MULTIDISCIPLINARY SCIENTIFIC RESEARCH
BJMSR VOL 11 NO 3 (2026) P-ISSN 2687-850X E-ISSN 2687-8518
Available online at https://www.cribfb.com
Journal homepage: https://www.cribfb.com/journal/index.php/BJMSR
Published by CRIBFB, USA

AN AI BENCHMARK SELECTION FRAMEWORK FOR SUSTAINABLE CYBERSECURITY: COMPARATIVE CHARACTERIZATION OF CIC-IDS2017, UNSW-NB15, AND IOT-23

Boumedyen Shannaq [(a)1]

[(a)]*Associate Professor, College of Business, Management Information System Department, University of Buraimi, Al Buraimi, Oman; E-mail: boumedyen@uob.edu.om*

ARTICLE INFO

Article History:

Received: 5th January 2026
Reviewed & Revised: 5th January 2026 to 25th June 2026
Accepted: 26th June 2026
Published: 30th June 2026

Keywords:

Artificial Intelligence, Intrusion Detection Systems, Benchmark Selection Framework, Explainable Artificial Intelligence, Sustainable Cybersecurity, Resilient Digital Infrastructure, SDG 9, SDG 16, Oman Vision 2040

JEL Classification Codes:

C6, C8, C9

Peer-Review Model:

External peer review was done through double-blind method.

ABSTRACT

AI is not just a tool for digital transformation, but a vital component in safeguarding the critical digital infrastructure and facilitating sustainable digital transformation. But the choice of inappropriate benchmark datasets can make AI-driven IDS less reliable, transparent, and reproducible, ultimately undermining their role in supporting resilient cybersecurity ecosystems aligned with the SDGs, especially SDG 9 (Industry, Innovation, and Infrastructure) and SDG 16 (Peace, Justice and Strong Institutions). This work proposed four complementary analytical dimensions that were comparatively analysed for three widely used benchmark datasets: dataset characterisation, attack diversity, Mutual Information (MI)-based feature importance, and feature correlation analysis. The findings show significant inter-dataset differences. There are 71,984,818 network records in the IoT-23, much larger than those of CIC-IDS2017 (2,830,743) and UNSW-NB15 (2,540,047), which makes it more suitable for large-scale deep learning research. In addition, CIC-IDS2017 offers the highest feature representation (80 features) and attack diversity (15 attack categories). In comparison, UNSW-NB15 is a good benchmark for feature representation (50 features) and attack diversity (9 attack categories) after class harmonization. The results of the feature importance analysis also reveal significant differences across datasets: statistical flow features are the most important in CIC-IDS2017, communication endpoint features are the most informative in IoT-23, and protocol-related features are the most important in UNSW-NB15. Based on these results, this study proposes a Benchmark Selection Matrix and a Benchmark Selection Framework to help translate the comparative analysis of datasets into an evidence-based decision-making process for specific AI research scenarios.

INTRODUCTION

With the ever-changing nature of cyber threats, there is a growing need for enterprise networks, cloud deployments, and Internet of Things (IoT) environments to have intelligent Intrusion Detection Systems (IDSs) that can identify known and unknown attacks. Cybersecurity is a core component of sustainable development as Governments and organizations move towards digital transformation. Secure digital infrastructure is essential for innovation, economic resilience, trusted digital services and responsible AI usage. This is why the United Nations Sustainable Development Goals (SDGs), including SDG 9 (Industry, Innovation and Infrastructure) and SDG 16 (Peace, Justice and Strong Institutions), highlight the need for resilient digital ecosystems that support secure technological advancement. Similarly, Oman Vision 2040 considers digital transformation, artificial intelligence, cybersecurity, and innovation as strategic priorities for building a competitive knowledge-based economy. Reliable benchmark datasets, in turn, serve as a crucial starting point for developing trustworthy AI-powered intrusion detection systems to safeguard next-generation digital infrastructure.

The security mechanisms that rely on signatures are becoming less effective as more advanced attacks emerge because they cannot detect new attack patterns. As a result, Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) have emerged as the primary tools used to create adaptive IDS solutions that boast high detection accuracy and low false alarm rates (Dhote & Agrawal, 2026; Hussein et al., 2025; Ren, 2026; Thomas et al., 2025; Rai et al., 2025). In recent years, numerous AI-based intrusion detection models have been proposed by Zhang and Tuo (2025)

[1]Corresponding author: *ORCID ID: 0000-0001-5867-3986*

https://doi.org/10.46281/bjmsr.v11i3.2899

To cite this article: Shannaq, B. (2026). AN AI BENCHMARK SELECTION FRAMEWORK FOR SUSTAINABLE CYBERSECURITY: COMPARATIVE CHARACTERIZATION OF CIC-IDS2017, UNSW-NB15, AND IOT-23. *Bangladesh Journal of Multidisciplinary Scientific Research, 11*(3), 45-63. https://doi.org/10.46281/bjmsr.v11i3.2899

And others, such as ensemble, hybrid deep learning architectures, explainable AI (XAI), federated, transfer learning, and transformer-based architectures (Hejazi et al., 2025; Kalyana, 2026; Sharma & Kumar, 2025). Similarly, researchers have also reported more on AI-based intrusion detection models (Hoa, 2025; Kaliyaperumal et al., 2024; Sagaran et al., 2025). However, most of them achieve good classification performance, with accuracies up to 95% on standard datasets such as CIC-IDS2017, UNSW-NB15, and IoT-23 (Hejazi et al., 2025; Hoa, 2025; Nzuva et al., 2024). Although these are promising results, they are difficult to compare across studies because each study uses a different dataset, preprocessing methods, feature engineering methods, evaluation procedures, and attack categories. Because of this, reported improvements might be due to dataset characteristics rather than to the actual methodology (Ren, 2026; Saini et al., 2023; Sravani et al., 2026). A perusal of the recent literature reveals another important constraint. Although considerable research has focused on developing increasingly powerful detection algorithms, little work has addressed the benchmark datasets themselves. Researchers often choose benchmark datasets not for the specific research they want to conduct, but because they are popular. Popular benchmark datasets such as MNIST handwritten digits, CIFAR-10, and CIFAR-100 are unrelated to the research objectives of explainable AI, IoT security, feature selection, cross-dataset generalization, or lightweight edge deployment. As a result, there has been no systematic comparison of the features, attack diversity, attack relevance, statistical properties, and usability of the most popular public IDS datasets (Chinnasamy & Subramanian, 2025; Kodete et al., 2025; Ren, 2026). To fill this need, a comprehensive comparative characterization of the three most popular intrusion detection datasets, CIC-IDS2017, UNSW-NB15, and IoT-23, is proposed in this study. The study does not propose another intrusion detection algorithm; rather, it examines the inherent properties of these benchmark datasets using statistical analysis, attack-distribution analysis, mutual information-based feature importance, and feature-correlation analysis. Based on these results, a practical framework is proposed for selecting the appropriate dataset for different IDS research objectives. This work focuses on benchmark-centric analysis rather than model-centric evaluation, providing a reference for future AI-based cybersecurity research and laying the foundation for building more powerful, explainable, and transferable intrusion detection systems. This study aims to create an extensive benchmark framework to help researchers select, understand, and evaluate intrusion detection benchmark datasets before constructing an artificial intelligence-based detection model. For this purpose, 3 popular benchmark datasets (CIC-IDS2017, UNSW-NB15, and IoT-23) are comparatively studied in various aspects, including statistical characteristics, attack diversity, feature importance, and feature correlation. The proposed study differs from existing studies, which have primarily focused on analysing the properties of the benchmark data sets and their applicability to the different research tasks in intrusion detection.

The main findings of this study are following: It offers a complete statistical analysis of the CIC-IDS2017, UNSW-NB15 and IoT-23 benchmark datasets, and gives analyses of the attack diversity, class distributions and traffic characteristics from these datasets. This study also analyzes feature relevance based on the Mutual Information, feature correlation and feature redundancy analysis and analyzes the advantages and disadvantages of each intrusion detection data for use in AI-based intrusion detection. It proposes a Benchmark Selection Framework for guiding the selection of datasets for various research goals, ranging from machine learning and deep learning, explainable AI and IoT security, feature selection, transfer learning, to cross-dataset generalization. Finally, it provides future research directions to make more powerful, explainable and generalizable IDSs.

