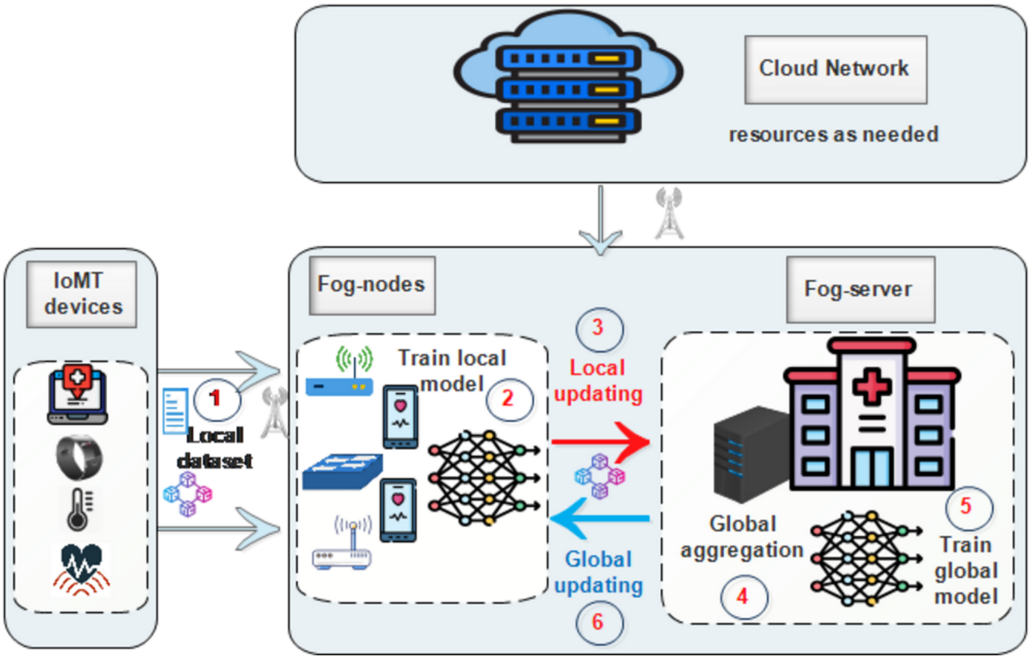


Figure 2 The proposed model. Icon credit: Hospital icon (Freepik, https://www.flaticon.com/free-icon/hospital_4320350, Flaticon license); Healthcare device icon (Smashicons, https://www.flaticon.com/free-icon/healthcare-device_2904470, Flaticon license); Blockchain icon (jojooid, https://www.flaticon.com/free-icon/blockchain_8757988, Flaticon license). Full-size DOI: 10.7717/peerj-cs.2414/fig-2

(*Saidi, Labraoui & Ari, 2022*). Fog nodes and fog servers can dynamically allocate resources based on their requirements during the FL process.

- **Fog network layer:** The intelligent NIDS based on FL and blockchain is deployed in the Fog layer. This layer supervises the network and makes decisions regarding traffic flow classification. It can be deployed at each hospital or clinic. The Fog Network Layer consists of two distinct sub-layers:
- **Fog server:** Serving as the central server that initiates and constructs a shared global model architecture among participating fog nodes. It is responsible for adding verified model weights to the blockchain, ensuring that local model updates are securely and transparently aggregated and shared across the network using the TLS protocol (*Möller et al., 2022*). Additionally, blockchain is deployed for identity management to ensure that only authenticated devices can participate in the FL.

2. **Fog nodes:** These distributed nodes consist of physical components such as mobile devices, gateways, routers, and switches. They act as clients in FL, performing local model training to protect sensitive medical data. Therefore, fog nodes are effective for local model training due to their proximity to the fog server and the IoMT layer, as well as their increased processing power, memory, and connection as compared to individual IoMT devices. Furthermore, each fog node is identified by a DID which is registered into the blockchain. DIDs are used to sign documents or transactions, create secure and

persistent communication channels, and send encrypted private messages (*Figueroa-Lorenzo, Benito & Arrizabalaga, 2021*).

3. **IoMT layer:** The Internet of Medical Things (IoMT) layer is used to sense, collect, encrypt, and upload medical data to the fog nodes for private local model training. The transmitted data can encompass both benign network traffic and potential cyber-attack classification. Moreover, each IoMT device is identified by a unique DID registered on the blockchain to handle the authentication process.

Challenges and design goal

FL systems are susceptible to various adversarial attacks, posing risks to their security, and integrity, and hindering their deployment in NIDS. We detail the following attacks that can be launched against FL systems:

- *Sybil attack*, a malicious participant creates multiple fake identities to disrupt FL, injecting biased or misleading information into aggregated model updates (*Lian et al., 2023*).
- *Eavesdropping attack*, involves unauthorized interception of communication between FL participants, potentially leading to privacy breaches by accessing sensitive information such as model updates or raw data (*Lian et al., 2023*).
- *Data poisoning attack*, in FL involves injecting adversarial data into the training set of participating devices, aiming to compromise the integrity and performance of the global model (*Lian et al., 2023*).
- *Data tempering* involves unauthorized modification of data in the FL process, aiming to compromise the integrity and reliability of the global model by injecting malicious or false information.

To tackle the challenges mentioned above, we are emphasizing the following design goals:

- Ensuring the security of FL.
- Ensuring that only authenticated IoMT devices can participate in the FL using blockchain-based DID to prevent Sybil attacks.
- Secure communication between IoMT devices, fog nodes, and fog servers, and prevent data tampering and eavesdropping during communication between nodes.
- Reduce the influence of outliers or malicious participants when computing the global model using a robust aggregation function to mitigate the impact of adversarial participants engaging in poisoning attacks and introducing noisy data.

SA-FLIDS SYSTEM

This section presents details of our proposed scheme, SA-FLIDS, which aims to ensure security and privacy preservation in an IoMT-enabled smart healthcare system.

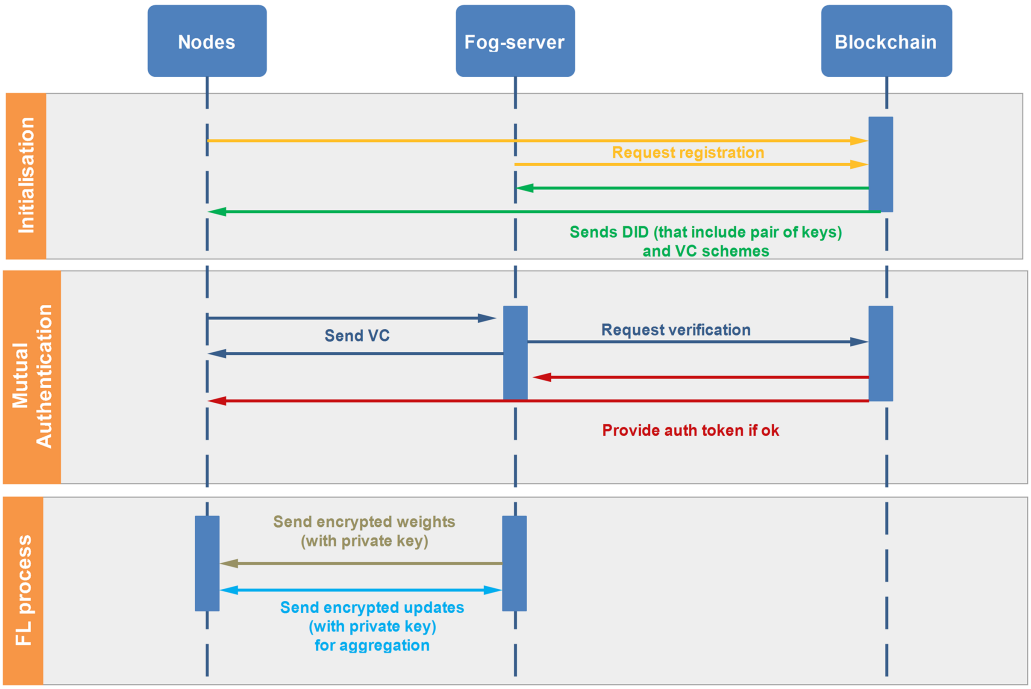


Figure 3 **Sequence diagram of our proposed SA-FLIDS.**
Full-size DOI: 10.7717/peerj-cs.2414/fig-3

SA-FLIDS identification and authentication approaches

Figure 3 illustrates a sequence diagram of our system throughout its lifecycle, involving nodes, a fog server (FS), and a blockchain (BC). The process contains three phases:

- **Initialization phase:**
  1. Nodes and the FS request registration from the BC.
  2. The BC responds by sending a DID, which includes a pair of keys and a VC scheme for each entity.
- **Mutual authentication phase:**
  1. Nodes send their VC to the FS, and the FS also sends its VC to the nodes. Both parties then request verification from the BC.
  2. The BC verifies the credentials and, if everything is in order, provides an authentication token to both the nodes and the FS.
- **Federated learning process phase:**
  1. The FS sends encrypted data, secured with its private key, to the Nodes, enabling them to work on the data securely.
  2. After processing, the nodes send encrypted updates, secured with their private keys, back to the FS. The FS then aggregates these updates.

