

Digital Twin Knowledge Distillation for Federated Semi-Supervised Industrial IoT DDoS Detection

Mohammed Ayalew Belay^{*✉}, Adil Rasheed^{†✉}, Pierluigi Salvo Rossi^{*‡✉}

^{*} Department of Electronic Systems, Norwegian University of Science and Technology, Trondheim, Norway

[†] Department of Engineering Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway

[‡] Department of Gas Technology, SINTEF Energy Research, Trondheim, Norway

Emails: mohammed.a.belay@ntnu.no, adil.rasheed@ntnu.no, salvorossi@ieee.org

Abstract—The widespread deployment of the Internet of Things (IoT) has transformed industries by enabling automation and connectivity across various application domains. However, this expansion has also introduced significant security challenges, such as distributed denial of service (DDoS) attacks, which can overwhelm network resources, causing service disruptions and financial losses. Traditional machine learning methods for DDoS detection often struggle with limited labeled data, high false alarm rates, and privacy concerns associated with centralized data processing. To address these limitations, we propose DTKD-Fed, a novel semi-supervised DDoS detection framework that integrates digital twin technology with federated knowledge distillation. By leveraging digital twins as virtual replicas of IoT devices, the proposed method enables continuous learning and decentralized model training without requiring labeled data or sharing raw data. The DTKD-Fed framework enhances real-time anomaly detection and mitigates DDoS attacks while maintaining data privacy. Experimental results using publicly available IoT datasets demonstrate the effectiveness of achieving high detection accuracy, competitive performance, and scalability in real-world environments.

Index Terms—Anomaly detection, Industrial IoT, Digital twins, Knowledge Distillation, Federated learning.

I. INTRODUCTION

THE Internet of Things (IoT) has significantly advanced connectivity and automation across various sectors, including industrial systems, healthcare, agriculture, and transportation [1]–[7]. With over 12 billion connected devices today and projections of 500 billion by 2030, the IoT rapid expansion presents significant security challenges, particularly distributed denial of service (DDoS) attacks [8], [9]. These attacks exploit vulnerabilities in IoT devices, which often lack strong security due to their limited computational and energy capacities [10], [11]. Recently, several statistical and deep learning methods are proposed for IoT anomaly detection in supervised, semi-supervised, and unsupervised learning methods [12]–[15]. Conventional DDoS detection systems face challenges such as limited labeled data and high false alarm rates, limiting their real-world effectiveness [16], [17].

Digital twins are virtual replicas of physical systems and provide an alternative solution for IoT security. They enable real-time monitoring, simulation, and analysis of IoT devices,

providing valuable insights into their operational states and potential vulnerabilities [18], [19]. In DDoS detection, digital twins simulate attack scenarios and monitoring devices in real time. Digital twin can help predict, detect, and respond to security threats more efficiently [20], [21]. However, conventional DDoS detection are also mostly centralized detection systems and face challenges in monitoring the vast data flows generated by a plethora of IoT devices. These centralized systems create a single point of failure, making them appealing targets for attackers. Furthermore, transmitting data to a central server for detection compromises user privacy, as sensitive information may be exposed during transit or while stored. To address these issues, federated learning (FL) has emerged as a decentralized approach, enabling collaborative model training without sharing raw data, thus preserving privacy [22]–[25]. FL is particularly advantageous in IoT environments, allowing robust DDoS detection systems to be developed across multiple devices while maintaining data privacy [26], [27].

II. RELATED WORK AND CONTRIBUTIONS

The demand for efficient, scalable, and robust DDoS detection mechanisms has emerged as a critical focus in IoT security research, driving the exploration of cutting-edge approaches such as federated learning and digital twin technologies. Recently, several federated learning-based DDoS detection frameworks have been proposed [28]–[30]. In addition, several studies have explored digital twin-based federated learning frameworks. Su *et al.* [31] proposed an intrusion detection method that integrates federated learning with LSTM models within a Digital Twin Network (DTN). Yigit *et al.* [32] introduced a digital twin-enabled DDoS detection architecture for ISP core networks using online learning techniques. De Benedictis *et al.* [33] proposed a conceptual architecture for anomaly detection in industrial IoT (IIoT) systems, combining digital twins with autonomic computing paradigms. Gupta *et al.* [34] introduced a hierarchical federated learning (HFL) model for anomaly detection in Vehicular IoT (V-IoT), while Gupta *et al.* [35] developed a digital twin-based hierarchical federated anomaly detection framework for smart healthcare applications.

