Appendix 1: Detailed Insights and Limitations of Recent IDS Studies in IoT

Article	Insights	Limitations
Yaras & Dener[39]	• Employed a hybrid deep learning algorithm combining CNN and LSTM, designed for efficient intrusion detection in IoT networks. • Achieved high accuracy levels on the CICIoT2023 and TON_IoT datasets, demonstrating the model's effectiveness in real-time attack detection. • Utilized a centralized training paradigm and deployed the model in a cloud environment, showcasing high performance and scalability.	• Significant computational demands may hinder deployment on resource-constrained IoT devices. • Lack of focus on real-time detection capabilities raises concerns about its applicability in dynamic IoT environments requiring immediate response.
Sun et al.[40]	• Introduced an optimized IDS using PSO for feature selection and AdaBoost for classification, tailored for the Internet of Medical Things (IoMT). • Demonstrated strong intrusion detection capabilities with high accuracy, precision, and recall on the NSL-KDD dataset. • Integrated machine learning IDS into smart health applications, enhancing patient care and security infrastructure.	• Use of the NSL-KDD dataset, which is not IoT-specific, may not effectively represent the unique challenges and attack vectors of IoMT environments. • Potential limitations in adapting the methodology to real-world IoMT applications due to dataset constraints.

Continued on next page

Article	Insights	Limitations
Morshedi et al.[41]	• Utilized a CNN-LSTM hybrid model for enhancing IDS capabilities in IoT networks, capturing both spatial and temporal data dependencies effectively. • Achieved remarkable accuracy on the CICIDS2017 dataset, indicating high model stability and robustness against various cyber attacks. • Demonstrated resilience to Gaussian noise, highlighting the model's reliability under adverse conditions.	• Reliance on the CICIDS2017 dataset, which may not fully encompass evolving IoT attack patterns, potentially reducing effectiveness in diverse IoT scenarios. • High computational requirements could limit deployment in IoT environments with constrained resources.
Mahadik et al.[53]	• Proposed a federated learning-based CNN classifier for detecting DDoS attacks within heterogeneous IoT applications, demonstrating effective adaptation to dynamic network conditions. • Achieved high accuracy rates across various conditions and configurations on the CICDDoS2019 dataset, showcasing the model's adaptability and performance. • Enhanced data privacy and security through federated learning by training locally on edge devices, reducing the risk of data breaches.	• Variability in performance under different data distributions and non-IID conditions, indicating challenges in model consistency across diverse IoT environments. • Dependence on the CICDDoS2019 dataset, which lacks IoT specificity, potentially affecting the model's generalizability to real-world IoT security applications.

Continued on next page

Article	Insights	Limitations
Latif et al.[42]	• Developed a CNN-GAN based IDS for IIoT with a focus on leveraging deep transfer learning and genetic algorithms for enhanced detection accuracy. • Utilized a multi-layer architectural approach to integrate various optimization techniques, significantly improving detection rates on the Edge-IIoTset dataset. • Demonstrated robustness and adaptability in IIoT environments, effectively addressing diverse cyber threats with high accuracy.	• Performance challenges with non-IID data which may affect the model's effectiveness in real-world applications with diverse and heterogeneous data sources. • Potential scalability issues in dynamic IoT environments due to the complexity of maintaining consistent model performance across multiple clients.
Kilichev et al.[43]	• Presented an advanced IDS for IoT-based EVCS by integrating CNN, LSTM, and GRU models. • Achieved high performance across binary, six-class, and fifteen-class classifications on the Edge-IIoTset dataset. • Efficiently captured both spatial and temporal features, enhancing detection and classification of various cyber threats in real-time.	• Significant computational complexity may hinder deployment on resource-constrained IoT devices. • Centralized training approach may not be suitable for scenarios where data privacy and decentralization are critical.