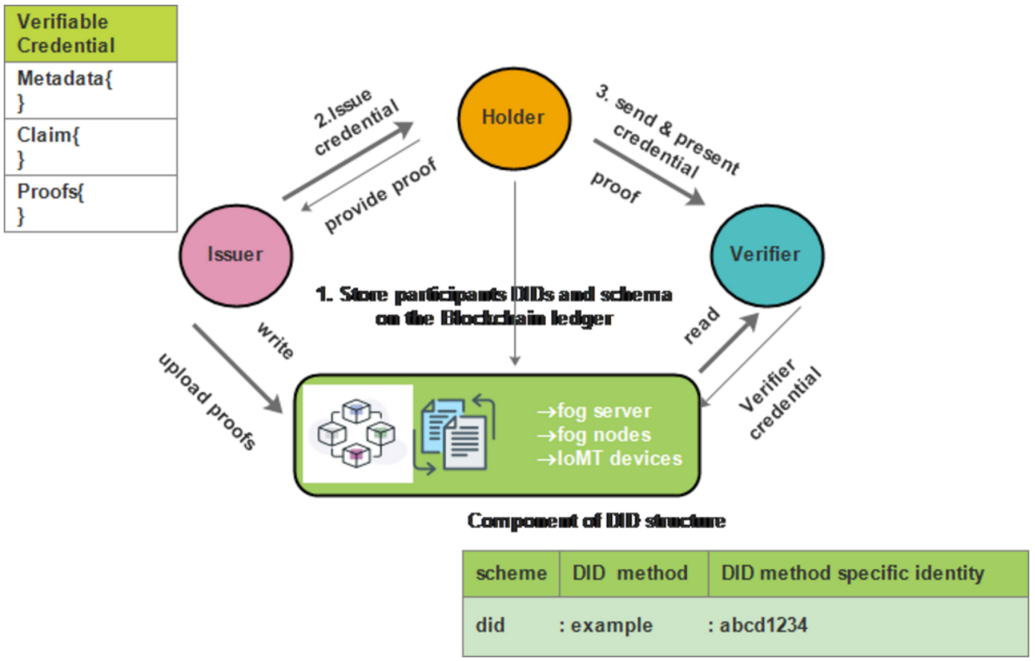


Figure 4 Trust triangle for VC. Icon credit: Blockchain icon: (© Abdul Basit Noohani Dreamstime.com, https://www.dreamstime.com/visualize-power-blockchain-symbolic-icon-representing-secure-digital-ledgers-decentralized-transactions-image291382086); Distributed Ledger icon (Kalashnyk, https://www.flaticon.com/free-icon/documents_12864624, Flaticon license).
Full-size DOI: 10.7717/peerj-cs.2414/fig-4

The SA-FLIDS model integrates a Blockchain-based DID and VC system to enhance participant identification within the FL process. Figure 4 demonstrates the triangle of trust and privacy in digital interactions, comprising three main parties: the issuer, holder, and verifier. In this model:

1. The issuer is the hospital, which manages the blockchain. The blockchain stores only the DIDs and VC schemes.
2. Nodes and FS act as both holders and verifiers depending on their role in mutual authentication.

- When the fog server receives data from nodes, it acts as the verifier.
- When nodes receive an authentication query from the FS, they act as the verifier, and the FS is the owner.

Initially, the issuer uploads proofs and stores participants' DIDs and schema on the blockchain ledger. Then, the holder receives and stores the issued credentials. Subsequently, the holder presents the credentials to the verifier, who reads and verifies them. This system ensures secure, private, and verifiable digital interactions.

Detectin process FL in SA-FLIDS system

After successful authentication and verification processes, the fog server sets up a global model architecture distributed across the involved fog nodes. During each training round,

Algorithm 1 **Federated learning.**

Initialize global model θ_0
for $t = 1$ to T **do**
  for each fog node i **do**
    Collect local data D_i from IoMT devices
    Train local model θ_t^i on D_i
    Send θ_t^i to server
  end for
  Aggregate local models: $\{\theta_t^1, \theta_t^2, \ldots, \theta_t^N\}$
  $\theta_{t+1} = \text{Trimmed Mean}(\{\theta_t^1, \theta_t^2, \ldots, \theta_t^N\})$
  for each fog node i **do**
    Distribute θ_{t+1} to fog node i
  end for
end for
Output: Final global model θ_T

every fog node updates its local model by conducting training on the data collected from IoMT devices within its proximity. Subsequently, these fog nodes transmit their updated models back to the central fog server. Then, the central fog server collects and aggregates these models from all fog nodes using the robust trimmed mean aggregation. Next, the fog server forwards the updated global model back again to all the fog nodes. This iterative process is repeated for each training round until the final global model is obtained and ready for use. Moreover, all those communications occur securely by using both the gRPC framework for effective communication among nodes and the TLS to ensure end-to-end encryption in the communication channels. The final global model is then used for the detection classification process which is deployed in the NIDS to differentiate between normal and potentially malicious traffic patterns. Algorithm 1 demonstrates the FL process.

Trimmed-mean aggregation method

The Trimmed-Mean method enhances the accuracy and reliability of the global model in FL by mitigating the influence of outliers and malicious participants. It removes a certain percentage of extreme values from local models and calculates a weighted average of the remaining models. The Trimmed-Mean formula is:

$$\text{Trimmed} - \text{Mean} = \frac{\sum_i (w_i \cdot m_i)}{\sum_i w_i} \tag{1}$$

where:

w_i represents the weight of the i-th local model.

m_i represents the i-th local model in the trimmed set M_t.

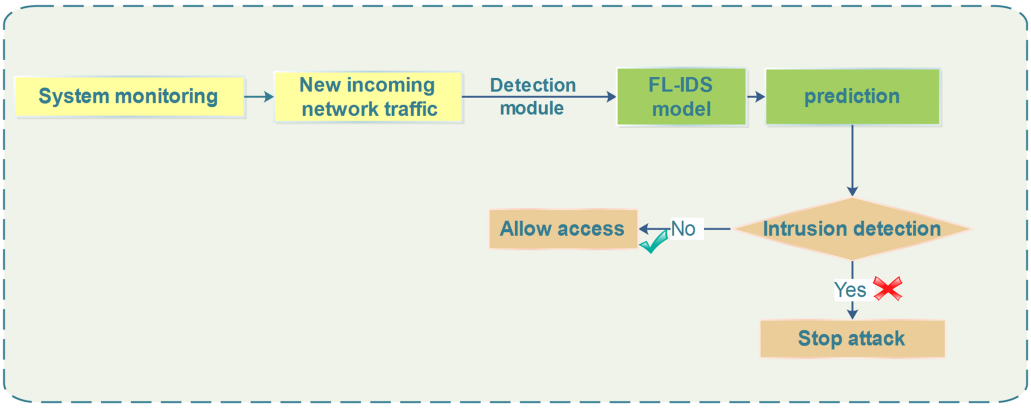


Figure 5 **SA-FLIDS detection and mitigation scheme.** Full-size DOI: 10.7717/peerj-cs.2414/fig-5

Intelligent NIDS-based SA-FL for mitigation process

In the SA-FLIDS model, The fog server plays a dual role in the system, not only detecting potential intrusions but also promptly and effectively responding to mitigate or minimize the consequences of these attacks. Hence, the SA-FL model which is trained through the robust process, is integrated into the NIDS and incorporates the intrusion detection and response mechanism into the fog server's architecture. The flow involves continually monitoring fresh incoming traffic and sending it through the FL model for prediction, which has previously been trained, to differentiate between normal and potentially malicious traffic patterns. If the FL model predicts normal traffic, access is granted, demonstrating the model's ability to make real-time decisions based on learned patterns of normal behavior. Otherwise, if the FL model identifies incoming traffic as indicative of an intrusion or attack, it triggers an alarm in the system monitoring, promptly alerting the system to the potential threat. Therefore, the NIDS responds in real-time, taking proactive measures to block and drop malicious packets, as illustrated in Fig. 5.

EXPERIMENTS AND RESULTS

This study explores the potential of SA-FLIDS in detecting intrusion in IoMT networks. As a result, this decentralized approach could be crucial for securing healthcare applications. In this section, we present the experimental setup along with the evaluation metrics and results.

Experimental setup

In this section, we delve into the experimental setup for both FL and blockchain-based SSI.

Experimental setup for federate learning

Table 1 presents the parameters used in federated deep learning. In our study, we conducted experiments deploying our model with client sets denoted as K, where K = 10. We employed the Independent and Identically Distributed (IID) approach, ensuring that the data distribution across the dataset matches the distribution of data for each client. Furthermore, To prevent overfitting, the following techniques were used:

Table 1 Federated deep learning classifier parameter setting.

	Parameter	Value
Federated deep learning classifier	Local epoch	10
	Global epoch	4
	Batch size	128
	Hidden layer	2
	Hidden nodes	128,64
	Activation function	Relu
	Regularization	L2
	Classification function	Sigmoid/softmax
	Optimizer, learning rate	Adam, 0.001
	Loss function	Binary_crossentropy Categorical_crossentropy

- **Stratified K-fold cross-validation:** Set k = 5 to split the data into five subsets, evaluating the model's performance.
- **L2 regularization:** Applied with a factor of 0.01 to the dense layers, adding a penalty to the loss function based on the weights' magnitude, simplifying the model.
- **Early stopping:** With the patience of three, training stops if the validation loss does not improve, monitoring the validation set's performance.

Experimental setup for blockchain-based SSI

The simulation is carried out on a Lenovo ThinkPad P51 -Core i7 2.9 GHz–SSD 1 To–32 Go, computer running Ubuntu 18.04 LTS. The simulation environment is based on Hyperledger Indy (*Banerjee et al., 2022*), a framework dedicated to self-sovereign identity management. It offers an abstraction that enables DIDs and VCs to be created, verified, and revoked. It embeds Hperledger ursa, a module that provides all the primitives for cryptographic operations. We also use Hyperledger Aries (*Manoj, Makkithaya & Narendra, 2022*). This is a library for creating agents that can manage the cryptographic wallet of each player. It also offers interfaces for creating functionalities that realize the behavior of the wallet owner.