Despite these advancements, existing digital twin-based federated IoT DDoS detection frameworks face several challenges. Many rely on supervised learning models that require

This work was partially supported by the Research Council of Norway under the project DIGITAL TWIN within the PETROMAKS2 framework (project nr. 318899).

labeled data for training, and most focus solely on digital twin models without adequately integrating them with real-world data. Additionally, few approaches address the communication overhead in federated learning using digital twin. To the best of our knowledge, no previous work has explored semi-supervised, digital twin-based knowledge distillation for federated learning in the context of IoT DDoS attack detection.

III. THE PROPOSED METHOD

A. Digital Twin-based Federated Learning Framework

We propose Digital Twin-based Knowledge Distillation Federated Learning (DTKD-Fed), a federated and semi-supervised framework for IoT network DDoS attack detection. The DTKD-Fed framework enables federated learning, where multiple edge devices, each with their own local data, collaboratively train a shared global model while preserving data privacy. Let $\mathcal{D}_{\text{twin}}$ and $\mathcal{D}_{\text{real}}$ represent the datasets from the digital twin and real systems, respectively. We first train digital twin-based model based on synthetic labeled data. The digital twin data is used to minimize the binary cross-entropy loss function ℓ_{BCE} through mini-batch stochastic gradient descent as follows:

$$\Theta^{(0)} = \arg \min_{\Theta} \frac{1}{|\mathcal{D}_{\text{twin}}|} \sum_{i=1}^{|\mathcal{D}_{\text{twin}}|} \ell_{\text{BCE}}(\Theta; \mathcal{D}_{\text{twin}}) \quad (1)$$

where $\Theta^{(0)}$ represent the parameters of the digital twin model and $|\mathcal{D}_{\text{twin}}|$ denotes the number of samples in the digital twin data.

The federated approach involves K edge devices, with each device k owning a local dataset $\mathcal{D}_k \subseteq \mathcal{D}_{\text{real}}$. The objective is to train a global model with parameters Θ by optimizing the following objective function:

$$\Theta^* = \arg \min_{\Theta} \frac{1}{\sum_{k=1}^K n_k} \sum_{k=1}^K \sum_{l=1}^{n_k} \ell(\Theta; d_{k,j}) \quad (2)$$

Here, Θ^* is parameter of global model after training, $\ell(\Theta; d_{k,j})$ represents the loss function evaluated at the global model parameters Θ for the j -th data sample $d_{k,j}$ from the local dataset $\mathcal{D}_k$ of the k -th edge device. The federated process proceeds in rounds, where the server selects a subset of edge devices to participate, and each edge device updates its local model based on the current global model parameters.

B. Digital Twin Knowledge Distillation

In the DTKD-Fed framework, a digital twin model is trained on synthetic IoT DDoS attack data. The digital twin generates soft targets, which are used to distill knowledge into local models on the edge devices. During local training, each edge device initializes its local model with the global model parameters $\Theta^{(t)}$ and trains it using its own data $\mathcal{D}_k$. The local training objective is to minimize the Kullback-Leibler (KL) divergence loss between the soft targets from the digital twin and the predictions of the local model.

Let $p_{\text{twin}}(y|\mathbf{x}_i)$ represent the soft target probabilities from the digital twin model for the input $\mathbf{x}_i$, and $p_k(y|\mathbf{x}_i)$ represent

the predictions from the local model on edge device k . The KL divergence loss is defined as:

$$\ell_{\text{KL}} = D_{\text{KL}}(p_{\text{twin}}(y|\mathbf{x}_i) || p_k(y|\mathbf{x}_i)) \quad (3)$$

The total KL loss for an edge device k on local data $\mathcal{D}_k$ is given by:

$$\mathcal{L}_{\text{KL}} = -\frac{1}{|\mathcal{D}_k|} \sum_{i=1}^{|\mathcal{D}_k|} \left[p_{\text{twin}}(y=1|\mathbf{x}_i) \log \left(\frac{p_k(y=1|\mathbf{x}_i)}{p_{\text{twin}}(y=1|\mathbf{x}_i)} \right) + p_{\text{twin}}(y=0|\mathbf{x}_i) \log \left(\frac{p_k(y=0|\mathbf{x}_i)}{p_{\text{twin}}(y=0|\mathbf{x}_i)} \right) \right] \quad (4)$$

The KL divergence measures how well the local model predictions match the digital twin soft targets, allowing the local models to gradually mimic the behavior of the digital twin model. This approach enhances local model training without requiring labeled data and enables real-time learning and updates in the federated framework.