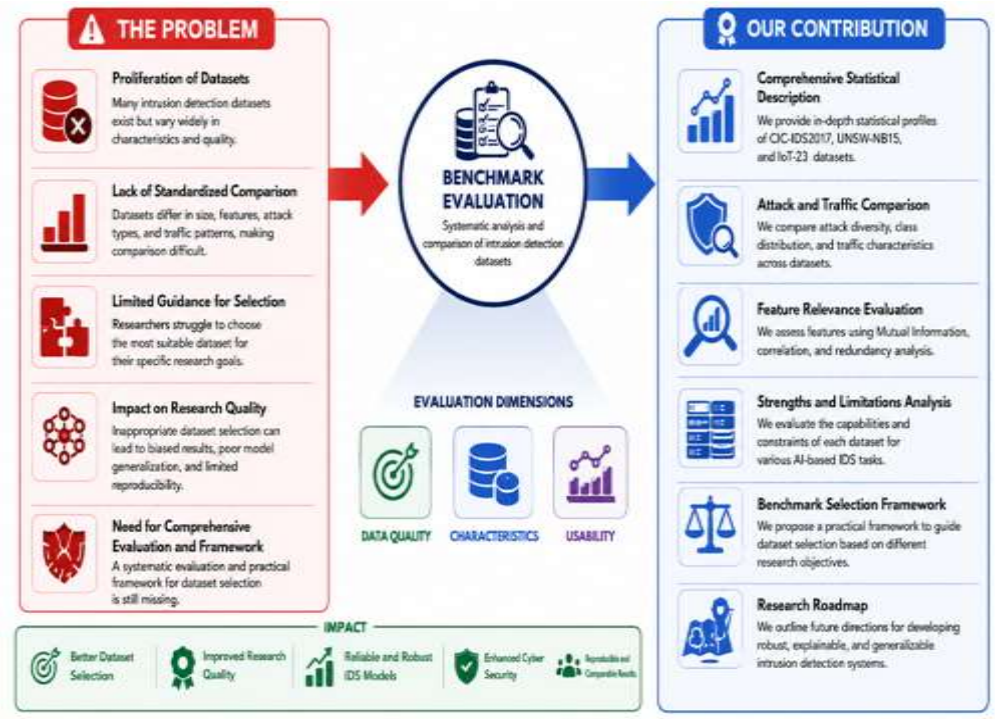

Figure 1. Overview of the proposed benchmark framework for AI-based intrusion detection
Source: Developed by the authors; AI-assisted visualization was used solely for graphical design

This study was motivated by the need to address what is going wrong in the school, as seen in figure 1 below, and to provide a proposed benchmark framework and possible contributions to improve the situation. The left side highlights the critical issues that are plaguing the state of intrusion detection research: unreliable use of datasets, inadequate guidance in using benchmarks, and unfair performance comparisons are some of the issues discussed. The proposed benchmark framework is presented by performing comparative statistical analysis, attack diversity analysis, feature importance evaluation, correlation analysis for three benchmark datasets. The right side emphasizes main contributions, with the ultimate goal being to provide practical guidelines for benchmark selection and a research roadmap for future studies on intrusion detection using artificial intelligence.

The rest of this paper is organized as follows. In Section 2, the related literature and research gaps are analyzed, in Section 3, the proposed benchmark framework and methodology are presented, and in Section 4, the comparative benchmark analysis and findings are reported. Finally, the implications, limitations, future research directions and a conclusion of this study are discussed in Sections 5 and 6.

LITERATURE REVIEW

In today's era, the world's most prevalent intrusion detection system (IDS) technique is an artificial intelligence approach capable of detecting attack patterns that would be impossible to detect with a signature-based approach. Early studies were mostly based on classic machine learning algorithms like Support Vector Machines (SVM), Decision Trees (DT), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), which are appealing due to their computational efficiency and interpretability (Hussein et al., 2025; Sharma & Kumar, 2025; Thomas et al., 2025). More recently, deep learning architectures have demonstrated greater prowess at extracting high-level representations from network traffic, enabling better detection of more advanced and previously unseen attacks.

Recent studies have shown increased interest in hybrid learning models that integrate the advantages of various learning approaches. CNNs are used in a great variety of applications for spatial feature extraction, while LSTM and GRU networks are used to model temporal dependence in network traffic. CNNs, LSTMs, attention mechanisms, and transformer-based methods have recently been combined into hybrid architectures that have consistently achieved high detection rates on benchmark datasets such as CIC-IDS2017 and UNSW-NB15 and enhanced detection performance against various types of attacks (Hejazi et al., 2025; Saini et al., 2023). Ensemble learning methods have attracted considerable interest for combining multiple classifiers to improve predictive stability and reduce false alarms (Sagaran et al., 2025; Sharma & Kumar, 2025).

In addition to enhancing predictive performance, recent research has expanded into new areas, including explainable AI, federated learning, transfer learning, and foundation models for cybersecurity. The goal of these approaches is to make the model's behaviour more transparent, ensure privacy protection, render it more adaptable, and enhance its ability to learn across domains, especially in cloud and IoT contexts (Chinnasamy & Subramanian, 2025; Zhang & Tuo, 2025). However, despite ongoing methodological developments, most studies to date test their models on only one or two benchmark datasets and focus primarily on classification accuracy. The variations in dataset characteristics, feature distributions, data preprocessing, and attack diversity are thus frequently ignored, which makes direct comparison of proposed methods difficult (Dhote & Agrawal, 2026; Sravani et al., 2026). Recent research has also shown that when classification algorithms are used in conjunction with feature selection mechanisms, intrusion detection performance can be greatly enhanced while reducing computational complexity. In some of these methods, the need to select informative network features before model development is emphasized, particularly for benchmark datasets with high-dimensional traffic records (Mada et al., 2024; Sah et al., 2024; Zhang et al., 2025).

From this observation, it can be concluded that the current research focus remains primarily on the modeling approach. Although significant advances have been made in algorithm development, relatively little effort has been put into studying the properties of the benchmark datasets themselves and assessing their appropriateness for various intrusion detection research goals. This restriction calls for a systematic, benchmark-oriented analysis before proposing new AI-based intrusion detection models.

Benchmark datasets are essential for training, testing, and comparing intrusion detection algorithms, ensuring a consistent environment for developing and assessing such AI systems. With low access costs, a large number of features, and diverse attack scenarios, several publicly available datasets, such as CIC-IDS2017, UNSW-NB15, and IoT-23, are being used as de facto benchmarks for evaluating machine learning and deep learning solutions (Hoa, 2025; Ren, 2026). These datasets have been extensively adopted in recent studies to evaluate classical machine learning, ensemble learning, deep learning, explainable AI, and transfer learning models.

Even though it does not have the most recent updates, it is still one of the most widely used benchmark datasets due to the statistical features provided by the collected flows and the abundance of common enterprise network attacks such as brute-force, denial-of-service, distributed denial-of-service, web attacks, botnets, infiltration, and Heartbleed attacks (Kaliyaperumal et al., 2024; Oziegbe et al., 2026). In contrast, UNSW-NB15 was designed to address certain shortcomings of previous versions of the benchmark set, including the inclusion of new attack categories (such as exploits, fuzzers, reconnaissance, shellcode, and backdoor attacks), additional protocol information, and modern network traffic, allowing for the testing of more general intrusion detection models (Kalyana, 2026; Nzuva et al., 2024). More recently, the IoT-23 benchmark has gained ground as a representative test of IoT security, featuring realistic IoT traffic that includes botnets, C2 communication, DDoS attacks, and large-scale scanning activities by connected devices (Kodete et al., 2025; Zhang & Tuo, 2025).