The simulation begins with the initialization phase, during which 10 nodes plus the server are each initialized in a Docker container as shown in Fig. 6. The server is the trusted authority that manages the blockchain. It creates the genesis block and initializes the chain. Each node is registered on the chain and receives a DID and an authentication scheme. Of course, each DID is accompanied by its key pair, which is stored in each node's wallet, and the public keys are known to everyone. Next comes the authentication phase, when each node sends its VC to the server, which verifies it and receives the server's VC (mutual authentication). The principle is to request the blockchain's response to the VC presented to it. Once each node's authenticity has been verified. The nodes can exchange data. Remember that confirmation of a VC leads to validation of a token, which has a lifetime

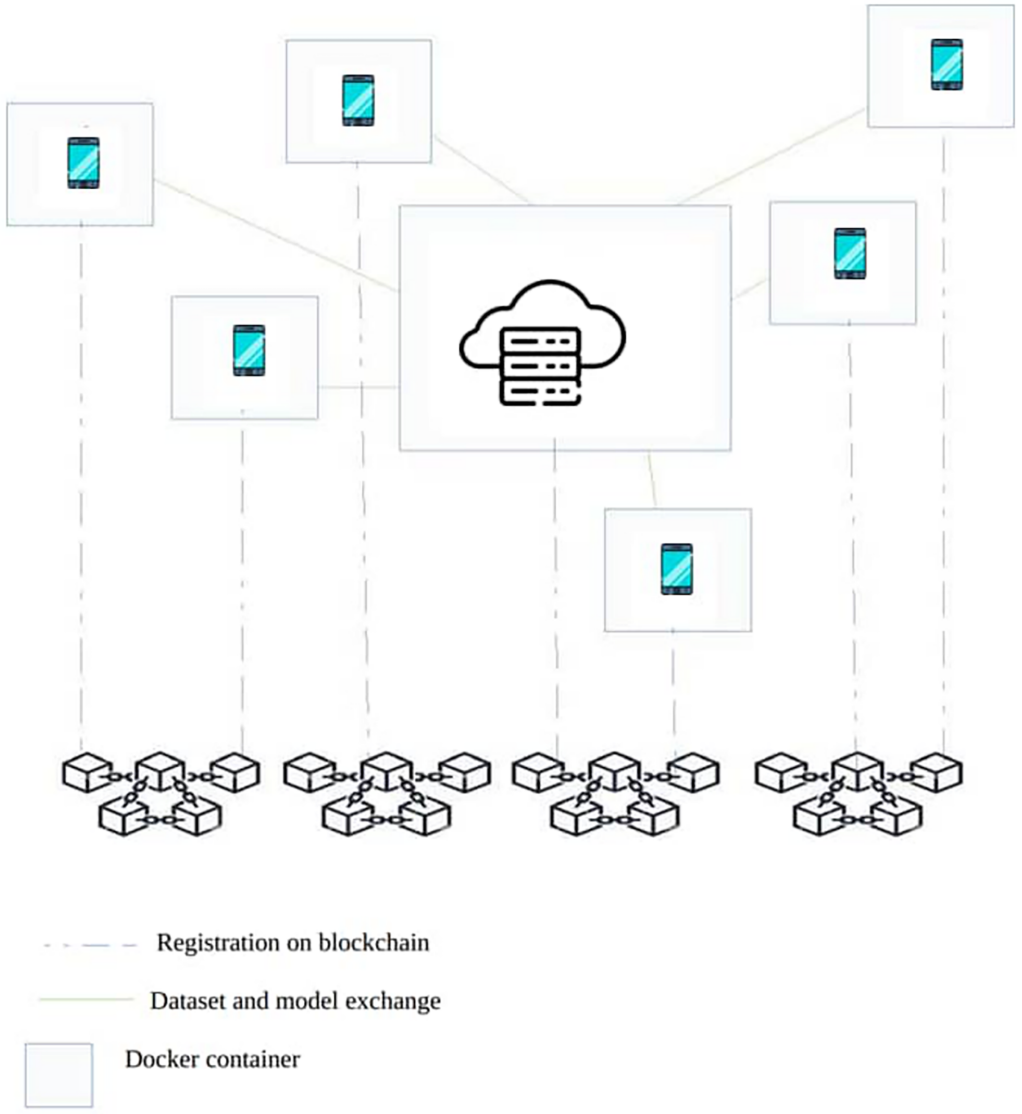


Figure 6 Blockchain network. Icon credit: Blockchain icon (© Stockoxinoxi | Dreamstime.com, https://www.dreamstime.com/blockchain-concept-symbol-vector-icon-image222346576).
Full-size DOI: 10.7717/peerj-cs.2414/fig-6

(that of the session). Each node in possession of its token can then use it for future exchanges. After authentication, the server sends the chunks of the dataset to the various nodes using its private key as demonstrated in Fig. 3. The nodes receive them and can proceed with processing. The different scores are exchanged using the same process. The communication protocol in this phase is gRPC. However, the solution would work with any other protocol, with security being guaranteed by the authentication token.

Datasets description

Datasets play an essential role in both training and assessing IDSs within IoT networks. The choice of suitable datasets tailored to particular tasks holds significant importance, particularly in evaluating the efficacy of FL approaches for IoT networks. In our

experiment, we incorporated two recent datasets specifically designed to mimic real-world conditions for IDSs: CICIoT2023 (*Neto et al., 2023*), made available in 2023, and the Edge-IIoTset dataset, released in 2022.

1. ***CICIoT2023 dataset:*** A *Neto et al. (2023)* novel and extensive IoT attack dataset to foster the development of security analytics applications in real IoT operations. To accomplish this, 33 attacks are executed in an IoT topology composed of 105 devices, and all attacks are executed by malicious IoT devices targeting other IoT devices. We analyzed a dataset containing 47 features (not including label and sublabel) based on 2,366,956 samples extracted from the first 10 CSV files provided by the Canadian Institute.
2. ***Edge-IIoTset dataset:*** It is tailored specifically for IIoT and IoT applications (*Ferrag et al., 2022*), providing an authentic test environment closely resembling real-world IoT/IIoT settings. Within this environment, we conducted simulations of genuine cyberattacks to collect datasets comprising both legitimate and malicious network traffic. This dataset includes data generated by various IoT devices, spanning from heart rate sensors to flame sensors, temperature, and humidity sensors. The testbed is structured into seven interconnected layers. We utilized the Selected dataset for ML and DL/DNN-EdgeIIoT-dataset CSV file (*Banerjee et al., 2022*), which contains 61 features and 2,219,201 samples, encompassing both normal traffic and 14 distinct attacks in the IoT and IIoT environment.

Preprocessing

The datasets undergo several preprocessing steps to ensure their suitability for analysis as demonstrated in Fig. 7. After cleaning the data we first, address imbalanced data by implementing SMOTE (Synthetic Minority Over-sampling Technique) and under-sampling techniques to enhance predictive performance, particularly for minority classes. Secondly, data transformation is conducted using the StandardScaler for standardization, adjusting data to have a mean of 0 and a standard deviation of one. Additionally, feature importance analysis is performed using insights from random forest and XGBoost experiments. Finally, the processed dataset is split into an 80% training set and a 20% testing set, ensuring no duplication between the two, contributing to refining the dataset for subsequent analysis and modeling tasks. In the case of the CICIoT dataset, we opt to eliminate the Brute and Web attack labels due to their limited number of samples, which could potentially skew the analysis and compromise the reliability of the results. The detailed features selected and the attacks used are outlined in Table 2 provided below.

Evaluation metrics

In this section, we introduce the metrics employed in our experiments to evaluate both FL and SSI-based DID.

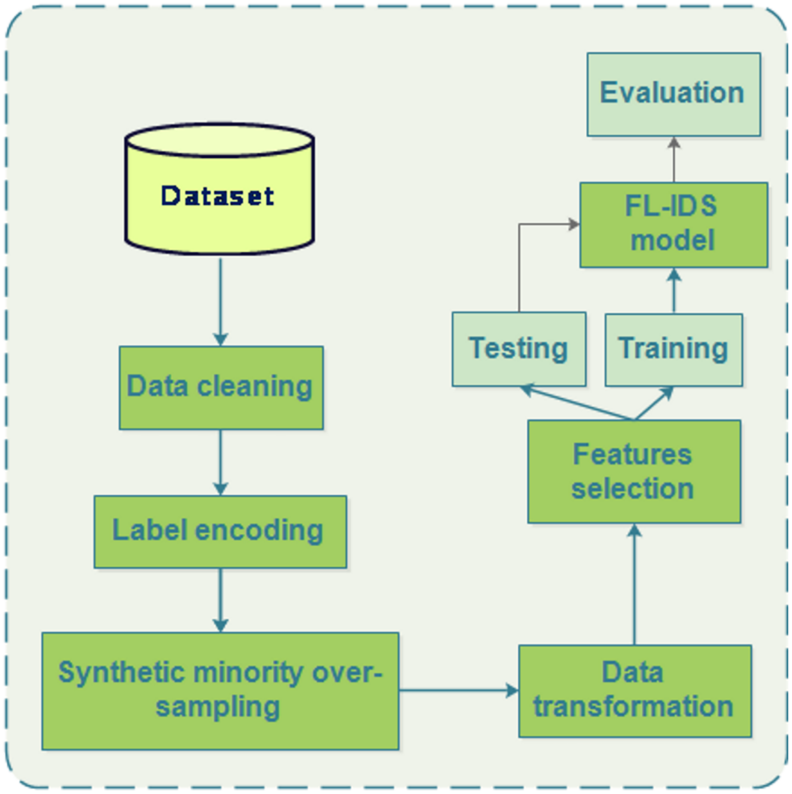


Figure 7 **Data preprocessing.** Full-size DOI: 10.7717/peerj-cs.2414/fig-7

Table 2 **Datasets description for experimental evaluation.**

	CICIoT2023	Edge-IIoTset
Features selected	'flow_duration', 'Header_Length', 'Protocol Type', 'Rate', 'Srate', 'syn_count','urg_count','rst_count', 'Tot sum', 'Min', 'Max', 'AVG', 'Tot size', 'IAT', 'Magnitue', 'Variance'	'http.content_length', 'http.request.method', 'http.referer', 'http.request.version', 'tcp.ack', 'tcp.ack_raw', 'tcp.checksum', 'tcp.flags', 'tcp. len', 'tcp.seq', 'udp.time_delta', 'dns.qry.name.len', 'mqtt.conack.flags', 'mqtt.protoname', 'mqtt.topic'
Label	'Benign', 'DDoS', 'DoS', 'Mirai', 'Recon', 'Spoofing'	'DoS/DDoS', 'Information gathering', 'Injection', 'Malware', 'Man in the middle', 'Normal'

Metrics used for federated learning evaluation

When conducting intrusion detection using federated deep learning performance analysis, the most common metrics used are:

- True negatives (TN): Benign network activity correctly classified as normal.
- True positives (TP): Malicious network activity correctly identified as an attack.
- False positives (FP): Benign network activity incorrectly classified as malicious.
- False negatives (FN): Malicious network activity is incorrectly classified as normal.

Table 3 Performance metrics.

Metric	Formula	Description
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	A measure that quantifies the proportion of instances correctly classified among the total number of observed samples.
Precision	$\frac{TP}{TP+FP}$	A metric that indicates the proportion of correctly predicted positive instances out of the total predicted positive instances.
Recall	$\frac{TP}{TP+FN}$	The proportion of correctly identified positive samples.
F-measure	$\frac{2 \times Precision \times Recall}{Precision + Recall}$	The harmonic mean of precision and recall.