C. Federated Learning with Knowledge Distillation

After each round of local training, edge devices send their updated model parameters to the central server. The server aggregates the parameters from all participating devices using federated averaging, resulting in a refined global model:

$$\Theta^{(t+1)} = \frac{1}{K} \sum_{k=1}^K \Theta_k^{(t+1)} \quad (5)$$

The new global model is then distributed to the edge devices for the next round of training. By integrating digital twin knowledge distillation into this federated learning process, the DTKD-Fed framework effectively reduces the communication overhead and computational load on the edge devices while improving model performance.

IV. EXPERIMENTAL SETUP

A. Datasets

We performed our experiments using datasets from an Industry 4.0 production line system and its digital twin under cyberattack [36]. Data was recorded during both normal operation and under attack conditions, and label information is provided. The dataset contains 58 features that represent sensor readings, actuator statuses, timing information, throughput time, and PLC parameters for the four stations.

B. Implementation Details and Evaluation Metrics

The digital twin and local models employ a simple neural network architecture with an input layer matching the feature size, an 8-neuron hidden layer with ReLU activation, and a single output neuron. Initial training on digital twin data is conducted with a mini-batch size of 32, a learning rate of 0.001, and 10 epochs to initialize the global model parameters. Each client device trains a similar model locally using the DTKD-Fed algorithm. In federated learning rounds, a fraction

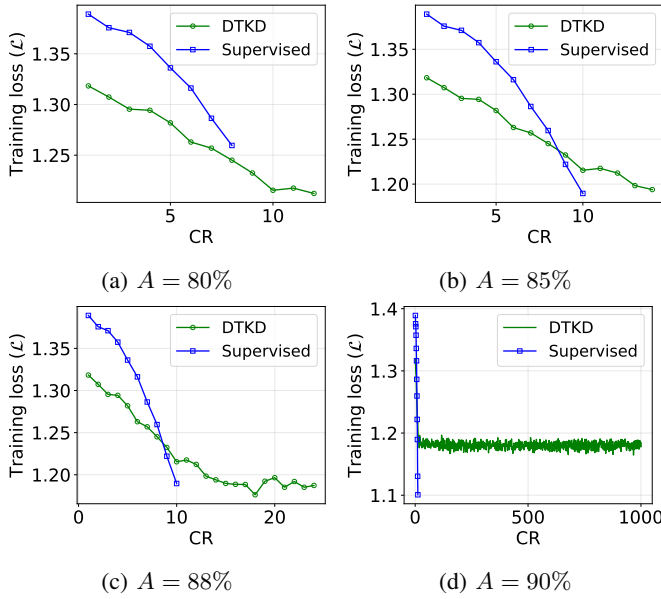


Fig. 1: Convergence analysis of DTKD-Fed for different target accuracy and fixed local batch size of $B = 10$, local epochs $E = 2$, client fraction $C = 0.2$.

of clients (C) is selected to perform local updates over epochs (E) with a mini-batch size (B), refining the global model.

Model performance is evaluated using accuracy, precision, recall, F1-score, and AUC, and training continues until the target accuracy or maximum rounds are reached.

V. RESULTS AND DISCUSSIONS

A. Convergence Analysis

We analyzed the convergence of the proposed DTKD-Fed model and compared it with a supervised, non-digital twin-based model by plotting communication rounds (CR) against training loss for varying target accuracies, as shown in Figure 1. Using fixed local epochs (E), batch sizes (B), and client fractions ($C = 0.1$), the DTKD-Fed model demonstrated faster convergence at lower target accuracies due to its ability to distill meaningful insights via digital twin-based knowledge distillation, reducing the required communication rounds. However, at a higher target accuracy ($A = 90\%$, Figure 2d), DTKD-Fed exhibited early stagnation in training loss, while the supervised model continued to improve.