Although these datasets are widely used, they are most often understood as equivalent assessment tools, as examined in the literature. Most studies select a benchmark dataset based on availability and/or popularity but do not

systematically examine its statistics, attack diversity, feature relevance, or appropriateness for specific studies. Moreover, few studies analyze variations in data distributions, class imbalance, feature engineering, and traffic behaviour before building models, making it difficult to compare model accuracy and fairness across published studies (Dhote & Agrawal, 2026; Saini et al., 2023). This makes it difficult to choose a suitable benchmark dataset, as it is still largely based on experience rather than evidence.

In addition to explainable AI, recent studies have sought to enhance benchmark quality through sophisticated data preprocessing and optimization methods. These include flow exporter optimization, handling class imbalance using the Synthetic Minority Oversampling Technique (SMOTE), feature optimization algorithms, generative adversarial networks (GANs) for data augmentation, and hybrid deep learning frameworks. These techniques have been shown to enhance the model's robustness and detection capabilities across a set of benchmark datasets. Still, most of them focus on optimizing the performance of the model in the selection of datasets, not the intrinsic features of the benchmark datasets themselves (Kamal & Mashaly, 2025; Memmesheimer et al., 2024; Nugroho et al., 2025), and are similar in other works (Pinto et al., 2024; Tian et al., 2024; Xie et al., 2025; Xu et al., 2026). Moreover, recent optimization research aimed at explaining the model has focused on reducing model latency while ensuring transparent decision-making, highlighting the paramount need for benchmark datasets that can support both high predictive accuracy and explainability (Hleha & Hol, 2025).

Recent works have further developed explainable intrusion detection by combining Federated learning, Explainable Boosting Algorithms (EBA), Neuro-symbolic Artificial Intelligence, and Transparent Machine Learning (TML) frameworks. These methods enhance the interpretability of the models while maintaining privacy and aiding analysts in making decisions in distributed cybersecurity settings. While they can improve the transparency and trustworthiness of AI models, their impact is highly reliant on the nature of the benchmark datasets they have been trained and evaluated on (AlMohamad, 2026; Assudani et al., 2025), and also in (Kwubeghari & Ezeji, 2025; Prajwalasimha et al., 2025). In recent research, the popular IoT-23 has been used to assess IDS performance in healthcare IoT, smart environments, anomaly detection, and industrial applications. Comparative investigations consistently demonstrate the effectiveness of machine learning and deep learning methods for IoT malware detection and highlight the importance of representative datasets that accurately reflect realistic IoT communication behaviours and attack patterns (Balega et al., 2024; Dash et al., 2024; Dubey et al., 2025). Similarly, other researchers have extended these efforts for realistic IoT cybersecurity research (Abdo et al., 2025; Mohamed et al., 2025). The fields of explainability, deep learning architectures for IoT traffic analysis, federated intrusion detection, fog computing security, and methods for managing highly imbalanced cybersecurity datasets are being explored, as well as areas beyond the scope of explainability. The studies mentioned above demonstrate the ongoing development of AI-based IDS, while also emphasizing that the choice of benchmark remains an important factor affecting the performance, scalability, and experimental repeatability of these systems in various deployment settings (Darwish & Roy, 2025; Li et al., 2024; Rajasa et al., 2023). Furthermore, emerging studies have explored cybersecurity challenges (Sreerenjith & Benitta, 2026; Tamuka et al., 2026).

Table 1 shows a summary of representative AI-based Intrusion Detection studies for the period 2024–2026. Table 1 provides an overview of representative studies on AI-based Intrusion Detection (2024- 2026).

Table 1. Summary of Representative AI-Based Intrusion Detection Studies (2024–2026)

Ref.	Dataset(s)	AI Technique	Classification	Main Finding	Limitation
Thomas et al. (2025)	UNSW-NB15	CNN	Binary	Achieved approximately 99% accuracy with superior feature extraction.	Evaluated on a single benchmark dataset.
Hussein et al. (2025)	UNSW-NB15, SmartGrid	CNN-GRU-LSTM	Binary/Multi-class	Hybrid deep learning improved detection accuracy and real-time performance.	Limited cross-dataset validation.
Sharma and Kumar (2025)	CIC-IDS2017, UNSW-NB15	CapsNet + BiLSTM	Multi-class	Improved robustness against diverse attack categories.	High computational complexity.
Kalyana (2026)	CIC-IDS2017, UNSW-NB15	CNN-LSTM	Binary/Multi-class	Combined spatial and temporal feature learning.	Limited explainability.
Huang et al. (2024)	CIC-IDS2017	Bagging-XGBoost	Binary	High F1-score using ensemble learning.	Requires extensive hyperparameter tuning.
Mada et al. (2024)	CIC-IDS2017, UNSW-NB15	IPOA-SVM	Binary/Multi-class	Feature optimization significantly improved classification performance.	Comparison limited to selected baseline models.

Alhassan et al. (2024)	NSL-KDD, CIC-IDS2017	Correlation-Based Feature Selection (CFS) + Autoencoder (AE)	Binary	Improved intrusion detection accuracy (94.32% on NSL-KDD and 97.71% on CIC-IDS2017) while reducing false alarms through feature selection and autoencoder-based classification	Evaluated only on binary classification and two benchmark datasets; multi-class performance and cross-dataset generalization were not investigated.
Balega et al. (2024)	IoT-23	XGBoost, SVM, DCNN	Binary	XGBoost achieved the best balance between accuracy and efficiency.	Limited evaluation on enterprise network traffic.
Dubey et al. (2025)	IoT-23	CNN-LSTM	Binary	Lightweight architecture suitable for healthcare IoT environments.	Application-specific validation only.
Abdo et al. (2025)	IoT-23	CNN, KNN	Multi-class	CNN outperformed conventional machine learning methods.	Limited comparison with recent transformer models.
Mohamed et al. (2025)	IoT-23	Hybrid ML-DL	Binary	Improved anomaly detection through hybrid learning.	Generalization across datasets has not been investigated.
AlMohamad (2026)	CIC-IDS2017, UNSW-NB15	Federated Transformer + SHAP	Binary/Multi-class	Combined explainability with federated learning.	Higher computational overhead.
Assudani et al. (2025)	CIC-IDS2017, UNSW-NB15	LightGBM + SHAP	Binary/Multi-class	Enhanced interpretability through feature attribution.	Dataset-specific preprocessing required.
Prajwalasimha et al. (2025)	CIC-IDS2017, UNSW-NB15	Transformer + Neuro-symbolic XAI (Causal Inference + Attention Visualization)	Binary/Multi-class	Achieved 97.5% detection accuracy with a 42% reduction in false positives while providing human-interpretable explanations and improving analyst decision-making speed by 35%	Higher computational complexity and increased implementation overhead due to the integration of transformer, neuro-symbolic reasoning, and explainable AI components.

The restrictions have underscored the need for a thorough benchmark characterization to assess the intrinsic characteristics of the most important intrusion datasets and to provide practical guidance for selecting the most suitable benchmark for AI-based intrusion detection.

The purpose of this is to provide a foundational understanding of the principles and components of EAI, including the requirements for benchmark-oriented IDSs.

With the growing complexity of AI-powered intrusion detection systems, one of the key research issues is model interpretability. In addition to accurate attack detection, security analysts need explanations that clarify the model's actions, provide recommendations for incident investigation, and help build trust in the automated cybersecurity system. To enhance the transparency of decisions, Explainable Artificial Intelligence (XAI) methods, such as SHAP (Sravani et al., 2026) and LIME (Chinnasamy & Subramanian, 2025), have been increasingly applied alongside machine learning and deep learning models, thus offering a method to reduce the opacity of decisions while maintaining their detection capabilities. Recent works have also integrated XAI with federated learning and transformer-like models to improve privacy protection, model interpretability, and collaborative threat detection in distributed settings (Kodete et al., 2025; Zhang & Tuo, 2025).