Moreover, we have used a variety of measures to evaluate our proposed model, including precision, recall, precision, F-score, and accuracy, to conduct a systematic comparative analysis with other relevant approaches as demonstrated in Table 3.

Metrics used for blockchain-based SSI evaluation

To compute the metrics outlined below, we utilize the following formulas:

1. **Startup duration (SD):** The duration for the system to initiate.
2. **Connect duration (CD):** The time required for the system to establish connections between nodes and the Fog server.
3. **Publish duration (PD):** The duration for the system to publish schema credentials and related settings.
4. **Issuing credential duration (ICD):** The time taken for the system to issue credentials.
5. **Completed credential exchanges duration (CCED):** The total time needed for all credential exchanges to conclude.Metrics such as SD, CD, PD, and ICD are used to assess the Initialization phase, while CCED is used to assess the Mutual Authentication phase.
6. **Average time per credential duration (ATCD):** The average time taken to issue a single credential.

$$\text{ATCD} = \frac{ICD}{N_{\text{credentials}}} \tag{2}$$

where:
ICD is the issuing credential duration.
$N_{credentials}$ is the total number of credentials issued.

7. **Average time per transaction duration (ATTD):** The average time taken per transaction.

$$\text{ATTD} = \frac{\sum_{i=1}^{N_{\text{transactions}}} T_i}{N_{\text{transactions}}} \tag{3}$$

where:

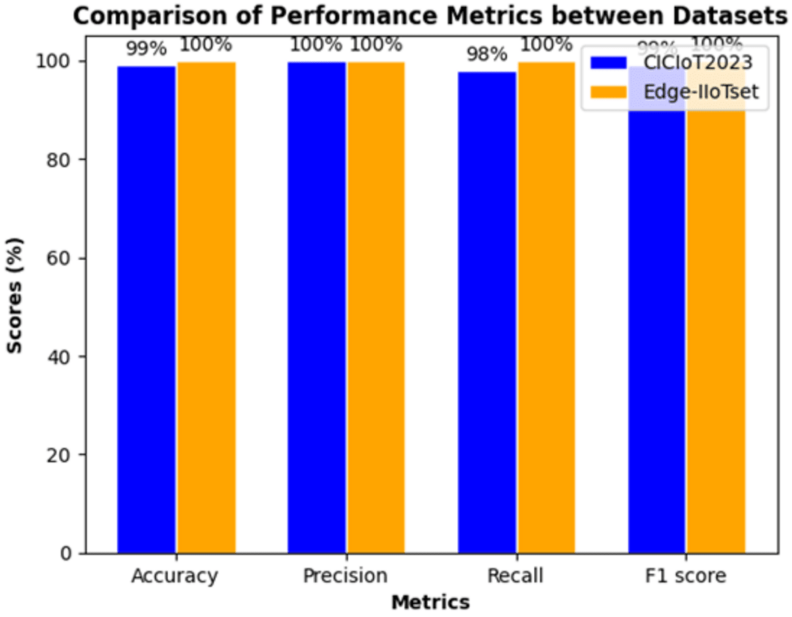


Figure 8 **Evaluation of performance in binary classification.**
Full-size DOI: 10.7717/peerj-cs.2414/fig-8

$N_{transactions}$ is the total number of transactions.
T_i is the duration of each transaction i.

Evaluation results

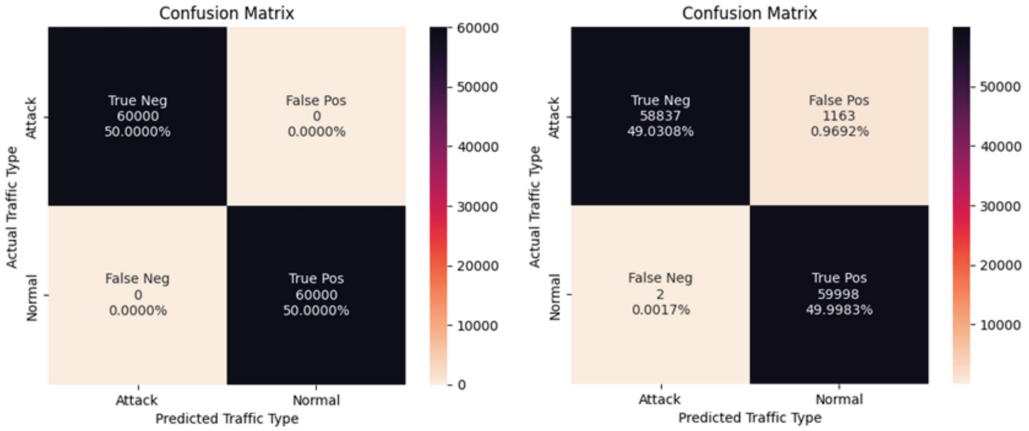
We utilize federated deep learning-based NIDS models to detect cyber-attacks in IoMT environments, specifically focusing on the networks of healthcare applications. Our training incorporates the most recent datasets for IDS, including CICIoT2023 and the Edge-IIoTset dataset. We conduct experiments employing binary and multi-class classification techniques for each dataset.

Binary classification

In this subsection, we present the evaluation results for binary classification scenarios using FL. In addition, we provide an evaluation of blockchain-based SSI.

1. **Federated learning evaluation results.** We employed 150,000 samples for both benign and attack instances in both datasets, ensuring a balanced dataset for a comprehensive and meaningful comparison. Remarkably, our model demonstrates impeccable performance, achieving perfect scores of 100% across all metrics for the Edge-IIoT dataset. In contrast, the results for the CICIoT2023 dataset remain highly promising, with an accuracy of 99.09%, indicating a low error rate in classifying both benign and malicious traffic. Furthermore, achieving a perfect precision of 100%, along with a recall of 98% and an F1-score of 99%, underscores the robust overall performance of the model, as shown in Fig. 8.

The classification performance of our model is depicted through the confusion matrix presented in Fig. 9, providing a concise summary of the model's accurate and erroneous



(a) EdgeIIoTset dataset. **(b)** CICIoT2023 dataset.

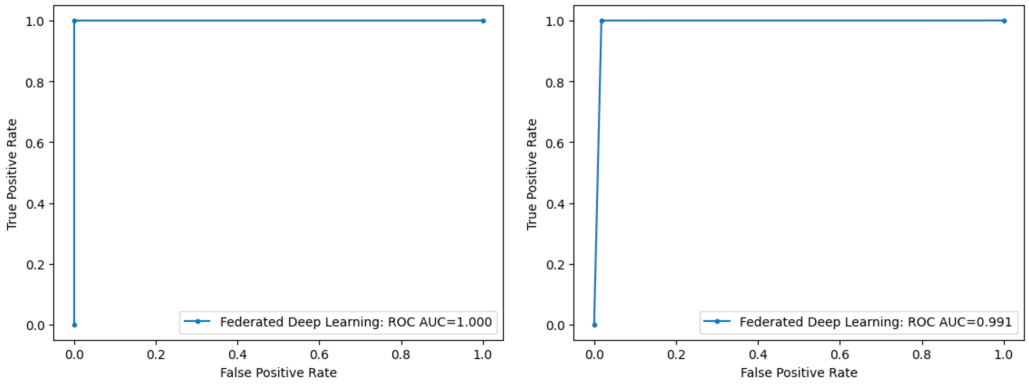
Figure 9 Confusion matrix in binary classification. Full-size DOI: 10.7717/peerj-cs.2414/fig-9

predictions The primary goal is to minimize both false positive and false negative rates, ensuring precise classification outcomes. Our proposed model effectively achieves this objective, exhibiting false positive and negative rates of 0% in the Edge-IIoTset dataset. For the CICIoT2023 dataset, we observe a negligible false negative rate of 0.0017%, alongside false positive rates of 0. 96%, which confirms the accuracy and efficiency of the model in mitigating classification errors.

Furthermore, the receiver operating characteristic (ROC) curve and the area under the curve (AUC) depicted in Fig. 10 offer a visual representation of our model's ability to distinguish between classes to highlight the model's effectiveness, we achieve an AUC of 99.1% for CICIIoT2023 notably 100% for the EdgeIIoTset dataset.

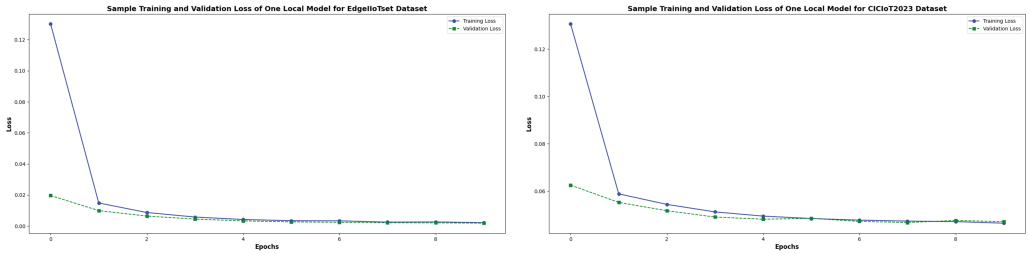
The training and validation loss curves for both the EdgeIIoTset and CICIoT2023 datasets demonstrate the promising performance of our federated learning model in binary classification. As illustrated in Fig. 11, the models exhibit rapid convergence within the initial epochs, followed by stable performance. The close alignment of training and validation loss curves, particularly in later epochs, indicates good generalization without significant overfitting. Both models achieve remarkably low final loss values (¡ 0.02) for training and validation sets, suggesting high predictive accuracy. The EdgeIIoTset model shows a slightly lower final loss, while the CICIoT2023 model displays smoother convergence between training and validation losses. These results collectively suggest that our approach effectively captures the underlying patterns in both datasets, promising strong performance on unseen data in real-world applications.

2. **SSI-based DID evaluation results.** As illustrated in Fig. 12 below both datasets exhibit similar startup durations (SD), with EdgeIIoTset demonstrating a marginally quicker performance by 0.01 s. Furthermore, the EdgeIIoTset dataset shows a shorter connect duration (CD) of 0.02 s compared to the CICIoT2023 dataset. Both datasets share the same publish duration (PD) of 9.15 s. However, in terms of issuing credential duration (ICD), the EdgeIIoTset dataset outperforms the CICIoT2023 dataset by 0.87 s. Regarding completed credential exchange duration (CCED), the EdgeIIoTset dataset exhibits a



(a) EdgeIIoTset dataset. **(b)** CICIoT2023 dataset.