Continued on next page

Article	Insights	Limitations
Karthikeyan et al.[44]	• Introduced a novel Firefly Algorithm with Machine Learning (FA-ML) technique for enhancing intrusion detection accuracy in WSN-IoT environments. • Achieved superior performance with maximum accuracy of 99.34% on the NSL-KDD dataset, demonstrating its effectiveness as a security solution. • Leveraged the optimization capabilities of the Firefly Algorithm for feature selection and Grey Wolf Optimizer for parameter tuning of the SVM classifier.	• Use of the NSL-KDD dataset, which is not IoT-specific, may reduce the model's effectiveness in real-world IoT scenarios. • Centralized training approach may pose privacy concerns and is not ideal for decentralized IoT environments.
Javeed et al.[58]	• Introduced a federated learning-based zero trust IDS for IoT, combining CNN for spatial feature extraction and BiLSTM for temporal feature extraction. • Achieved high performance in detecting and classifying network intrusions on the Edge-IIoT and CICIDS2017 datasets. • Enhanced privacy and scalability through federated learning, outperforming traditional centralized methods.	• Performance variations with increased numbers of edge devices indicate a need for further optimization in large-scale IoT deployments. • Reliance on the CICIDS2017 dataset, which is not IoT-specific, may limit the model's ability to capture unique characteristics of IoT traffic.
Hazman et al.[45]	• Introduced IDS-SIoDL, a deep learning-based IDS for IoT-based smart cities using LSTM and feature engineering techniques. • Achieved high accuracy, precision, and recall with efficient training and classification times. • Utilized tensor processing units (TPUs) to optimize performance, making it suitable for large-scale IoT data.	• Reliance on the NSL-KDD dataset, which is not IoT-specific, may reduce effectiveness in real-world IoT scenarios. • High computational complexity may limit deployment on resource-constrained IoT devices.

Continued on next page

Article	Insights	Limitations
Elnakib et al.[91]	• Proposed a deep learning model for IoT intrusion detection, evaluating centralized and federated learning approaches. • Achieved high accuracy with centralized learning but compromised data privacy. • Federated learning approach balanced accuracy and privacy, making it viable for IoT intrusion detection.	• Significant performance drop with non-IID data distributions in federated learning. • Centralized aggregation of model updates may still pose privacy risks and may not be suitable for complete decentralization.
Bhavsar et al.[54]	• Introduced a federated learning-based IDS (FL-IDS) for vehicular networks in IoT environments using logistic regression and CNN classifiers. • Achieved high accuracy rates on the NSL-KDD and Car-Hacking datasets, significantly enhancing transportation IoT security. • Leveraged federated learning to protect data privacy by sharing only model updates instead of raw data.	• Reliance on the NSL-KDD dataset, which is not IoT-specific, may limit effectiveness in real-world IoT scenarios. • High computational complexity may hinder deployment on resource-constrained IoT devices in transportation IoT environments.
Ayoob et al.[46]	• Presented a heterogeneous machine learning-based stack classifier model for IoT intrusion detection. • Integrated the K-Best algorithm for feature selection and an ensemble model combining traditional ML models. • Achieved exceptional accuracy, precision, recall, and F1-score on the ToN-IoT dataset.	• Reliance on the ToN-IoT dataset, which may not encompass the full diversity of IoT attack scenarios. • Computational complexity may pose challenges for deployment on resource-constrained IoT devices.

Continued on next page

Article	Insights	Limitations
Alazab et al.[92]	• Developed an LSTM-based IDS with a DAC algorithm to detect and defend against intrusions at IoT nodes. • Achieved high validation accuracy and precision, recall, and F1 scores on the CIC-IDS2017 dataset. • Provided effective defense mechanisms with an average response time under 1.2 seconds.	• Reliance on the CIC-IDS2017 dataset, which is not IoT-specific, may limit effectiveness in real-world IoT scenarios. • High computational complexity may hinder deployment on resource-constrained IoT devices.
Zakariyya et al.[55]	• Introduced a Resource-Efficient Federated Deep Neural Network (REFDNN) for IoT security monitoring. • Combined regularization and simulated micro-batching to train DNNs in a resource-friendly manner. • Achieved significant reductions in memory and time usage while maintaining or improving accuracy.	• Reliance on specific datasets (N-BaIoT, Kitsune, WUSTL) may limit effectiveness in general IoT scenarios. • Computational complexity, despite optimizations, may still pose challenges for deployment on extremely resource-constrained IoT devices.
Wang et al.[93]	• Presented S2CGAN-IDS, integrating Siamese networks and autoencoders with CGAN for data augmentation. • Achieved significant improvement in F1-score on the CICIDS2017 dataset, enhancing detection of minority categories. • Demonstrated robustness in maintaining high detection precision for both majority and minority classes.	• Reliance on the CICIDS2017 dataset, which may not capture the full diversity of IoT attack scenarios. • Computational complexity may pose challenges for deployment on resource-constrained IoT devices.