B. Local Computation Analysis

We analyzed the effects of varying local epochs and mini-batch sizes on DDoS detection performance for different target accuracy. Figures 2 and 3 illustrate the trade-offs between communication rounds (CR) and local computation. Figure 2 shows the impact of increasing local epochs ($E = 2, 4, 6, 8$) with a fixed batch size ($B = 10$) and client fraction ($C = 0.2$). Increasing local epochs reduces the number of communication rounds required to achieve the target accuracy for both DTKD-Fed and supervised models. DTKD-Fed benefits significantly,

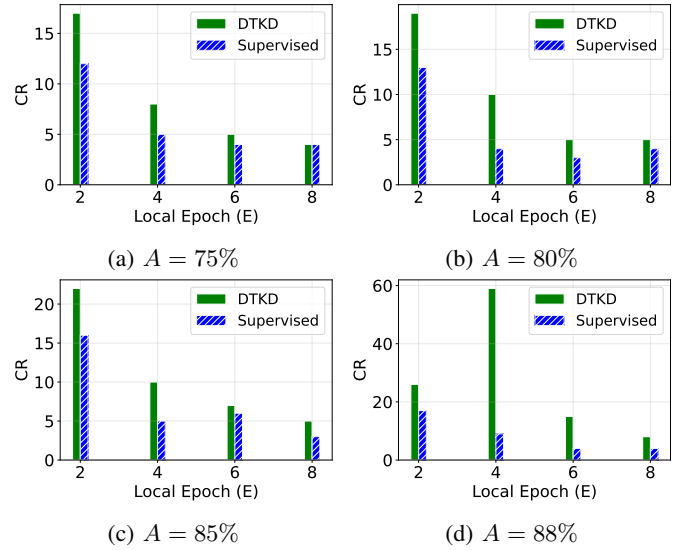


Fig. 2: Local epoch analysis of DTKD-Fed for different target accuracy and fixed local batch size of $B = 10$ and client fraction $C = 0.2$.

as additional local computation enhances the effectiveness of digital twin-based knowledge distillation, reducing the need for frequent communication rounds. Figure 3 presents the effect of varying batch sizes (B) on communication rounds, with local epochs fixed at $E = 2$ and client fractions $C = 0.2$. Smaller batch sizes lead to faster convergence, requiring fewer communication rounds to reach the desired accuracy for both models. Overall, the DTKD-Fed model effectively leverages knowledge distillation to offset the need for labeled data, demonstrating that increased local computation can maintain competitive performance while reducing reliance on extensive labeled datasets in IoT environments.

C. Scalability Analysis

The scalability of the DTKD-Fed model was evaluated by varying the fraction of active systems (C), as shown in Figure 5. The number of communication rounds (CR) required to achieve the target accuracy was measured for different client fractions, with the experiments conducted using a fixed number of clients ($K = 20$), local epochs ($E = 1$), and batch size ($B = 20$). The results for target accuracy of $A = 80\%$ and $A = 88\%$ are discussed below. The DTKD-Fed model demonstrates strong scalability as the fraction of active systems increases. While the supervised model benefits slightly more from a larger number of active clients, DTKD-Fed remains efficient, achieving comparable performance with a minimal increase in communication rounds, even without access to labeled data.

D. Global Model Performance

We evaluated the global model performance of DTKD-Fed and the supervised model using metrics such as accuracy, F1-score, and AUROC, with a fixed client fraction ($C = 0.1$),

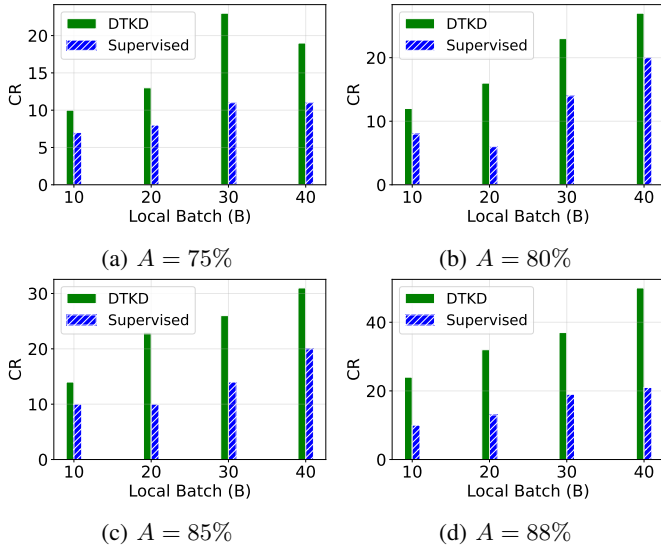


Fig. 3: Local batch analysis of DTKD-Fed for different target accuracy and fixed local epochs $E = 2$ and client fraction $C = 0.2$.