While these are important steps toward building trustworthy, model-based assessments, they are still predominantly prevalent. Many studies address the explainability of predictions from a single model trained on a particular benchmark dataset, and the role of dataset characteristics in explainability is seldom explored. However, in practice, the quality of explanations depends not only on the learning algorithm but also on the statistics of the underlying dataset, the feature distributions, the redundancy, and the variety of attacks used on the data. As a result, explainability results derived from a benchmark dataset are not always transferable to another dataset (Ren, 2026; Saini et al., 2023).

Likewise, recent benchmark studies have mostly focused on comparing classification accuracy across different datasets, with no attempt to examine the nature of those datasets. Selection of datasets is often based on popularity rather than on objective Information about feature richness, attack coverage, statistical diversity, or their suitability for different research goals. Detecting improvements in detection performance is complicated by the difficulty of distinguishing between improvements due to better learning algorithms and those due to more favourable data sets (Dhote & Agrawal, 2026; Hoa, 2025).

For this reason, a perspective grounded in benchmarks has to complement model-oriented research. Researchers

can only develop increasingly complex intrusion detection algorithms if they are familiar with the properties of benchmark datasets, including their strengths and weaknesses. This enables better dataset selection, more equitable comparisons in experiments, and stronger support for explainable, reproducible, and generalizable AI-based intrusion detection systems.

Despite the advancement of intrusion detection systems based on artificial intelligence, there are some basic issues that have not yet been solved. First, most of the previous research studies remain on predictive performance, with more and more sophisticated machine learning and deep learning architectures, and comparative under-representation of benchmark data sets used to evaluate these. This implies that the improvements are often reported as the classification accuracy, precision, recall, or F1-score without considering the impact of dataset characteristics (Hussein et al., 2025; Sagaran et al., 2025).

Secondly, all benchmark datasets have very different distributions of traffic, types of attacks, feature engineering, class imbalance, and statistical distributions. These differences in turn affect the selection of features, the complexity of the model, computational requirements, and the generalization capabilities of the model. In most of the previous studies, however, the datasets used for model development have not been systematically compared before but, rather, used to benchmark the models, which makes it difficult to conclude whether the better performance was due to the algorithms themselves or to the favourable characteristics of the datasets (Nzuva et al., 2024; Ren, 2026). Thirdly, new research challenges have emerged that demand more than predictive accuracy, including Explainable Artificial Intelligence, transfer learning, federated learning, and lightweight intrusion detection systems. A set of benchmark datasets with sufficient feature richness, a range of attack categories, statistical diversity, and realistic traffic behaviour is needed for these emerging paradigms. However, little guidance is available in the literature on the appropriateness of benchmark datasets for these research types (Sravani et al., 2026; Zhang & Tuo, 2025). Table 2 discusses the research gaps and the motivation for the proposed benchmarking framework. The most recent research in intrusion detection is investigating AI-driven cyber defense through systematic benchmarking, reinforcement learning, foundation models, lightweight deep learning architectures, and large language model-based cybersecurity. These new trends also underscore the need to select benchmark datasets to enable robust, explainable, transferable, and scalable AI models. However, prior research still prioritizes algorithmic advances over the development of objective methods for selecting benchmark datasets across tasks (Chizari et al., 2026; García et al., 2025; He et al., 2024). Furthermore, recent studies have explored multi-model deep learning frameworks, highlighting the growing need for benchmark datasets (Arul et al., 2026; Mittal & Rajvanshi, 2025).

Table 2. Research Gaps and Motivation for the Proposed Benchmark Framework

Research Gap	**Evidence from Literature**	**Why It Matters**	**How This Study Addresses the Gap**
Existing studies emphasize AI model development rather than analysis of benchmark datasets.	Alhassan et al. (2024), Kalyana (2026), and Thomas et al. (2025)	It makes it difficult to make a fair comparison among IDS models.	Performs comprehensive benchmark characterization of three major IDS datasets.
Benchmark datasets are selected mainly based on popularity.	Alhassan et al. (2024), Kalyana (2026), and Thomas et al. (2025)	Dataset choice may bias reported performance.	Introduces evidence-based benchmark selection guidelines.
Limited comparative analysis of CIC-IDS2017, UNSW-NB15, and IoT-23.	Abdo et al. (2025), Dhote and Agrawal (2026), and Nzuva et al. (2024)	Researchers lack an objective dataset comparison.	Provides statistical, feature, and attack diversity comparison.
Explainability studies mainly focus on AI models rather than benchmark datasets.	AlMohamad (2026), Assudani et al. (2025), Prajwalasimha et al. (2025)	Dataset characteristics also influence interpretability.	Evaluates feature importance and feature correlation across datasets.
No practical framework exists for selecting datasets according to research objectives.	Balega et al. (2024) and Ren (2026)	Researchers rely on trial-and-error dataset selection.	Develops a Benchmark Selection Framework for ML, DL, XAI, IoT, and transfer learning research.

Table 2 links the literature review with the proposed framework. In addition, while CIC-IDS2017, UNSW-NB15, and IoT-23 are the most widely used public benchmark datasets, comparative studies typically report only the detection performance of specific algorithms. Their statistical properties, feature importance, correlation patterns, strengths, weaknesses, and usability remain largely under-researched. Consequently, researchers still choose benchmark datasets based on their popularity or prior studies conducted on them, rather than on evidence and well-defined evaluation (Dhote & Agrawal, 2026; Kodete et al., 2025).

This study presents a framework that shifts the focus of research from comparing models to characterizing datasets, thereby overcoming these limitations. The proposed framework systematically compares the statistical properties, attack distributions, feature importance, correlation patterns, strengths, and weaknesses of CIC-IDS2017, UNSW-NB15, and IoT-23, and offers practical guidelines for selecting a benchmark; it also lays the groundwork for the creation of more explainable, reproducible, and generalizable AI-based intrusion detection systems.

In light of the identified research gaps, this study proposes a benchmark AI-based intrusion detection framework that prioritises dataset characterisation prior to model development. The framework proposed in this paper differs from

existing work, which mainly compares intrusion detection algorithms; it systematically analyses the intrinsic properties of benchmark datasets and, based on the evidence, helps select appropriate datasets for various research objectives. Figure 2 demonstrates the evolution of AI-based Intrusion Detection Research and the Position of the Proposed Benchmark Framework.