Figure 10 **ROC curve AUC in binary classification.** Full-size DOI: 10.7717/peerj-cs.2414/fig-10



(a) EdgeIIoTset dataset. **(b)** CICIoT2023 dataset.

Figure 11 **Train and validation loss of one local model in binary classification.**
Full-size DOI: 10.7717/peerj-cs.2414/fig-11

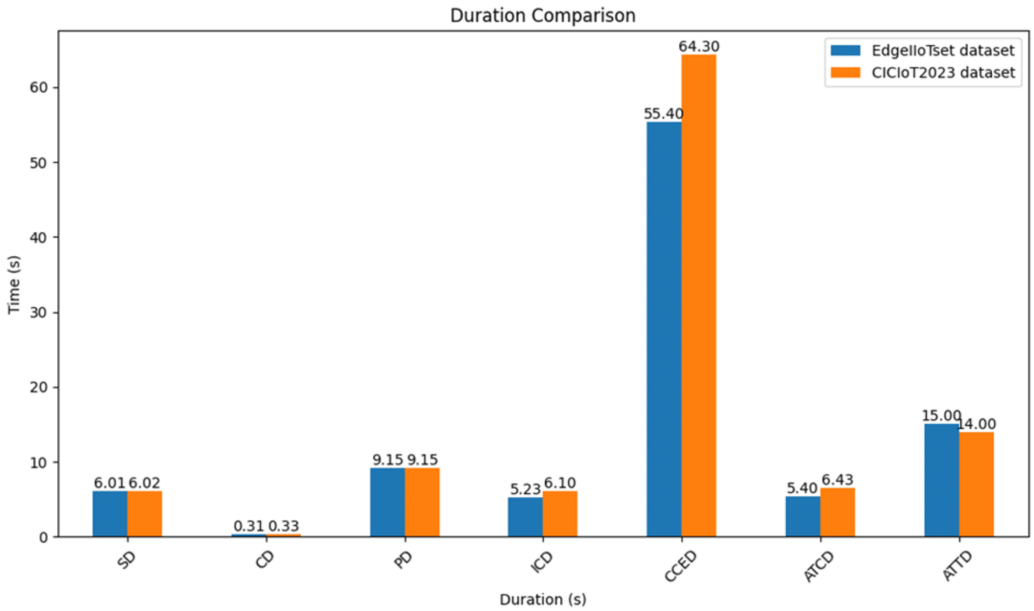


Figure 12 **Comparative performance analysis of blockchain-based SSI for binary classification.**
Full-size DOI: 10.7717/peerj-cs.2414/fig-12

Table 4 Classification report.

Dataset	Class	Precision	Recall	F1-score
CICIoT2023	Benign	83%	85%	84%
	DDoS	100%	100%	100%
	DoS	100%	100%	100%
	Mirai	100%	100%	100%
	Information gathering	81%	84%	83%
	Spoofing	87%	80%	84%
Edge-IIoTset	DoS/DDoS	91%	94%	92%
	Information gathering	88%	88%	88%
	Injection	77%	91%	84%
	Malware attacks	97%	50%	74%
	Man in the middle	100%	100%	100%
	Normal	100%	100%	100%

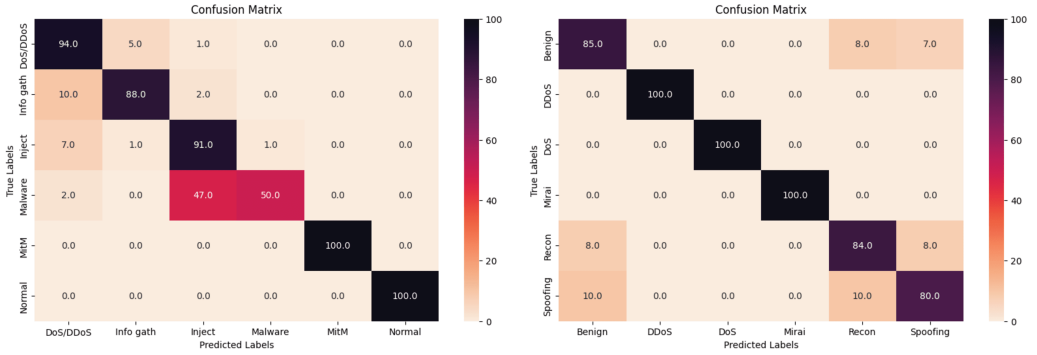
reduction of 8.9 s compared to the CICIoT2023 dataset. the CCED should normally be higher because this is the mutual authentication phase since the FS must authenticate all the nodes, and each node must authenticate the FS. Additionally, the average time per credential duration (ATCD) is shorter for the 'EdgeIIoTset' dataset in comparison to the 'CICIoT2023' dataset. Finally, the average time per transaction duration (ATTD) is marginally higher for the 'EdgeIIoTset' dataset when contrasted with the 'CICIoT2023' dataset.

Multiclass classification

In this subsection, we present the evaluation results for both the CICIoT2023 and EdgeIIoTset datasets in multiclass classification scenarios using FL. Additionally, we provide an evaluation of SSI-based DID.

1. **Federated learning evaluation results.** The evaluation of both the CICIoT2023 and Edge-IIoTset datasets reveals strong performance across diverse classes, as demonstrated in Table 4. Within the CICIoT2023 dataset, each class achieves good performance in most metrics. Notably, the DDoS, DoS, and Mirai attack classes in the CICIoT2023 dataset exhibit great classification capabilities, demonstrating perfect performance across all metrics with complete precision, recall, and F1-scores of 100%. Furthermore, the Spoofing, Benign, and Information Gathering classes show good precision with 87%, 83%, and 81%, respectively. However, their recall and F1 scores vary. The Benign class achieves a high recall of 85% and an F1-score of 84%. The Information Gathering class presents a relatively good recall of 84% and a corresponding F1-score of 83%. The Spoofing class achieves a moderate recall at 80% and an F1-score of 84%.

Transitioning to the Edge-IIoTset dataset Man in the Middle, and Normal classes also exhibit high performance, achieving perfect precision, recall, and F1-scores of 100%. followed by the DoS/DDoS class with high precision 91% and very high recall 99%, resulting in a strong F1-score of 92%. Furthermore, other classes show more variability.



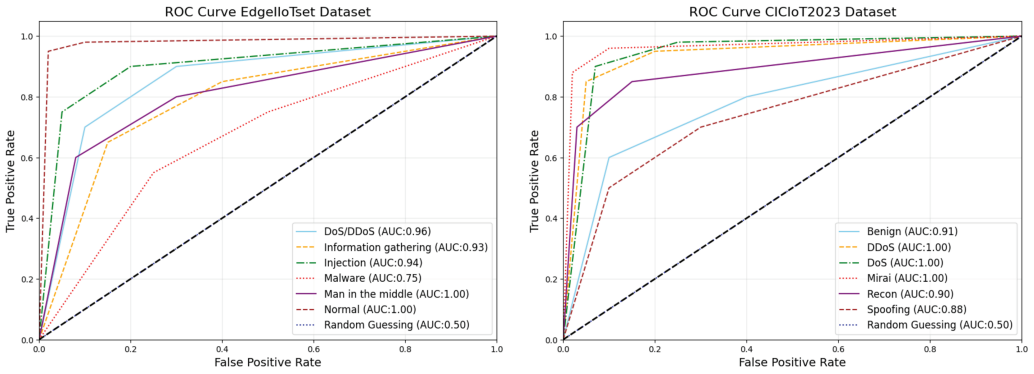
(a) EdgeIIoTset dataset. **(b)** CICIoT2023 dataset.

Figure 13 Confusion matrix in multiclass classification.
Full-size DOI: 10.7717/peerj-cs.2414/fig-13

The Information Gathering class has an excellent precision, recall, and f1-score with 88%. The Injection class shows a precision of 77% and a high recall of 91%, leading to an F1-score of 84%. The Malware Attacks class, despite its high precision of 97%, suffers from a low recall of 50%, resulting in a lower F1-score of 74%.

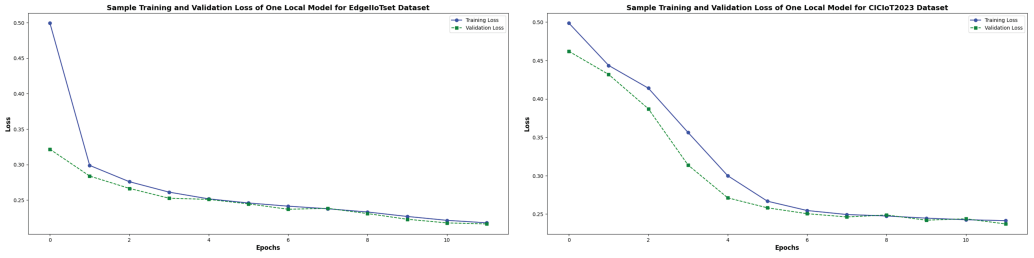
The confusion matrix depicted in Fig. 13 provides valuable insights into the prediction frequency for each class compared to their actual occurrences. Within the CICIoT dataset, remarkable performance was observed for classes such as DoS, DDoS, and Mirai, with the model accurately predicting all samples, achieving a flawless prediction rate of 100%. Additionally, the classification of Reconnaissance exhibited high accuracy, with 84% of samples correctly classified followed by Benign traffic with 85%. However, in subsequent classes such as Spoofing, accuracy decreased, with only 80% respectively, of the samples accurately classified. Likewise, within the Edge-IIoTset dataset, classes like Man in the Middle, and Normal traffic demonstrated robust performance, as the model accurately predicts all samples, resulting in a 100% prediction rate. The classification of the DoS/DDoS, Injection, and Information gathering class followed suit, with 94%, 91%, and 88% respectively of samples correctly classified. Nevertheless, as we delve into subsequent classes such as Malware, declined, with only 50%, respectively, of samples accurately classified.