Continued on next page

Article	Insights	Limitations
Okey et al.[94]	• Introduced an efficient-lightweight ensemble transfer learning (ELETL-IDS) model based on CNN for detecting a wide range of IoT attacks. • Achieved high accuracy, precision, recall, and F1-score on the CIC-IDS2017 and CSE-CICIDS2018 datasets. • Demonstrated the model's efficiency and reliability for deployment in cloud IoT systems.	• Reliance on the CIC-IDS2017 and CSE-CICIDS2018 datasets, which may not cover the full diversity of IoT attack scenarios. • Transformation of numeric tabular data into image format for CNN-based analysis introduces significant computational complexity.
Jullian et al.[95]	• Explored distributed deep learning-based intrusion detection in IoT environments through a federated learning framework. • Achieved high accuracy across different setups, significantly improving detection of diverse cyber-attacks. • Ensured data privacy by keeping data localized on edge devices and only sharing model updates.	• Reliance on the NSL-KDD and BoT-IoT datasets, which may limit effectiveness in real-world IoT scenarios. • High computational complexity for preprocessing and training deep learning models may hinder deployment on resource-constrained IoT devices.
Jithish et al.[56]	• Introduced a federated learning-based approach for anomaly detection in smart grids, ensuring data privacy and reducing network overhead. • Achieved high accuracy with the federated 1D-CNN model, demonstrating efficiency in terms of memory, CPU usage, bandwidth, and power consumption. • Demonstrated suitability for real-world deployment in smart grids with robust performance across multiple datasets.	• Reliance on non-IoT-specific datasets (KDD99, NSL-KDD, CIDDS-001) may reduce effectiveness in real-world IoT scenarios. • High computational complexity of implementing federated learning on resource-constrained devices poses challenges.

Continued on next page

Article	Insights	Limitations
Hussen et al.[96]	• Introduced a Fully Streaming Big Data Framework for Cyber Security (FSBDL) using an optimized CNN model for real-time intrusion detection. • Leveraged parallel optimization algorithms (Adam and RMSprop) to enhance efficiency and stability. • Achieved high accuracy and reliability in detecting intrusions promptly, making it a promising solution for enhancing cybersecurity.	• Reliance on the NSL-KDD dataset, which is not IoT-specific, may reduce effectiveness in real-world IoT scenarios. • High computational complexity of implementing parallel optimization algorithms may pose challenges for deployment on resource-constrained IoT devices.
Belarbi et al.[97]	• Designed an FL-based IDS tailored for IoT systems using DBNs and DNNs, addressing non-IID data issues by partitioning data according to IP addresses. • Achieved significant detection improvements with pre-trained models, demonstrating enhanced detection stability and accuracy in heterogeneous IoT networks. • Evaluated performance using the TON-IoT dataset, achieving a performance boost of up to 30% with FedProx and FedYogi aggregation methods.	• Reliance on the TON-IoT dataset, which may not encompass the full diversity of potential attack scenarios. • High computational complexity due to the use of various aggregation methods and the need for pre-training to mitigate data heterogeneity effects.