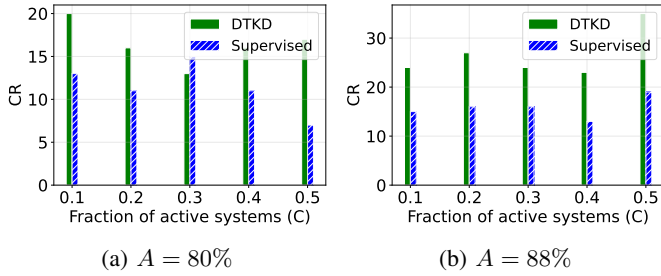


Fig. 4: Scalability analysis for different target accuracy.

local epochs ($E = 2$), and mini-batch size ($B = 10$). Figure 5 presents the results for different target accuracies. The supervised model achieves faster convergence due to its reliance on labeled data, particularly at lower communication rounds. However, DTKD-Fed progressively catches up, demonstrating its capability to perform well without labeled data. At higher target accuracy ($A = 88\%$), DTKD-Fed requires more communication rounds to match the supervised model but remains competitive by leveraging digital twin knowledge. Overall, DTKD-Fed achieves comparable accuracy, F1-score, and AUROC values with sufficient communication rounds, validating the effectiveness of digital twin knowledge distillation.

VI. CONCLUSIONS AND FUTURE WORK

In this study, we introduced the DTKD-Fed framework, a semi-supervised digital twin-based federated learning approach for DDoS attack detection in IoT environments. Through various analyses, including convergence, local computation, scalability, and global model performance, we demonstrated the effectiveness of digital twin-based knowledge distillation in achieving comparable performance to supervised models, even in the absence of labeled data. The

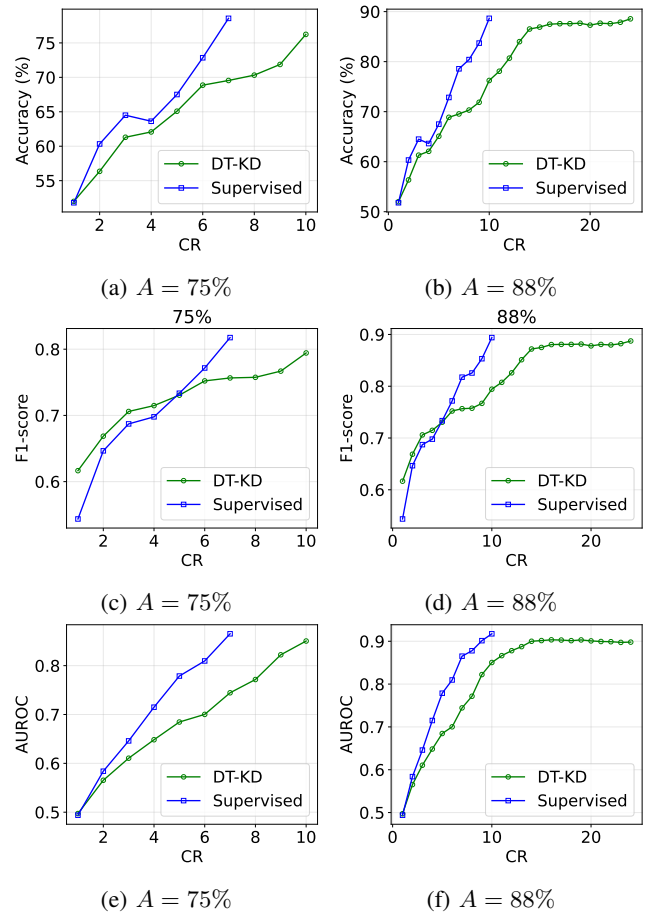


Fig. 5: DDoS attack detection accuracy, F1-score and AUROC for different target accuracy

DTKD-Fed model showcased strong potential in reducing communication rounds, leveraging local computation to improve learning efficiency, and maintaining robust scalability as the fraction of active clients increased. For future study, several areas for improvement and exploration remain. First, enhancing the fidelity of the digital twin model to better capture the complex real-world IoT environments could further improve performance, particularly at higher target accuracies where the model currently stagnates. Additionally, investigating privacy-preserving techniques, such as differential privacy, in conjunction with digital twin models could enhance the security of federated learning in sensitive IoT environments. In general, DTKD-Fed presents a promising approach to addressing the challenges of data privacy and limited labeled data in IoT networks.