The ROC curves and AUC values for both datasets demonstrate excellent detection capabilities across various attack types as presented in Fig. 14. In the EdgeIIoTset dataset, most attacks show very good performance with AUC values above 0.90, with “Man in the middle” and “Normal” attacks achieving perfect detection (AUC 1.00). While “Malware” has the lowest performance (AUC 0.75), other attacks like “DoS/DDoS”, “Information gathering”, and “Injection” exhibit strong performance (AUC 0.96-0.93-.94). The second CICIoT2023 dataset similarly shows outstanding results, with DDoS, DoS, and Mirai attacks reaching perfect detection (AUC 1.00). “Benign” and “Recon” categories perform very well (AUC 0.90-0.91), and even the lowest performing “Spoofing” category maintains good detection ability (AUC 0.88). Notably, all attack types in both datasets significantly



(a) EdgeIIoTset dataset. **(b)** CICIoT2023 dataset.

Figure 14 ROC Curve AUC in multiclass classification.
Full-size DOI: 10.7717/peerj-cs.2414/fig-14



(a) EdgeIIoTset dataset. **(b)** CICIoT2023 dataset.

Figure 15 Train and validation loss of one local model in multiclass classification.
Full-size DOI: 10.7717/peerj-cs.2414/fig-15

outperform the random guessing baseline, indicating robust and effective detection across the board.

Figure 15 illustrates the training and validation loss curves for both the EdgeIIoTset and CICIoT2023 datasets demonstrating the promising performance of our federated learning model in multi-class classification. For the EdgeIIoTset, we observe a rapid initial decrease in both training and validation loss, followed by a gradual convergence. The CICIoT2023 dataset shows a more gradual, consistent decrease in both losses across epochs. Importantly, neither dataset exhibits signs of overfitting, as the validation loss continues to decrease alongside the training loss, with only minimal divergence in later epochs. The EdgeIIoTset model achieves slightly lower final loss values (around 0.22) compared to the CICIoT2023 model (about 0.24), suggesting robust performance across different IoT datasets. The close alignment between training and validation losses, particularly in later epochs, indicates good generalization capabilities of our federated learning approach. These results suggest that our model effectively learns from local data without compromising privacy while maintaining strong predictive performance across diverse IoT classification tasks.

2. **SSI-based DID evaluation results.** As illustrated in Fig. 16 both datasets exhibit similar startup durations (SD), with EdgeIIoTset demonstrating a slight advantage of

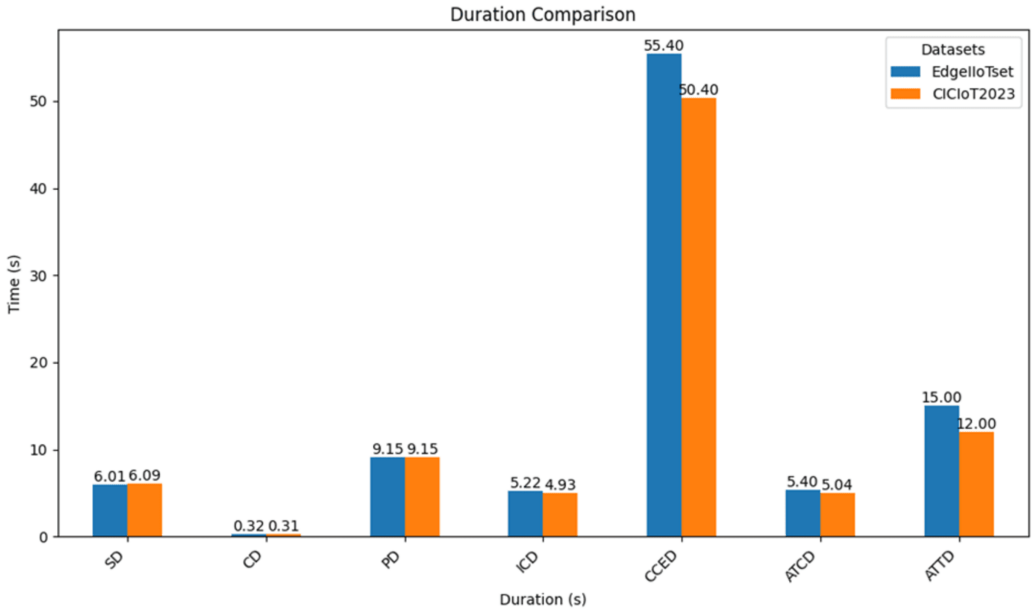


Figure 16 **Comparative performance analysis of blockchain-based SSI for multiclass classification.** Full-size DOI: 10.7717/peerj-cs.2414/fig-16

0.08 s. The CICIoT2023 dataset shows a shorter connect duration (CD) by 0.01 s compared to EdgeIIoTset. Additionally, both datasets share the same publish duration (PD) of 9.15 s. However, the CICIoT2023 dataset boasts a shorter issuing credential duration (ICD) of 0.29 s compared to EdgeIIoTset. Regarding the completion of the credential exchange duration (CCED), the CICIoT2023 dataset surpasses EdgeIIoTset by 5 s. The CCED should normally be higher because this is the mutual authentication phase since the FS must authenticate all the nodes, and each node must authenticate the FS. Furthermore, the CICIoT2023 dataset demonstrates a shorter average time per credential duration (ATCD) by 0.36 s compared to EdgeIIoTset. Lastly, the CICIoT2023 dataset shows a shorter average time per transaction duration (ATTD) by 3 s compared to the EdgeIIoTset.

In Table 5 we provide a comparison between the performance of our work with other FL-based state-of-the-art IDS. The proposed SA-FLIDS demonstrates superior performance compared to existing state-of-the-art FL-based IDS approaches. It achieves the highest accuracy of 100% for binary classification on the EdgeIIoTSet dataset with a standard deviation σ = 0.00%, outperforming previous methods. For multiclass classification, SA-FLIDS attains 93.48% accuracy and σ = 0.12% on EdgeIIoTSet and 92% and σ = 0.47% on CICIoT2023, surpassing earlier works. The model's consistently high performance across different datasets containing emerging new cyber attacks on IoT networks and classification tasks underscores its robustness and effectiveness in intrusion detection for IoT environments.

Limitation of the experimental design

1. **Scalability considerations:** Our experiments involved a relatively small number of fog nodes (clients K = 10). While this setup demonstrated the effectiveness of our approach,

Table 5 Comparisons between SA-FLIDS and State-of-the-art works. The bold values represent our method's performance.

FL-based IDS	Dataset	Classifier	Accuracy (%)
Begum et al. (2024)	EdgeIIoTset binary	BiLSTM	96
	EdgeIIoTset multiclass	BiLSTM	83
	EdgeIIoTset binary	CNN	85.31
	EdgeIIoTset multiclass	CNN	97
Baucas, Spachos & Plataniotis (2023)	Human activity recognition	CNN	91.75
Chatterjee & Hanawal (2021)	NSL-KDD	MLP	88
Schneble & Thamilarasu (2019)	PhysioNet	ANN	99
Ashraf et al. (2022)	Ba-IoT binary	ANN	99
Our	EdgeIIoTSet binary	LSTM	**100, (σ = 0.00)**
	EdgeIIoTSet multiclass		**93.48, (σ = 0.12)**
	CICIoT2023 binary		**99.12, (σ = 0.02)**
	CICIoT2023 multiclass		**92, (σ = 0.47)**

it may not fully represent the scalability challenges in larger, more complex IoMT networks. Future work should explore the performance and efficiency of SA-FLIDS in larger-scale deployments.

2. **IID data assumption:** Our current implementation assumes Independent and Identically Distributed (IID) data across clients. This assumption may not hold in all real-world scenarios, potentially impacting the model's performance in non-IID settings. Further investigation into non-IID data distributions is necessary.

Communication overhead and system performance

Our proposed architecture employs several strategies to minimize communication overhead and optimize system performance. The integration of IoMT devices and DIDs in the fog computing environment has been carefully designed to reduce network load:

1. **Data exchange optimization:** IoMT devices share data only during the initialization phase, with subsequent communications limited to updates. This approach significantly reduces the volume of data transferred across the network.
2. **Efficient authentication:** The system utilizes a session-based authentication mechanism. A token is generated once per session, eliminating the need for repeated DID authentications and thereby reducing associated overheads.
3. **Two-stage communication process:** The system operates in two distinct stages:
   (a) Initial authentication using VC and DID communication.
   (b) Subsequent update exchanges using the authenticated token.
   This separation ensures that DID communication and network protocol operations do not run concurrently, further optimizing resource usage.
4. **Blockchain utilization:** Blockchain's role is to provide authentication support only, handling DID and VC registration, verification, and revocation. It does not store FL

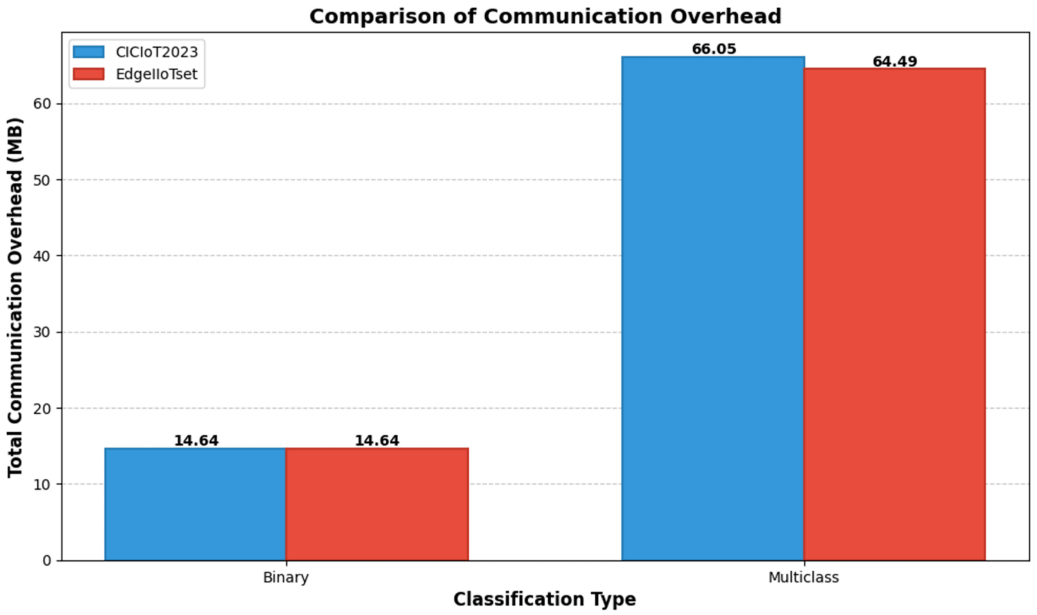


Figure 17 Comparison of communication overhead. Full-size DOI: 10.7717/peerj-cs.2414/fig-17

data. Consequently, its workload is limited to operations during the initialization phase (allocation of DIDs), the authentication phase (creation and verification of VCs), and the node revocation phase (revocation of VCs). Furthermore, no mining operations, which are known to consume a lot of energy and computing power, are carried out. The only situation requiring additional work is when a new node is added to the network, which significantly reduces its workload and associated overheads.