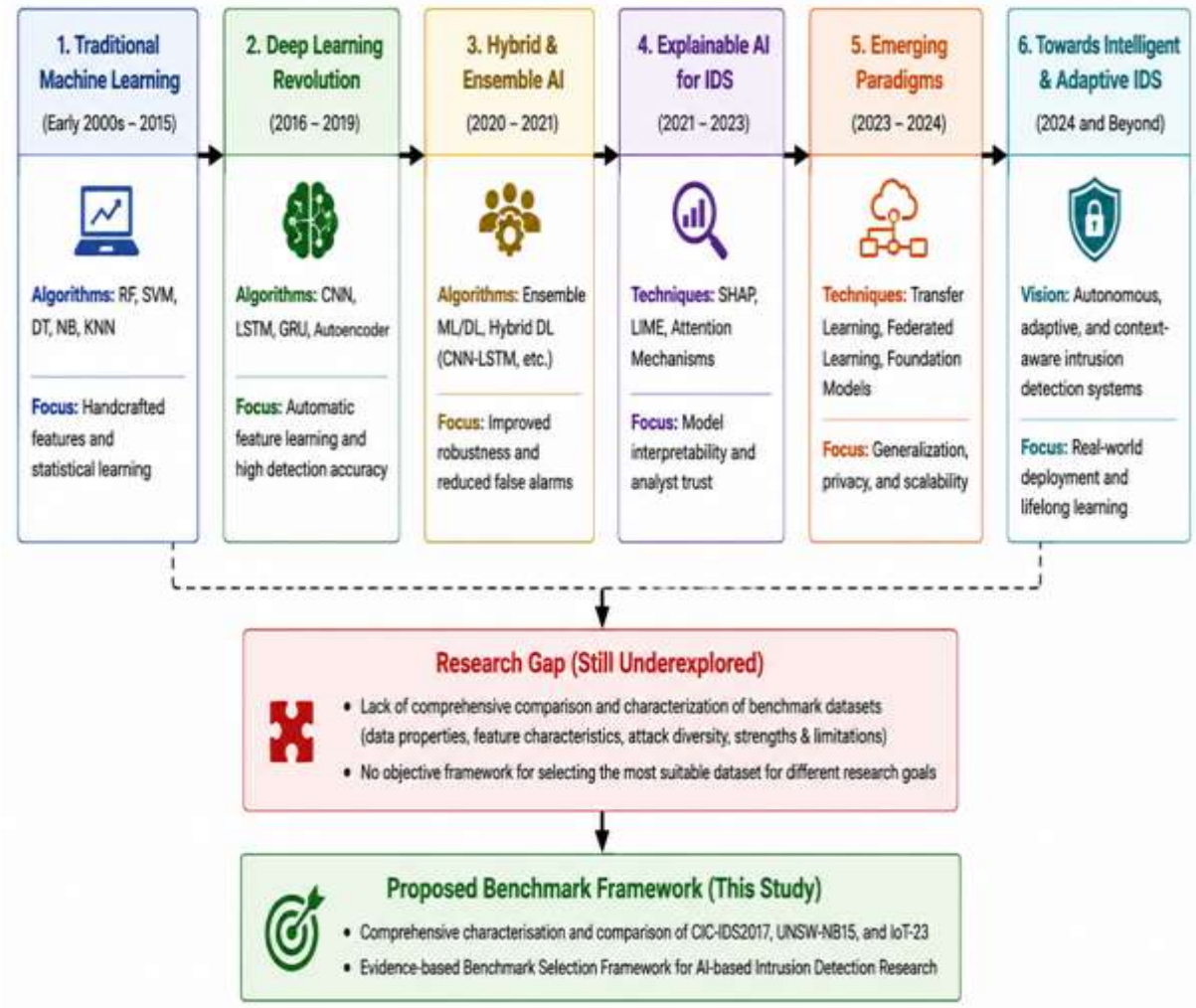

Figure 2. Evolution of AI-Based Intrusion Detection Research and Positioning of the Proposed Benchmark Framework
Source: Developed by the authors; AI-assisted visualization was used solely for graphical design

The framework is made up of five complementary stages. First, each benchmark dataset is statistically profiled to analyze data quality, class distribution, traffic composition, and attack diversity.

Second, the concept of feature relevance is introduced to determine the most informative features in intrusion detection. Third, a correlation analysis is conducted to examine feature redundancy and inter-feature relationships that could affect the model's complexity and interpretability. Fourth, each benchmark dataset is compared with the others using a range of parameters, including attack coverage, feature richness, statistical diversity, computational complexity, and suitability for other AI paradigms. Finally, the results are compiled into a Benchmark Selection Framework that provides practical guidance for choosing the right dataset for various research tasks, including classical machine learning, deep learning, explainable AI, IoT security, feature selection, transfer learning, and cross-dataset generalization.

This study does not present a new intrusion detection algorithm; rather, it introduces a model-independent point of view based on a benchmark approach that complements research on model-oriented approaches. The proposed framework is designed to ensure that future intrusion detection studies are conducted in a fairer, more reproducible, and more practically relevant manner, allowing researchers to go beyond popularity or convention to select benchmark datasets that are validly chosen using objective analytical evidence. Thus, the framework can serve as a reference for designing strong, explainable, and generalizable AI-based IDS.

MATERIALS AND METHODS

The proposed benchmark framework was implemented through a systematic five-stage methodology, as illustrated in Figure 3. The framework guides the comparative characterization, evaluation, and selection of benchmark intrusion detection datasets prior to AI model development.

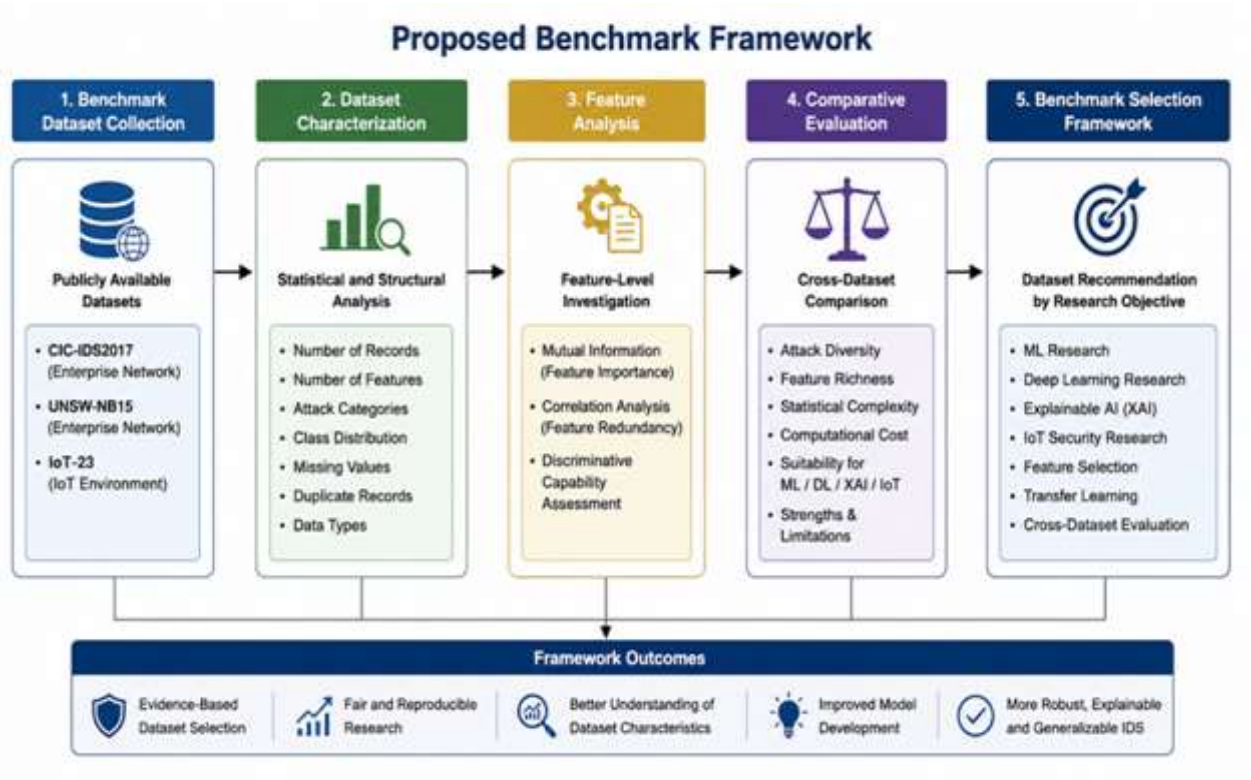

Figure 3. The proposed Benchmark framework
Source: Developed by the authors; AI-assisted visualization was used solely for graphical design

There are five sequential steps in the framework. First, benchmark datasets are gathered from publicly available, well-known cybersecurity repositories. The datasets were chosen to represent enterprise network traffic, modern network attacks, and Internet of Things (IoT) environments, with three datasets representing these, respectively: CIC-IDS2017, UNSW-NB15, and IoT-23.

The second approach is to perform a detailed statistical characterization of each dataset to assess its data quality, including the number of records, feature dimensions, attack categories, class distributions, missing values, duplicate records, and overall dataset composition. This stage provides quantitative insight into the structural differences among the benchmark datasets.

Third, feature-level analysis is used to examine the discriminative power of individual features. Mutual Information is used to identify the most informative features for attack detection, and correlation analysis is performed to assess inter-feature relationships and feature redundancy that can affect the model's complexity and explainability.

Fourth, the analytical results from the three benchmark datasets are compared to one another, and the advantages and disadvantages of the datasets and their appropriateness are determined for various intrusion detection research situations. This comparative evaluation considers several criteria, including attack diversity, feature richness, statistical complexity, computational requirements, and applicability to machine learning, deep learning, explainable AI, and IoT security research.