Continued on next page

Article	Insights	Limitations
Almuqren et al.[98]	• Presented the HMMLB-BND method, integrating MFFO for feature selection and a CNN-QRNN model for efficient botnet detection in cloud-aided IoT environments. • Achieved high performance metrics, demonstrating effectiveness in detecting and classifying botnet attacks. • Improved predictive performance and robustness against botnet attacks through a hybrid approach using CNN and QRNN along with CBOA for hyperparameter tuning.	• Reliance on the N-BaIoT dataset, which focuses on specific botnet attacks and may not encompass the full diversity of potential attack scenarios. • High computational complexity may pose challenges for deployment on resource-constrained IoT devices.
Alghamdi & Bellaiche[77]	• Introduced a two-stage hybrid IDS combining ANN, CNN, and LSTM, trained on the IoT-23 dataset for detecting and classifying network intrusions in IoT environments. • Achieved high accuracy of 99.93%, significantly outperforming individual models. • Enhanced system throughput and efficiency using Lambda architecture for real-time traffic monitoring.	• High computational complexity may limit deployment on resource-constrained IoT devices. • Reliance on the IoT-23 dataset may limit generalizability to other IoT scenarios.

Continued on next page

Article	Insights	Limitations
Aldaej et al.[47]	• Proposed a deep learning-inspired IoT-IDS mechanism optimized for edge computing environments, using a hybrid RNN and Bi-LSTM model. • Leveraged attribute selection at the edge to reduce dataset size significantly without compromising detection accuracy. • Demonstrated high recall, F1-measure, accuracy, and precision on the BoT-IoT dataset.	• Reliance on the BoT-IoT dataset, which may not cover the full diversity of potential attack scenarios in IoT environments. • High computational complexity of implementing RNN and Bi-LSTM models may limit deployment on resource-constrained IoT devices.
Rey et al.[99]	• Presented a federated learning approach enhancing privacy and security by keeping data localized and sharing model updates. • Demonstrated robustness in detecting malware across different IoT devices using supervised and unsupervised learning techniques. • Improved applicability in real-world scenarios by training and evaluating models on IoT-specific datasets.	• Reliance on a limited number of devices in the N-BaIoT dataset may affect generalizability. • Federated learning introduces complexities in model synchronization and aggregation, increasing computational overhead and communication latency.
Ponniah & Retnaswamy[100]	• Proposed MWKF-LSTM model demonstrating superior performance in identifying various types of intrusions with high accuracy and low false-positive rates. • Enhanced IoT-Cloud security using blockchain-based authentication and elliptic curve cryptography. • Robust and efficient intrusion detection making it valuable for enhancing security in IoT-Cloud scenarios.	• Reliance on the NSL-KDD dataset, which is not IoT-specific, may limit applicability to real-world IoT environments. • Centralized training and deployment may raise scalability and resilience concerns.

Continued on next page

Article	Insights	Limitations
Nguyen et al.[78]	• Developed Realguard, a lightweight IDS demonstrating high accuracy in real-time intrusion detection for IoT environments. • Achieved an average detection accuracy of 99.57% with minimal false positives. • Efficient feature extraction and attack detection model suitable for resource-constrained IoT devices.	• Reliance on the CICIDS2017 dataset, which is not IoT-specific, may limit relevance to real-world IoT environments. • Potential susceptibility to adversarial attacks due to lack of barrier layers mitigating adversarial sample impact.
Mothukuri et al.[101]	• Presented a federated learning-based approach addressing privacy concerns by keeping data localized and sharing model updates. • Demonstrated robustness in detecting anomalies across various IoT devices using GRUs for efficient training. • Achieved high accuracy rates and reduced false alarms, outperforming non-federated approaches.	• Reliance on the Modbus-based network dataset may limit the model's ability to generalize to other IoT environments. • Federated learning introduces complexities in model synchronization and aggregation, increasing computational overhead and communication latency.
Liang et al.[102]	• Proposed OICS-VFSL model leveraging few-shot learning to efficiently detect novel attack types with limited training data. • Optimized intra-class and inter-class distances through variational feature representation, enhancing classification accuracy. • Demonstrated robust performance in handling imbalanced datasets and out-of-distribution issues in IoT environments.	• Reliance on the NSL-KDD dataset, which is not IoT-specific, may reduce effectiveness in real-world IoT scenarios. • Centralized nature of model training may be a bottleneck in distributed and dynamic IoT environments.