ACKNOWLEDGMENT

This work was partially supported by the Research Council of Norway under the project DIGITAL TWIN within the PETROMAKS2 framework (project nr. 318899).

REFERENCES

- [1] L. Atzori, A. Iera, G. Morabito, [The Internet of Things: A survey](#), *Computer Networks*, 54:2787–2805 (2010).
- [2] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, [Internet of Things \(IoT\): A vision, architectural elements, and future directions](#), *Future Generation Computer Systems*, 29:1645–1660 (2013).
- [3] L. D. Xu, W. He, S. Li, [Internet of things in industries: A survey](#), *IEEE Transactions on Industrial Informatics*, 10:2233–2243 (2014).
- [4] R. S. Peres, X. Jia, J. Lee, K. Sun, A. W. Colombo, J. Barata, [Industrial Artificial Intelligence in Industry 4.0 -Systematic Review, Challenges and Outlook](#), *IEEE Access* (2020).
- [5] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, M. Gidlund, [Industrial Internet of Things: Challenges, Opportunities, and Directions](#), *IEEE Transactions on Industrial Informatics*, 14:4724–4734 (2018).
- [6] R. W. Liu, Y. Guo, Y. Lu, K. T. Chui, B. B. Gupta, [Deep Network-Enabled Haze Visibility Enhancement for Visual IoT-Driven Intelligent Transportation Systems](#), *IEEE Transactions on Industrial Informatics*, 19:1581–1591 (2023).
- [7] N. Ahmed, D. De, I. Hussain, [Internet of Things \(IoT\) for Smart Precision Agriculture and Farming in Rural Areas](#), *IEEE Internet of Things Journal*, 5:4890–4899 (2018).
- [8] C. Kolias, G. Kambourakis, A. Stavrou, J. Voas, [DDoS in the IoT: Mirai and other botnets](#), *Computer*, 50:80–84 (2017).
- [9] N. Mishra, S. Pandya, [Internet of Things Applications, Security Challenges, Attacks, Intrusion Detection, and Future Visions: A Systematic Review](#), *IEEE Access*, 9:59353–59377 (2021).
- [10] H. Wu, H. Han, X. Wang, S. Sun, [Research on Artificial Intelligence Enhancing Internet of Things Security: A Survey](#), *IEEE Access*, 8:153826–153848 (2020).
- [11] X. Yuan, C. Li, X. Li, [DeepDefense: Identifying DDoS Attack via Deep Learning](#), *2017 IEEE International Conference on Smart Computing, SMARTCOMP 2017* (2017).
- [12] M. A. Belay, A. Rasheed, P. S. Rossi, [Sparse Non-Linear Vector Autoregressive Networks for Multivariate Time Series Anomaly Detection](#), *IEEE Signal Processing Letters*, pp. 1–5 (2024).
- [13] M. A. Belay, A. Rasheed, P. S. Rossi, [Multivariate Time Series Anomaly Detection via Low-Rank and Sparse Decomposition](#), *IEEE Sensors Journal*, 24:34942–34952 (2024).
- [14] M. A. Belay, A. Rasheed, P. Salvo Rossi, [Self-Supervised Modular Architecture for Multi-Sensor Anomaly Detection and Localization](#), *IEEE Conference on Artificial Intelligence (CAI)*, pp. 1278–1283 (2024).
- [15] M. A. Belay, A. Rasheed, P. Salvo Rossi, [MTAD: Multiobjective Transformer Network for Unsupervised Multisensor Anomaly Detection](#), *IEEE Sensors Journal*, 24:20254–20265 (2024).
- [16] V. Chandola, A. Banerjee, V. Kumar, [Anomaly detection: A survey](#), *ACM Computing Surveys*, 41 (2009).
- [17] M. A. Belay, S. S. Blakseth, A. Rasheed, P. Salvo Rossi, [Unsupervised Anomaly Detection for IoT-Based Multivariate Time Series: Existing Solutions, Performance Analysis and Future Directions](#), *Sensors*, 23:2844 (2023).
- [18] A. Rasheed, O. San, T. Kvamsdal, [Digital twin: Values, challenges and enablers from a modeling perspective](#), *IEEE Access*, 8:21980–22012 (2020).
- [19] F. Tao, H. Zhang, A. Liu, A. Y. C. Nee, [Digital Twin in Industry: State-of-the-Art](#), *IEEE Transactions on Industrial Informatics*, 15:2405–2415 (2019).