Lastly, the results are incorporated into a Benchmark Selection Framework that offers practical guidance on selecting the most suitable dataset for specific research needs. The proposed framework does not introduce a new intrusion detection algorithm but rather provides a benchmark-oriented view that enables the design of fair experiments, enhances the reproducibility of AI-based intrusion detection research, and promotes the development of more powerful, explainable, and transferable cybersecurity solutions.

Benchmark Datasets

In this work, we evaluate three popular public benchmark datasets used in intrusion detection research, namely CIC-IDS2017, UNSW-NB15, and IoT-23. The datasets were chosen to cover the most common scenarios for testing enterprise networks, modern network environments, and Internet of Things (IoT) networks with AI-based intrusion detection systems.

The Canadian Institute for Cybersecurity (CIC) developed CIC-IDS2017 to provide realistic enterprise network traffic that includes a range of benign activities and cyberattack scenarios, such as Distributed Denial-of-Service (DDoS), Denial-of-Service (DoS), brute-forcing, web attacks, botnets, infiltration, and Heartbleed attacks. This dataset includes over 2.8 million network flows, characterized by flow attributes measured using about 80 flow statistics collected by CICFlowMeter.

The Australian Center for Cyber Security (ACCS) created UNSW-NB15 to address the drawbacks of previous versions of benchmark datasets, including the inclusion of recent network traffic and current attack categories. It has about 2.54 M records and 49 network traffic features, and includes attack types such as Generic, Exploits, Fuzzers, DoS, Reconnaissance, Analysis, Backdoor, Shellcode, and Worms.

IoT-23 is a new, large-scale benchmark just for IoT security research. The dataset comprises about 72 million

network connections from realistic IoT environments. It includes not only a benign traffic set but also three sets of network traffic representing numerous malware families, botnet activity, command-and-control (C&C) activity, DDoS traffic, and large-scale scanning behaviour. IoT-23 focuses on real-world IoT network behaviour and provides a useful basis for assessing intrusion detection techniques in resource-limited environments compared with other benchmark datasets.

The three benchmark datasets vary significantly in scale, features, types of attacks, and application areas. They therefore offer complementary views on the appropriateness of benchmark datasets for various research goals in AI-based intrusion detection. Results and discussion for these datasets are described and compared in detail.

Dataset Characterization

To ensure a consistent basis for comparison, each dataset was thoroughly characterized prior to feature analysis. The goal of this stage was to explore the characteristics of the benchmark datasets, not to evaluate the performance of a single intrusion detection algorithm.

The characterization process was carried out in several steps. First, descriptive statistics were computed to summarise the overall structure of each dataset, including the number of records, features, traffic classes, attack categories, missing values, and duplicate records. These statistics are just preliminary measures of dataset quality and data complexity.

Second, the distribution of benign and malicious Traffic was analyzed alongside the relative frequency of attack types. The analysis was performed to assess the class imbalance, attack diversity, and the representativeness of each benchmark set for intrusion detection research.

Thirdly, feature-level statistics were examined to gain insight into the available attributes of network traffic. The three benchmark datasets have different feature representations and applications, so this analysis helps to compare their structural properties consistently prior to feature evaluation.

The statistical summarisation results provide the basis for the comparative analysis in the Results and Discussion section, where similarities, differences, strengths, and weaknesses of the benchmark datasets are systematically explored.

Feature Analysis

For the three benchmark datasets, the discriminative capability and structural relationships among network traffic attributes were evaluated through feature analysis. Rather than the modelling results, this step focused on determining which features are most informative and exploring potential redundancy among the model variables that may affect the subsequent development of an AI-based intrusion detection model.

Firstly, Mutual Information (MI) was used to measure the dependence between each feature and the class label. Unlike linear correlation measures, Mutual Information can capture both linear and nonlinear relationships and assess the relevance of features in network traffic for intrusion detection. Features with higher MI values were deemed more indicative of the distinction between normal and malicious Traffic and were then ranked by relative importance.

Secondly, feature correlation analysis was conducted using the Pearson correlation coefficient to assess correlations among features that contained the most Information. A correlation matrix was created independently for each benchmark dataset to assess the extent of correlations among attributes, which can introduce redundancy and complexity and affect the interpretability of an AI-based intrusion detection system.

The results obtained from the combined use of Mutual Information and correlation analysis are complementary in assessing feature relevance and feature dependencies. The overlap of these informative features, or the provision of complementary Information, is measured using correlation analysis, while the contribution of each feature to attack discrimination is determined using Mutual Information. The results of these comparisons contribute an invaluable component to the proposed benchmark framework, which includes a support system for evidence-based dataset evaluation and selection, as well as features for future intrusion detection research.

Benchmark Evaluation Framework

The statistical analyses and feature investigation conducted in this study have led to the development of a Benchmark Evaluation Framework that enables objective comparison of benchmark intrusion datasets. The framework was developed to evaluate the suitability of each benchmark dataset for various research goals of intrusion detection systems using Artificial Intelligence (AI) rather than the predictive performance of a particular learning algorithm.

The evaluation framework combines the results from dataset characterization, attack diversity analysis, feature importance ranking, and feature correlation analysis. These supplementary analyses provide a complete picture of the structural characteristics of each benchmark dataset, as well as a systematic comparison of their analytical benefits and drawbacks.

The framework evaluates benchmark datasets based on several criteria relevant to today's intrusion detection research. The criteria address data volume, attack diversity, feature richness, statistical complexity, feature redundancy, explainability, computational demands, and applicability across research areas such as explainable AI, classical machine learning, deep learning, Internet of Things (IoT) security, feature selection, transfer learning, and cross-dataset generalization.

The proposed evaluation framework does not quantify or rank benchmark datasets through a single analytical lens; instead, it enables assessment from multiple perspectives, providing a balanced evaluation of the suitability of specific benchmark datasets for particular research contexts. This multidimensional assessment allows researchers to choose benchmark datasets based on objective evidence, research needs, and other factors, rather than solely on popularity or use in the literature.

The results of this framework are combined to form a Benchmark Scorecard and a Benchmark Selection Matrix in

the Discussion section. In sum, they offer concrete advice on selecting relevant benchmark sets and lay the groundwork for the proposed Benchmark Selection Framework for AI intrusion detection research.

RESULTS

This section compares the three benchmark intrusion detection datasets used in this work. Their goal is to study their structural characteristics, such as dataset size, feature richness, traffic composition, and other data quality metrics, to provide an objective evaluation of their appropriateness for various research scenarios in intrusion detection using Artificial Intelligence.

Table 3. Comparative Dataset Overview

Dataset	Records	Features	Attack Classes	Normal %	Attack %	Missing Values	Duplicate Rows
CIC-IDS2017	2,830,743	80	15	80.3	19.7	0	0
UNSW-NB15 (Original)	2,540,047	50	10*	87.35	12.65	2,778,024**	480,629
IoT-23	71,984,818	18	8	16.09	83.91	7,526	1,650,422

Notes: "For consistency, the "Backdoor" and "Backdoors" categories in the original UNSW-NB15 dataset will be merged into a single "Backdoor" class in the subsequent analyses. ** The reported missing values in the original UNSW-NB15 dataset are primarily associated with the attack_cat attribute for normal traffic records and represent expected labelling characteristics rather than data quality issues.

The three benchmark datasets. The main features are presented in Table 3. The largest dataset is IoT-23, comprising about 72 million network connections, which is significantly bigger than CIC-IDS2017 (2.83 million records) and UNSW-NB15 (2.54 million records). This is a significant difference, making IoT-23 well suited to large-scale IoT network traffic studies and deep learning applications. In contrast, CIC-IDS2017 offers the highest number of features (80), UNSW-NB15 offers 50 static flow-based features, and IoT-23 offers 18 connection-based features.

Each dataset also differs significantly in the attacks used. CIC-IDS2017 includes the highest number of attack categories (15) and offers more comprehensive coverage of enterprise network threats than UNSW-NB15, which has 10 modern attack categories representing a modern network environment. The eight traffic categories are primarily used in Internet of Things security research, and IoT-23 primarily focuses on IoT malware and botnet activities. In addition, there are apparent differences in class distributions across the datasets. The Traffic for CIC-IDS2017 and UNSW-NB15 is mostly benign, while IoT-23 has a much higher percentage of malicious Traffic. The original UNSW-NB15 dataset also has missing values in the attack category attribute and a relatively large number of duplicate records; CIC-IDS2017 has complete data quality, with no duplicate or missing records.