Continued on next page

Article	Insights	Limitations
Labiod et al.[103]	• Proposed a two-level IDS with a lightweight hybrid VAE-MLP model for fog nodes and centralized cloud IDS for further classification. • Achieved efficient binary classification of traffic in IoT networks using a 6-layer MLP. • Utilized VAE to encode data into low-dimensional representation, enhancing classification accuracy.	• High computational complexity may pose challenges for deployment on resource-constrained IoT devices. • Potential limitations in real-world scenarios where efficiency and low latency are critical.
Kaushik et al.[74]	• Introduced MI2G algorithm for efficient feature selection, significantly improving IDS performance. • Achieved high accuracy, precision, recall, and F1-score across different classifiers, with DT and RF showing excellent performance. • Lightweight nature of IDS makes it suitable for resource-constrained IoT environments.	• Reliance on the CICIDS2018 and UNSW-NB15 datasets may limit generalizability due to dataset limitations. • Centralized training approach may not be ideal for distributed IoT environments.
Li et al.[104]	• Proposed FLEAM using GRU models with iterative model averaging for detecting and mitigating DDoS attacks in IIoT environments. • Demonstrated high detection accuracy and efficient mitigation with reduced response times. • Ensured privacy and security by keeping data localized and only sharing model parameters.	• Reliance on the UNSW-NB15 dataset, which may limit applicability to real-world IIoT scenarios. • Federated learning introduces complexities in synchronization and aggregation, increasing overhead and latency.

Continued on next page

Article	Insights	Limitations
Gupta et al.[48]	• Developed MUSE system using deep hierarchical stacked ANN for detecting unknown attacks in multi-cloud healthcare systems. • Achieved high training and test accuracies, demonstrating robustness and efficiency. • Leveraged layers of trained edge cloud models for faster training times.	• Hierarchical training approach requires careful synchronization and management of models across different cloud layers. • Potential challenges in highly dynamic and distributed environments.
Guezzaz et al.[75]	• Proposed a hybrid IDS combining PCA and K-NN for improved detection accuracy and false alarm rates. • Achieved high accuracy and detection rates with low false alarm rates on the NSL-KDD and Bot-IoT datasets. • Effective in securing IIoT environments by reducing dimensionality and enhancing classifier efficiency.	• Reliance on the NSL-KDD and Bot-IoT datasets may limit generalizability to broader IoT environments. • Centralized training approach may not be ideal for highly distributed IoT environments.
Gad et al.[49]	• Proposed a distributed IDS using various ML algorithms for detecting attacks within IoT environments. • Leveraged the ToN-IoT dataset for accurate classification of a wide range of cyber threats. • Enhanced performance using Chi2 technique for feature selection and SMOTE method for addressing class imbalance.	• Complexity of managing distributed training and deployment architecture across cloud, fog, and edge layers. • Reliance on the ToN-IoT dataset may limit the system's efficacy if dataset biases or missing attack types are present.

Continued on next page

Article	Insights	Limitations
Dat-Thinh et al.[105]	• Proposed MidSiot IDS for detecting and classifying attacks in IoT networks, achieving high accuracy across various datasets. • Utilized a multi-stage approach for analyzing different aspects of network traffic, leading to high detection rates. • Demonstrated robustness across different IoT scenarios using multiple datasets for evaluation.	• Network bandwidth consumption and delays when deploying the model from the cloud to local gateways may impact real-time detection performance.
Alferaidi et al.[50]	• Proposed a CNN-LSTM model for IoV intrusion detection, leveraging CNN for spatial features and LSTM for temporal dependencies. • Achieved high detection accuracy on the NSL-KDD and UNSW-NB15 datasets. • Utilized Apache Spark framework for efficient large-scale data processing, suitable for real-time IoV environments.	• Reliance on the NSL-KDD and UNSW-NB15 datasets may limit effectiveness in real-world IoV scenarios. • Operational complexity and resource intensity of managing distributed training process across multiple nodes.
Alasmary et al.[51]	• Introduced ShieldRNN for detecting DDoS attacks using lightweight classifiers at IoT nodes and more powerful classifiers at servers. • Leveraged sequence majority voting to enhance model robustness and accuracy. • Demonstrated superior performance on CIC-IDS2017 and CIC-IoT2022 datasets, ensuring scalability and timely detection.	• Reliance on the CIC-IDS2017 dataset may limit effectiveness in exclusively IoT scenarios. • Inclusion of the IoT-specific CIC-IoT2022 dataset partially mitigates this issue.