5. **Ensuring privacy and FL model implementation:** What's special about DIDs is that they operate on the same principle as SSL/TLS certificates. Obtaining a DID implies having a pair of public and private keys stored in a wallet, which is used to guarantee the confidentiality of exchanges.

Figure 17 illustrates the total communication overhead for binary and multiclass classification tasks using the CICIoT2023 and EdgeIIoTset datasets. The chart demonstrates that multiclass classification requires significantly higher communication overhead compared to binary classification for both datasets. Interestingly, while the overhead for binary classification is identical (14.64 MB) for both datasets, there is a slight difference in multiclass classification, with CICIoT2023 requiring marginally more overhead (66.05 MB) than EdgeIIoTset (64.49 MB). This comparison provides insights into the computational demands of different classification tasks and datasets in federated learning-based intrusion detection systems.

SECURITY AND PRIVACY ANALYSES

Our research focuses on detecting potential attacks on smart healthcare systems enabled by IoMT networks. The aim is to mitigate the risks associated with unauthorized access to sensitive patient data through malicious IoMT devices. Thus, in this section, we

Table 6 Comprehensive comparison between existing works and our model.

Models	FL	IDS	Environment	Security techniques of FL			
				Against adversarial	Secure communication	Blockchain	SSI authentication
Schneble & Thamilarasu (2019)	✓	✓	Medical CPS	X	X	X	X
Chatterjee & Hanawal (2021)	✓	✓	IoT network	X	X	X	X
Man et al. (2021)	✓	✓	IoT network	X	X	X	X
Rey et al. (2022)	✓	X	IoT network	✓	X	✓	X
Ruzafa-Alcázar et al. (2021)	✓	✓	Industrial IoT network	✓	✓	X	X
Zhao et al. (2019)	✓	X	General purpose	X	X	X	X
Friha et al. (2022)	✓	✓	Agriculture IoT	X	✓	✓	X
Ashraf et al. (2022)	✓	✓	Healthcare IoT network	✓	X	✓	X
Preuveneers et al. (2018)	✓	✓	Industrial IoT network	✓	X	✓	X
Lakhan et al. (2022)	✓	X	Healthcare IoT network	X	X	✓	X
OUR MODEL	✓	✓	Healthcare IoT network	✓	✓	✓	✓

thoroughly investigate the SA-FLIDS model's privacy and security features. It is important to note that without appropriate security measures, people may be reluctant to participate in healthcare applications, which ultimately hinders the success of these technological advancements. Our SA-FLIDS framework addresses these concerns by providing robust security measures, instilling confidence in users, and promoting widespread adoption and sustainability of healthcare technologies. Additionally, the analysis process is rooted in a theoretical exploration of SA-FLIDS's resilience against potential attacks outlined in the adversary model ("Detectin Process FL in SA-FLIDS system").

Table 6 provides a detailed comparative analysis between the existing systems and our proposed model. In this analysis, we examine the security commitments of the SA-FLIDS model compared to other FL-based IDS. We also compare the SA-FLIDS system with the broader context of security considerations in FL. Notably, the SA-FLIDS model advances it further by incorporating additional layers of security, such as user authentication and communication channels security during the FL process, by using a reliable aggregation technique. Furthermore, our comprehensive approach enhances the level of sensitive data protection and stands out as the only scheme incorporating SSI for user authentication in the FL process. This feature strengthens the system's security posture by ensuring authorized access and preventing unauthorized participation.

Data privacy and security analysis

SA-FLIDS leverages a fog-based NIDS powered by FL for data privacy-preserving. This allows the NIDS to effectively classify and detect malicious network traffic in real-time, protecting patient privacy and health data confidentiality. On the other hand, Also, the SA-FLIDS system deploys its security shield close to the source, which leads to a seamless granting of access to normal traffic while automatically blocking malicious intrusions, ensuring a secure and trustworthy healthcare environment.

Table 7 Summary of attack scenarios and countermeasures.

Adversarial attacks scenario on FL	Description	Countermeasures
Sybil attacks	Adversary creates multiple fake identities to disrupt the FL process.	Utilizes blockchain-based DIDs and VCs for unique identification and authentication of nodes, preventing fake identities.
Data poisoning attacks	Adversary injects malicious data to degrade global model performance.	Employs trimmed mean aggregation to minimize the influence of outliers and malicious data.
Eavesdropping and data tampering attacks	Unauthorized interception or modification of communication data.	Ensures secure communication using gRPC framework with TLS for end-to-end encryption.
Unauthorized access and authentication attacks	Adversaries gain unauthorized access by exploiting weak authentication.	Incorporates blockchain-based SSI technologies to ensure that only authenticated devices can participate in the FL process.

Moreover, our system ensures FL protection through the implementation of blockchain-based SSI technologies. These technologies play a crucial role in securing the system against authorization and privacy concerns using the DID and VC techniques. By leveraging the immutability and integrity features of blockchain, it becomes practically impossible for any entity to manipulate, replace, or falsify user identities stored on the blockchain.

Analysis of FL attacks

Table 7 provides adversarial attack scenarios targeting the FL process and how our models resist those attacks.

Analysis of cyber-attacks in healthcare systems

After applying the countermeasures, we achieve a secure and authenticated FL (SA-FL) system. This SA-FL system is then integrated into an IDS to identify and mitigate cyber-attacks on IoMT network traffic, specifically for healthcare applications. Resisting a range of attacks such as DoS, DDoS, Information Gathering, Web-Based Vulnerabilities, Communication Spoofing, Brute-Force, and Mirai IoT threats. The model ensures data security and integrity by preventing unauthorized access, maintaining fog server access, and safeguarding patient lives. By implementing robust security measures, SA-FLIDS promotes the sustainability of the smart healthcare system by fostering user adoption and confidence in healthcare technologies.

CONCLUSION AND FUTURE WORK

This article introduces SA-FLIDS, a Secure and Authenticated Federated Learning-based Network Intrusion Detection System designed for Fog-IoT-enabled smart healthcare systems. Our research aims to detect and counter cyber attacks on IoMT by harnessing fog computing capabilities. Additionally, we aim to preserve data privacy and reduce communication overhead, while addressing vulnerabilities like poisoning and Sybil attacks inherent in decentralized FL paradigms. We achieve this by employing a blockchain-based SSI model for client authentication and using trimmed mean aggregation in FL. In addition, secure communication transfer is ensured through TLS and gRPC protocols.

Performance evaluation demonstrates that SA-FLIDS not only detects attacks on the Internet of Medical Things (IoMT) but also meets criteria for privacy preservation, scalability, and sustainability. Furthermore, Our SA-FLIDS framework achieves high accuracy with negligible false positives and false negatives, particularly in binary classification scenarios. Our future endeavors will focus on evaluating the performance of our proposed model across various domains of IoT applications. Additionally, we aim to explore the application of FL with non-distributed IID data distributions.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

Leandros Maglaras is an Academic Editor for PeerJ.

Author Contributions

- Radjaa Bensaid conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Nabila Labraoui conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Ado Adamou Abba Ari performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Hafida Saidi analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Joel Herve Mboussam Emati analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Leandros Maglaras conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The source code (complete implementation of the SA-FLIDS framework, including the Federated Learning process, the blockchain-based Self-Sovereign Identity (SSI) component, and the necessary scripts for, model training, and evaluation) for the SA-FLIDS framework is available at Zenodo: teamflssi. (2024). teamflssi/Secure-and-Authenticated-Federated-Learning-based-intelligent-NIDS-for-smart-healthcare: v1.4 (v1.3). Zenodo. https://doi.org/10.5281/zenodo.11316425.

The raw data used for experimental results is available at Zenodo: STIC Laboratory. (2024). preprocessed datasets. Zenodo. https://doi.org/10.5281/zenodo.11315294.

The third-party datasets are available at:

- CICIoT2023 dataset: (Neto, Dadkhah, and Ferreira 2023), https://www.unb.ca/cic/datasets/iotdataset-2023.html.

- Edge-IIoTset dataset: (*Ferrag et al., 2022*), https://www.kaggle.com/datasets/mohamedamineferrag/edgeiiotset-cyber-security-dataset-of-iot-iiot.

REFERENCES

Ali M, Karimipour H, Tariq M. 2021. Integration of blockchain and federated learning for internet of things: recent advances and future challenges. *Computers & Security* **108(5)**:102355 DOI 10.1016/j.cose.2021.102355.

Ashraf E, Areed NF, Salem H, Abdelhay EH, Farouk A. 2022. Fidchain: federated intrusion detection system for blockchain-enabled iot healthcare applications. *Healthcare* **10(6)**:1110 DOI 10.3390/healthcare10061110.

Banerjee A, Dutta B, Mandal T, Chakraborty R, Mondal R. 2022. Blockchain in iot and beyond: case studies on interoperability and privacy. In: *Blockchain based Internet of Things*. Cham: Springer, 113–138.

Baucas MJ, Spachos P, Plataniotis KN. 2023. Federated learning and blockchain-enabled fog-iot platform for wearables in predictive healthcare. *IEEE Transactions on Computational Social Systems* **10(4)**:1732–1741 DOI 10.1109/TCSS.2023.3235950.

Begum K, Mozumder MAI, Joo M-I, Kim H-C. 2024. BFLIDS: blockchain-driven federated learning for intrusion detection in iomt networks. *Sensors* **24(14)**:4591 DOI 10.3390/s24144591.