Figure 4 provides a comparative overview of the benchmark intrusion detection datasets in terms of the number of data packets, feature diversity, and traffic diversity.

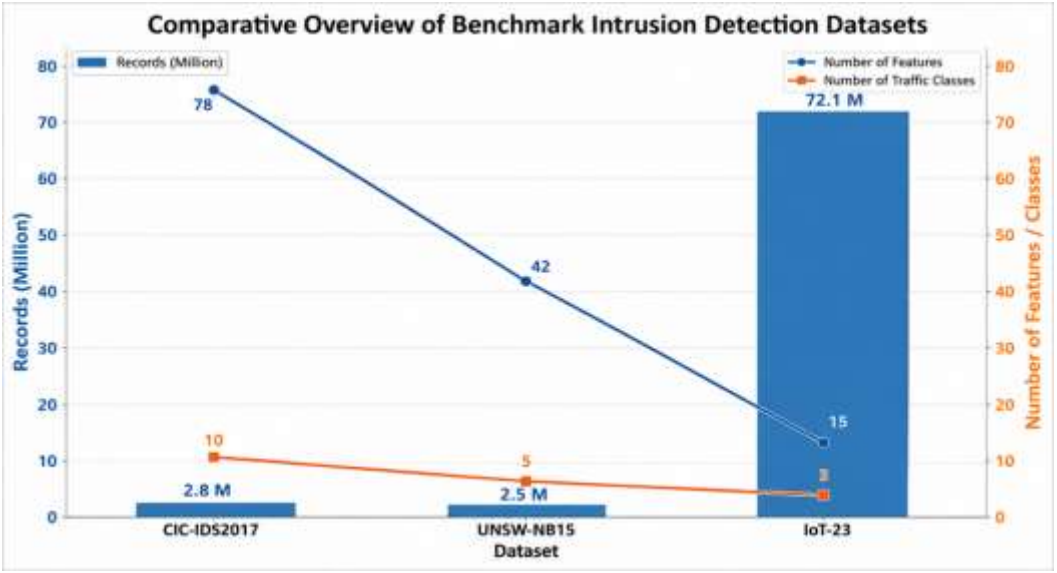

Figure 4. Comparative overview of the benchmark intrusion detection datasets in terms of data volume, feature richness, and traffic diversity

Source: Developed by the authors; AI-assisted visualization was used solely for graphical design

A visual comparison of the benchmark datasets, based on data volume, number of features, and traffic diversity, is shown in Figure 4. It is clear from this figure that no single benchmark dataset contains the most records, features, or attacks. The difference between IoT-23 and CIC-IDS2017 is that, while IoT-23 provides an unparalleled volume of data for developing computationally intensive deep learning models, the latter offers the most comprehensive feature space and the broadest attack coverage. UNSW-NB15 lies somewhere in the middle in terms of dataset size, state-of-the-art attack usage, and feature complexity.

The results above demonstrate that benchmark datasets exhibit analytical features that directly impact AI-based intrusion detection research. So, it is important to choose benchmarks based on the study's purposes, not because they are popular. The proposed Benchmark Selection Framework outlined in this paper is based on this observation.

Attack Diversity Analysis

The diversity of attacks is an important attribute of benchmark intrusion detection datasets, as it defines the scope of cyber threats for training and testing Artificial Intelligence (AI)- based intrusion detection systems. This section compares the attack composition of CIC-IDS2017, UNSW-NB15, and IoT-23 in terms of the ratio of normal to malicious Traffic and the variety and distribution of attack categories. Table 4 compares the distribution of attacks among the three datasets.

Table 4. Attack distribution comparison across the three datasets

Dataset	Normal Traffic (%)	Attack Traffic (%)	Largest Attack Category	Largest Attack (%)	Smallest Attack Category	Smallest Attack (%)	Attack Categories
CIC-IDS2017	80.3	19.7	DoS Hulk	8.16	Heartbleed	0.0004	15
UNSW-NB15	87.35	12.65	Generic	8.48	Worms	0.0069	9*
IoT-23	16.09	83.91	PartOfAHorizontal PortScan	61.01	C&C-PartOfAHorizontalPort Scan	0.0012	8

Note: BackDoor and Backdoors were merged into a single "Backdoor" category for consistency.

The distribution of attacks in the three benchmark datasets is shown in Table 4. There is a significant difference in the composition of the Traffic and the diversity of attacks. CIC-IDS2017 has 80.30% benign traffic and 19.70% malicious traffic across 15 attack categories, providing broad coverage of enterprise network threats. DoS Hulk is the most common Attack, while Heartbleed is the least common, reflecting a highly imbalanced set of attack classes. When the Backdoor classes are merged, UNSW-NB15 has an even higher percentage of normal Traffic (87.35%) and 12.65% attack Traffic across 9 attack classes. Of these attacks, the most common are in the Generic category, and the Worms are very small in comparison. Conversely, the malicious Traffic in IoT-23 makes up a significant majority of all Traffic, at 83.91%. Part of the Horizontal Port Scan traffic is the predominant Traffic in the dataset, while command-and-control scanning activities occur infrequently. Figure 5 compares the attack distributions.