Continued on next page

Article	Insights	Limitations
Abosata et al.[106]	• Proposed FT-CID model leveraging federated and transfer learning to address privacy concerns and heterogeneous nature of IIoT environments. • Achieved high detection accuracy for various intrusion types using CNN for local learning. • Enhanced data privacy by keeping data localized and only sharing model parameters.	• Reliance on a simulated dataset focused on a specific IIoT application area may limit generalizability to other IoT environments. • Potential limitations in capturing nuanced attack patterns found in real IIoT deployments.
Tabassum et al.[52]	• Proposed a distributed IDS using SAE and CNNs for efficient and accurate intrusion detection in IoT systems. • Achieved high detection rates with minimal latency and efficient resource utilization. • Leveraged distributed computing to enhance efficiency and accuracy in real-time IoT applications.	• Reliance on non-IoT-specific datasets (KDD 99, NSL-KDD) may reduce effectiveness in real-world IoT scenarios. • Limited evaluation of attack types specific to different IoT applications.
Kumar et al.[34]	• Proposed an ensemble IDS combining kNN, XGB, NB, and RF for detecting attacks in IoT environments using fog computing. • Achieved high detection rates for various attack types on modern datasets (UNSW-NB15, DS2OS). • Distributed approach reduced latency and improved real-time detection capabilities in dynamic IoT environments.	• Reliance on the UNSW-NB15 dataset, which may not cover the full spectrum of IoT attacks. • Complexities in synchronization and data aggregation in distributed environments.

Continued on next page

Article	Insights	Limitations
Khater et al.[107]	• Proposed a lightweight IDS using MVSR N-gram techniques and MLP for high accuracy and efficiency in IoT environments. • Achieved significant reductions in space complexity and computational overhead, making it suitable for resource-constrained IoT devices. • Evaluation on a Raspberry Pi demonstrated high accuracy, low false positive rates, and minimal energy consumption.	• Reliance on the ADFA-LD dataset, which is not specifically tailored for IoT environments, may limit generalizability. • Centralized training approach may introduce challenges in scalability and real-time adaptation.
Houng et al.[108]	• Proposed LocKedge using PCA and ANN models for high accuracy and low complexity in edge computing environments. • Enhanced data privacy through federated learning by keeping data localized on edge nodes. • Demonstrated robustness and efficiency in adapting to different IoT scenarios and attack types using the BoT-IoT dataset.	• Federated learning introduces challenges in model synchronization and communication latency. • Further research needed to improve detection rates for less frequent attack types and optimize resource usage.
Gavel et al.[35]	• Proposed a dual-axis dimensionality reduction technique combining Kalman Filter and Salp Swarm Algorithm (SSA) for efficient feature selection. • Achieved high detection accuracy and low computational overhead using KELM for classification. • Demonstrated scalability and real-time responsiveness in dynamic IoT environments.	• Reliance on non-IoT-specific datasets (NSL-KDD, KYOTO 2006+) may limit effectiveness in real-world IoT scenarios. • Complexities in data synchronization and model aggregation in distributed environments.