Benfriha S, Labraoui N, Bensaid R, Bany Salameh H, Saidi H. 2023. Fuba: a fuzzy-based unmanned aerial vehicle behaviour analytics for trust management in flying ad-hoc networks. *IET Networks* **13(3)**:n/a DOI 10.1049/ntw2.12108.

Bensaid R, Labraoui N, Abba Ari AA, Maglaras L, Saidi H, Abdu Lwahhab AM, Benfriha S. 2024. Toward a real-time TCP SYN flood DDoS mitigation using adaptive neuro-fuzzy classifier and SDN assistance in fog computing. *Security and Communication Networks* **2024(1)**:6651584.

Chatterjee S, Hanawal MK. 2021. Federated learning for intrusion detection in iot security: a hybrid ensemble approach. Arxiv preprint DOI 10.48550/arXiv.2106.15349.

Djenne A, Saïdouni DE. 2018. Cyber attacks classification in iot-based-healthcare infrastructure. In: *2018 2nd Cyber Security in Networking Conference (CSNet)*. Piscataway: IEEE, 1–4.

Elayan H, Aloqaily M, Guizani M. 2021. Deep federated learning for iot-based decentralized healthcare systems. In: *2021 International Wireless Communications and Mobile Computing (IWCMC)*. Piscataway: IEEE, 105–109.

Emati JHM, Mboussam HP, Tchendji VK. 2023. Feasibility study of improving blockchain-based self-sovereign identity security using artificial intelligence and lightweight cryptography. In: *2023 IEEE AFRICON*. Nairobi, Kenya, 1–3.

Ferrag MA, Friha O, Hamouda D, Maglaras L, Janicke H. 2022. Edge-IIoTset: a new comprehensive realistic cyber security dataset of iot and iiot applications for centralized and federated learning. *IEEE Access* **10**:40281–40306 DOI 10.1109/ACCESS.2022.3165809.

Figueroa-Lorenzo S, Benito JA, Arrizabalaga S. 2021. Modbus access control system based on ssi over hyperledger fabric blockchain. *Sensors* **21(16)**:5438 DOI 10.3390/s21165438.

Friha O, Ferrag MA, Shu L, Maglaras L, Choo K-KR, Nafaa M. 2022. Felids: federated learning-based intrusion detection system for agricultural internet of things. *Journal of Parallel and Distributed Computing* **165(15)**:17–31 DOI 10.1016/j.jpdc.2022.03.003.

Gamblin J. 2017. Mirai botnet. *Available at https://github.com/jgamblin/Mirai-Source-Code*.

Iwendi C, Anajemba JH, Biamba C, Ngabo D. 2021. Security of things intrusion detection system for smart healthcare. *Electronics* **10(12)**:1375 DOI 10.3390/electronics10121375.

Jensen M, Gruschka N, Herkenhöner R. 2009. A survey of attacks on web services. *Computer Science—Research and Development* **24(4)**:185–197 DOI 10.1007/s00450-009-0092-6.

Kumari A, Tanwar S, Tyagi S, Kumar N. 2018. Fog computing for healthcare 4.0 environment: opportunities and challenges. *Computers & Electrical Engineering* **72(5)**:1–13 DOI 10.1016/j.compeleceng.2018.08.015.

Lakhan A, Mohammed MA, Nedoma J, Martinek R, Tiwari P, Vidyarthi A, Wang W. 2022. Federated-learning based privacy preservation and fraud-enabled blockchain iomt system for healthcare. *IEEE Journal of Biomedical and Health Informatics* **27(2)**:664–672 DOI 10.1109/JBHI.2022.3165945.

Lee D, Yoon SN. 2021. Application of artificial intelligence-based technologies in the healthcare industry: opportunities and challenges. *International Journal of Environmental Research and Public Health* **18(1)**:271 DOI 10.3390/ijerph18010271.

Lian Z, Zhang C, Nan K, Su C. 2023. Spoil: sybil-based untargeted data poisoning attacks in federated learning. In: *International Conference on Network and System Security*. Cham: Springer Nature Switzerland, 235–248.

Lim WYB, Luong NC, Hoang DT, Jiao Y, Liang Y-C, Yang Q, Niyato D, Miao C. 2020. Federated learning in mobile edge networks: a comprehensive survey. *IEEE Communications Surveys & Tutorials* **22(3)**:2031–2063 DOI 10.1109/COMST.2020.2986024.

Man D, Zeng F, Yang W, Yu M, Lv J, Wang Y. 2021. Intelligent intrusion detection based on federated learning for edge-assisted internet of things. *Security and Communication Networks* **2021**:1–11 DOI 10.1155/2021/9361348.

Manoj T, Makkithaya K, Narendra V. 2022. A blockchain based decentralized identifiers for entity authentication in electronic health records. *Cogent Engineering* **9(1)**:2035134 DOI 10.1080/23311916.2022.2035134.

Mboussam Emati JH, Mboussam HP. 2023. A block mining based machine learning scheme for data integrity in blockchain based iot solutions. In: *2023 IEEE AFRICON*. 1–6.

McMahan B, Moore E, Ramage D, Hampson S, Arcas BA. 2017. Communication-efficient learning of deep networks from decentralized data. In: *Artificial intelligence and statistics*. Proceedings of Machine Learning Research, 1273–1282.

Mondejar ME, Avtar R, Diaz HLB, Dubey RK, Esteban J, Gómez-Morales A, Hallam B, Mbungu NT, Okolo CC, Prasad KA, She Q, Garcia-Segura S. 2021. Digitalization to achieve sustainable development goals: steps towards a smart green planet. *Science of the Total Environment* **794**:148539 DOI 10.1016/j.scitotenv.2021.148539.

Möller B, Karlsson M, Antelius F, van den Berg T, Wood D. 2022. Hla 4 federate protocol-requirements and solutions. In: *2022 Simulation Innovation Workshop*.

Neto ECP, Dadkhah S, Ferreira R, Zohourian A, Lu R, Ghorbani AA. 2023. CICIoT2023: a real-time dataset and benchmark for large-scale attacks in IoT environment. *Sensors* **23(13)**:5941 DOI 10.3390/s23135941.

Preuveneers D, Rimmer V, Tsingenopoulos I, Spooren J, Joosen W, Ilie-Zudor E. 2018. Chained anomaly detection models for federated learning: an intrusion detection case study. *Applied Sciences* **8(12)**:2663 DOI 10.3390/app8122663.

Qu Y, Gao L, Luan TH, Xiang Y, Yu S, Li B, Zheng G. 2020. Decentralized privacy using blockchain-enabled federated learning in fog computing. *IEEE Internet of Things Journal* **7(6)**:5171–5183 DOI 10.1109/JIOT.2020.2977383.

Radjaa B, Nabila L, Salameh HB. 2023. Federated deep learning-based intrusion detection approach for enhancing privacy in fog-iot networks. In: *2023 10th International Conference on Internet of Things: Systems, Management and Security (IOTSMS)*. Piscataway: IEEE, 156–160.

Rey V, Sánchez PMS, Celdrán AH, Bovet G. 2022. Federated learning for malware detection in iot devices. *Computer Networks* **204(7)**:108693 DOI 10.1016/j.comnet.2021.108693.

Ruzafa-Alcázar P, Fernández-Saura P, Mármol-Campos E, González-Vidal A, Hernández-Ramos JL, Bernal-Bernabe J, Skarmeta AF. 2021. Intrusion detection based on privacy-preserving federated learning for the industrial IoT. *IEEE Transactions on Industrial Informatics* **19**:1 DOI 10.1109/TII.2021.3126728.

Saidi H, Labraoui N, Ari AAA. 2022. A secure health monitoring system based on fog to cloud computing. *International Journal of Medical Engineering and Informatics* **1(1)**:1 DOI 10.1504/IJMEI.2022.10050253.

Saidi H, Labraoui N, Ari AAA, Maglaras LA, Emati JHM. 2022. Dsmac: privacy-aware decentralized self-management of data access control based on blockchain for health data. *IEEE Access* **10(3)**:101011–101028 DOI 10.1109/ACCESS.2022.3207803.

Sarhan M, Layeghy S, Moustafa N, Portmann M. 2023. Cyber threat intelligence sharing scheme based on federated learning for network intrusion detection. *Journal of Network and Systems Management* **31(1)**:3 DOI 10.1007/s10922-022-09691-3.

Schneble W, Thamilarasu G. 2019. Attack detection using federated learning in medical cyberphysical systems. In: *28th International Conference on Computer Communications and Networks (ICCCN)*. Piscataway: IEEE, 1–8.

Sindhusaranya B, Yamini R, Manimekalai M, Geetha K. 2023. Federated learning and blockchain-enabled privacy-preserving healthcare 5.0 system: a comprehensive approach to fraud prevention and security in iomt. *Journal of Internet Services and Information Security* **13(4)**:199–209 DOI 10.58346/JISIS.2023.I4.014.

Stiawan D, Idris M, Malik R, Nurmaini S, Alsharif N, Budiarto R. 2019. Investigating brute force attack patterns in iot network. *Journal of Electrical and Computer Engineering* **2019**:4568368 DOI 10.1155/2019/4568368.

Thomas AM, Ramaguru R, Sethumadhavan M. 2022. Distributed identity and verifiable claims using Ethereum standards. In: *Proceeding of the Inventive Communication Computional Technologies (ICICCT)*. Vol. 311, 621–636.

van der Merwe J, Zubizarreta X, Luk cin I, Rugamer A, Felber W. 2018. Classification of spoofing attack types. In: *2018 European Navigation Conference (ENC)*. Gothenburg, Sweden, 91–99.

Wang R, Lai J, Zhang Z, Li X, Vijayakumar P, Karuppiah M. 2022. Privacy-preserving federated learning for internet of medical things under edge computing. *IEEE Journal of Biomedical and Health Informatics* **27(2)**:854–865 DOI 10.1109/JBHI.2022.3157725.

Wu Q, Chen X, Zhou Z, Zhang J. 2020. Fedhome: cloud-edge based personalized federated learning for in-home health monitoring. *IEEE Transactions on Mobile Computing* **21(8)**:2818–2832 DOI 10.1109/TMC.2020.3045266.

Zhao Y, Chen J, Wu D, Teng J, Yu S. 2019. Multi-task network anomaly detection using federated learning. In: *Proceedings of the Tenth International Symposium on Information and Communication Technology*. 4–6.