- [20] S. Mihai, M. Yaqoob, D. V. Hung, W. Davis, P. Towakel, M. Raza, M. Karamanoglu, B. Barn, D. Shetve, R. V. Prasad, H. Venkataraman, R. Trestian, H. X. Nguyen, [Digital Twins: A Survey on Enabling Technologies, Challenges, Trends and Future Prospects](#), *IEEE Communications Surveys and Tutorials*, 24:2255–2291 (2022).
- [21] A. Castellani, S. Schmitt, S. Squartini, [Real-World Anomaly Detection by Using Digital Twin Systems and Weakly Supervised Learning](#), *IEEE Transactions on Industrial Informatics*, 17:4733–4742 (2021).
- [22] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, H. Janicke, [Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning](#), *IEEE Access*, 10:40281–40306 (2022).
- [23] J. Li, L. Lyu, X. Liu, X. Zhang, X. Lyu, [FLEAM: A Federated Learning Empowered Architecture to Mitigate DDoS in Industrial IoT](#), *IEEE Transactions on Industrial Informatics*, 18:4059–4068 (2022).
- [24] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. y. Arcas, [Communication-Efficient Learning of Deep Networks from Decentralized Data](#), *International Conference on Artificial Intelligence and Statistics* (2017).
- [25] T. Li, A. K. Sahu, A. Talwalkar, V. Smith, [Federated Learning: Challenges, Methods, and Future Directions](#), *IEEE Signal Processing Magazine*, 37:50–60 (2020).
- [26] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y. C. Liang, Q. Yang, D. Niyato, C. Miao, [Federated Learning in Mobile Edge Networks: A Comprehensive Survey](#), *IEEE Communications Surveys and Tutorials*, 22:2031–2063 (2020).
- [27] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, H. Vincent Poor, [Federated Learning for Internet of Things: A Comprehensive Survey](#), *IEEE Communications Surveys and Tutorials*, 23:1622–1658 (2021).
- [28] Z. Liu, C. Guo, D. Liu, X. Yin, [An Asynchronous Federated Learning Arbitration Model for Low-Rate DDoS Attack Detection](#), *IEEE Access*, 11:18448–18460 (2023).
- [29] S. S. Mahadik, P. M. Pawar, R. Muthalagu, [Edge-Federated Learning-Based Intelligent Intrusion Detection System for Heterogeneous Internet of Things](#), *IEEE Access*, 12:81736–81757 (2024).
- [30] Y. Alhasawi, S. Alghamdi, [Federated Learning for Decentralized DDoS Attack Detection in IoT Networks](#), *IEEE Access*, 12:42357–42368 (2024).
- [31] D. Su, Z. Qu, [Detection DDoS of Attacks Based on Federated Learning with Digital Twin Network](#), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13370 LNAI:153–164 (2022).
- [32] Y. Yigit, B. Bal, A. Karameşoğlu, T. Q. Duong, B. Canberk, [Digital Twin-Enabled Intelligent DDoS Detection Mechanism for Autonomous Core Networks](#), *IEEE Communications Standards Magazine*, 6:38–44 (2022).
- [33] A. De Benedictis, F. Flammini, N. Mazzocca, A. Somma, F. Vitale, [Digital Twins for Anomaly Detection in the Industrial Internet of Things: Conceptual Architecture and Proof-of-Concept](#), *IEEE Transactions on Industrial Informatics*, 19:11553–11563 (2023).
- [34] D. Gupta, S. S. Moni, A. S. Tosun, [Integration of Digital Twin and Federated Learning for Securing Vehicular Internet of Things](#), *2023 Research in Adaptive and Convergent Systems RACS 2023* (2023).
- [35] D. Gupta, O. Kayode, S. Bhatt, M. Gupta, A. S. Tosun, [Hierarchical Federated Learning based Anomaly Detection using Digital Twins for Smart Healthcare](#), *IEEE International Conference on Collaboration and Internet Computing*, pp. 16–25 (2021).
- [36] L. Shi, [A Industry 4.0 production line system and its Digital Twin under cyber attack](#), *IEEE DataPort* (2023).