Continued on next page

Article	Insights	Limitations
Chathoth et al.[57]	• Proposed a federated learning approach combined with differential privacy for enhancing security and privacy in IoT devices. • Demonstrated superior performance with continual learning-based DP methods, effectively handling heterogeneous privacy requirements and non-identical data distributions. • Ensured robustness to hyperparameter changes and mitigated catastrophic forgetting.	• Reliance on the CSE-CIC-IDS2018 dataset, which is not specifically tailored for IoT environments, may limit effectiveness. • Federated learning and DP algorithms reduce performance on rare classes with few training samples.
Li et al.[38]	• Proposed a federated deep learning approach for enhancing detection of cyber threats in industrial CPSs. • Combined CNN and GRU models for effective handling of various types of attacks. • Ensured data privacy by preserving data on local devices and only sharing model parameters.	• Reliance on a single industrial CPS dataset may limit effectiveness in diverse industrial environments. • Federated learning framework introduces computational and communication overhead.
Rahman et al.[33]	• Proposed IDS architectures enhancing scalability and performance for IoT-enabled smart cities using semi-distributed and distributed approaches. • Achieved high detection accuracy with significant improvements in CPU time and efficient handling of large-scale IoT networks. • Leveraged collaborative analytics and parallel processing to address latency and processing time issues.	• Reliance on the AWID dataset, which may not fully capture the diversity of IoT attack scenarios. • Complexities in synchronization and aggregation in semi-distributed and distributed approaches.

Continued on next page

Article	Insights	Limitations
D. L. T. Parra et al.[109]	• Proposed a distributed deep learning approach balancing computational load between IoT devices and backend servers. • Achieved real-time detection capabilities with data privacy using CNNs on IoT devices and LSTM networks on backend servers. • Leveraged edge and cloud computing strengths, making it suitable for large-scale IoT deployments.	• Reliance on the CICIDS2017 dataset, which may not cover all possible IoT attack scenarios. • Complexities in synchronization and data aggregation in distributed training process.
Pacheco et al.[110]	• Demonstrated the effectiveness of ANNs within an ABA-IDS framework for characterizing normal behaviors and detecting anomalies in IoT fog nodes. • Achieved high detection rates with low false positive rates and minimal overhead in terms of execution time, memory, and CPU usage. • Suitable for real-time applications in fog computing environments.	• Lack of detailed information about datasets used for evaluation limits the generalizability of the proposed IDS. • Centralized training process may introduce bottlenecks and reduce robustness in large-scale deployments.
Khoa et al.[111]	• Proposed a collaborative learning model enhancing intrusion detection accuracy while preserving data privacy and reducing network traffic. • Achieved significant improvements in detection accuracy, learning speed, and reduction in network overhead compared to conventional methods. • Demonstrated effective training on local data while collaborating on a global model, making it suitable for IoT Industry 4.0 applications.	• Reliance on specific datasets (KDD, NSL-KDD, UNSW-NB15, N-BaIoT) may limit generalizability to real-world IoT environments. • Challenges in synchronizing and aggregating models across multiple IoT gateways, increasing computational overhead and communication latency.

Continued on next page

Article	Insights	Limitations
Abdel-Basset et al.[112]	• Proposed Deep-IFS model combining LocalGRU and MHA layers to capture local and long-term dependencies in IIoT traffic data. • Achieved high accuracy in identifying intrusions and cyber-attacks. • Enhanced scalability and response times by processing data closer to the source in a fog computing environment.	• Does not address data privacy concerns critical in industrial applications. • Messaging complexity in broadcasting data across the IIoT environment may hinder intrusion detection and alarm aggregation efficiency.
Patil et al.[113]	• Proposed HLDNS framework for enhancing detection of intrusions in dynamic and distributed cloud computing environments using BBA for feature selection and RF for classification. • Achieved robust anomaly detection capabilities with efficient handling of high network traffic. • Distributed nature allows for efficient handling of high network traffic and correlation of alerts from multiple servers.	• Dependence on the hypervisor for monitoring and detection may introduce latency and computational overhead. • Quality and comprehensiveness of training datasets (UNSW-NB15, CICIDS-2017) may impact generalizability.
Nguyen et al.[114]	• Proposed DIOT leveraging federated learning for detecting anomalies in IoT device communication patterns. • Ensured privacy by keeping data localized while benefiting from collective learning. • Enhanced accuracy of anomaly detection with device-type-specific models, significantly reducing false alarms and improving detection rates.	• Reliance on specific datasets (KDD99, NSL-KDD, UNSW-NB15) may limit generalizability to diverse IoT environments. • Challenges in synchronization and aggregation of models across multiple IoT gateways.

Continued on next page

Article	Insights	Limitations
Gajewski et al.[115]	• Proposed a distributed IDS architecture addressing the security challenges of Smart Home environments by combining local and centralized detection capabilities. • Utilized a two-layer network architecture for efficient preliminary analysis at the user's premises and comprehensive data analysis at the ISP's infrastructure. • Ensured robust defense against both local and distributed attacks by leveraging the strengths of host-based and network-based IDS. • Enhanced overall security posture through real-time monitoring and advanced threat detection by collaborating between the ISP and the user.	• Dependency on the ISP for deeper analysis and detection might introduce latency in the detection process. • Heavy reliance on the Home Gateway's processing capabilities for preliminary analysis. • Distributed model requires efficient communication and data transfer between the user's premises and the ISP, which can be challenging regarding bandwidth and security. • Lack of specific IoT-related datasets and ML/DL methods limits the applicability of the IDS to general Smart Home scenarios rather than specific IoT applications.

Continued on next page

Article	Insights	Limitations
Alrashdi et al.[71]	• Proposed the AD-IoT system leveraging distributed fog nodes for enhanced detection of cyberattacks on IoT devices in a smart city environment. • Utilized the Random Forest algorithm for high accuracy in anomaly detection, achieving a classification accuracy of 99.34 • Addressed unique IoT security challenges by distributing the detection process closer to data sources, suitable for limited resources and heterogeneous IoT devices. • Provided real-time alerts to administrators, ensuring prompt responses to potential threats and enhancing the overall security posture of smart cities.	• The AD-IoT system significantly enhances the detection of cyberattacks on IoT devices within a smart city environment by leveraging the computational power of distributed fog nodes. • The use of the Random Forest algorithm allows for high accuracy in anomaly detection, achieving a classification accuracy of 99.34 • This approach addresses the unique challenges of IoT security, such as the limited resources and heterogeneous nature of IoT devices, by distributing the detection process closer to the data sources. • Additionally, the AD-IoT system's capability to provide real-time alerts to administrators ensures prompt responses to potential threats, thereby enhancing the overall security posture of smart cities.

Appendix 2: Detailed Summary of Acronyms Used in the Study

Acronym	Definition
AE	Auto Encoder
AI	Artificial Intelligence
AIDS	Anomaly-based Intrusion Detection System
ANN	Artificial Neural Network
AUC	Area Under the ROC Curve
BBA	Binary Bat Algorithm
BiLSTM	Bidirectional Long Short-Term Memory
CGAN	Conditional Generative Adversarial Network
CNN	Convolutional Neural Network
CPS	Cyber-physical Systems
DBN	Deep Belief Network
DDoS	Distributed Denial of Service
DL	Deep Learning
DNN	Deep Neural Network
DoS	Denial of Service
DR	Detection Rate
DT	Decision Tree
FAR	False Alarm Rate
FF	Firefly Algorithm
FPR	False Positive Rate
GAN	Generative Adversarial Network
GBM	Gradient Boosting Machine
GRU	Gated Recurrent Unit
HIDS	Host-based Intrusion Detection System
IDS	Intrusion Detection System
IIoT	Industrial Internet of Things
IoT	Internet of Things
KNN	k-Nearest Neighbours
LDA	Linear Discriminant Analysis
LR	Logistic Regression
LSTM	Long Short-term Memory
MCC	Matthews Correlation Coefficient
ML	Machine Learning
MLP	Multi-Layer Perceptron
NB	Naïve Bayes
NIDS	Network-based Intrusion Detection System
PSO	Particle Swarm Optimization
QDA	Quadratic Discriminant Analysis
REFDNN	Resource Efficient Federated Deep Neural Network
RF	Random Forest
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic Curve
SLR	Systematic Literature Review
SVM	Support Vector Machines

Continued on next page

Acronym	Definition
TPR	True Positive Rate
WSN	Wireless Sensor Network
XGB	Extreme Gradient Boosting