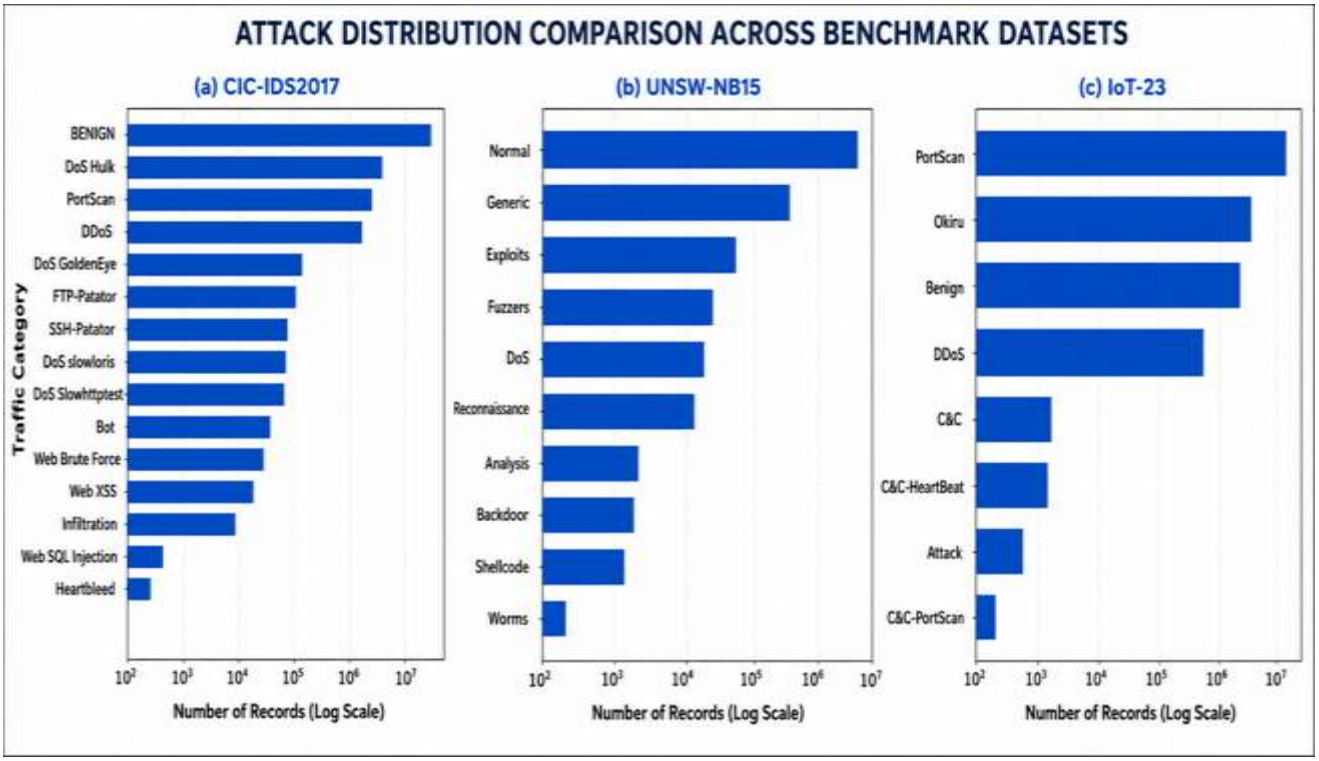

Figure 5. Attack distribution comparison

The attack distributions for the three benchmark datasets are shown in Figure 5. The figure clearly shows that the sets are not different versions of the same benchmark but are different cybersecurity environments. CIC-IDS2017 covers the most diverse enterprise attacks and can be used to test multi-class intrusion detection algorithms and feature selection methods. UNSW-NB15 is a set of network attacks that are reasonably well-balanced and that have a relatively low level of class imbalance, suitable for general-purpose intrusion detection research. On the other hand, IoT-23 is attack-centric and

captures realistic IoT malware behaviour, thus being highly beneficial for botnet detection, IoT security research, and large-scale anomaly detection research.

These observations indicate that attack variety is a function not only of the number of attack categories but also of the composition and class distribution of the Traffic. When choosing a benchmark dataset, the types of attacks included, as well as the proportion of benign and malicious Traffic, impact the training process, the dataset's fairness, and the overall ability of the AI-based intrusion detection system (IDS) to generalize. The results obtained from this research also justify the need for the proposed Benchmark Selection Framework presented in this study.

Feature Importance Analysis

Mutual Information (MI) was used to quantify the feature importance of network traffic attributes and assess their discriminative power in each benchmark dataset. Unlike traditional correlation methods, Mutual Information considers both linear and nonlinear associations between individual features and the target class and can identify the features that best help detect intrusions. Figure 6 shows the details of the Feature Comparison Based on Mutual Information.

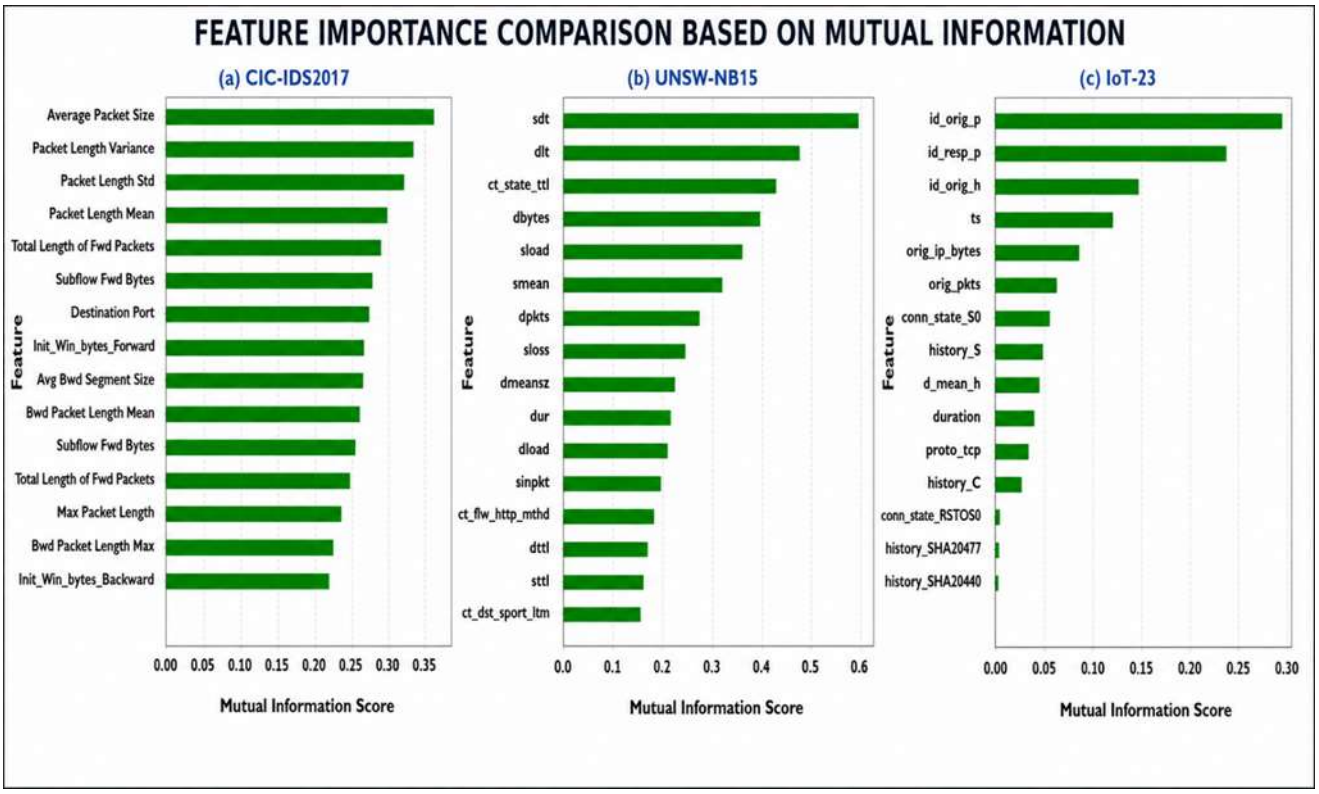

Figure 6. Feature Comparison Based on Mutual Information

Comparisons of the top features across CIC-IDS2017, UNSW-NB15, and IoT-23 are shown in Figure 6. The results show that each benchmark dataset highlights different aspects of network behaviour. The top three statistics in CIC-IDS2017 are packet size and packet length; the top four are Average Packet Size, Packet Length Variance, Packet Length Standard Deviation, and Packet Length Mean, all of which have the highest Mutual Information scores. Other high-level properties, such as backward packet statistics, destination port, and TCP window size, suggest that statistical flow properties are good discriminators of benign from malicious Traffic.

The most informative attributes of UNSW-NB15 are more related to protocol behaviour and traffic dynamics. The lifetime of a packet, protocol state information, and traffic volume are among the top indicators that show up as features with the highest importance scores, such as sttl, dttl, ct_state_ttl, sbytes, and sload, indicating that these are important metrics to look at when trying to distinguish contemporary network attacks. Unlike CIC-IDS2017, UNSW-NB15 is more focused on network protocol characteristics than on statistical packet measurements.

In IoT-23, the two most informative attributes are communication endpoint information: id_orig_p and id_resp_p. ts and orig_ip_bytes are also ranked as top features. The results show that IoT attacks are well correlated with communication patterns, network endpoints, and temporal characteristics, which are less complex than statistical flow descriptors. This is in line with the fact that IoT malware and botnets rely on communication.

Overall, the comparison shows that feature importance is highly dependent on the dataset and reflects each benchmark's design goals. CIC-IDS2017 offers a wide variety of statistical flow features suitable for feature engineering and explainable AI studies; UNSW-NB15 highlights protocol-aware network behaviour relevant to current intrusion detection research; and IoT-23 aims to describe communication features of particular interest to the field of Internet of Things security. The results in this study highlight the importance of taking into account the nature of the features (the feature space) available to the AI model because this will affect feature selection strategies, model interpretability, and the

performance of the AI model, which is also dependent on the size and diversity of the benchmark dataset used.

Feature Correlation Analysis

The relationships among some of the most informative attributes identified by the Mutual Information analysis were explored using feature correlation analysis. Feature dependency is crucial because if features are highly correlated, it can introduce noise and computational complexity, making the model less interpretable. On the other hand, a balanced feature space with complementary Information can enhance the effectiveness of feature selection and boost the confidence of AI-based intrusion detection systems. Figure 7. Shows the Feature Correlation of Top Information Features.

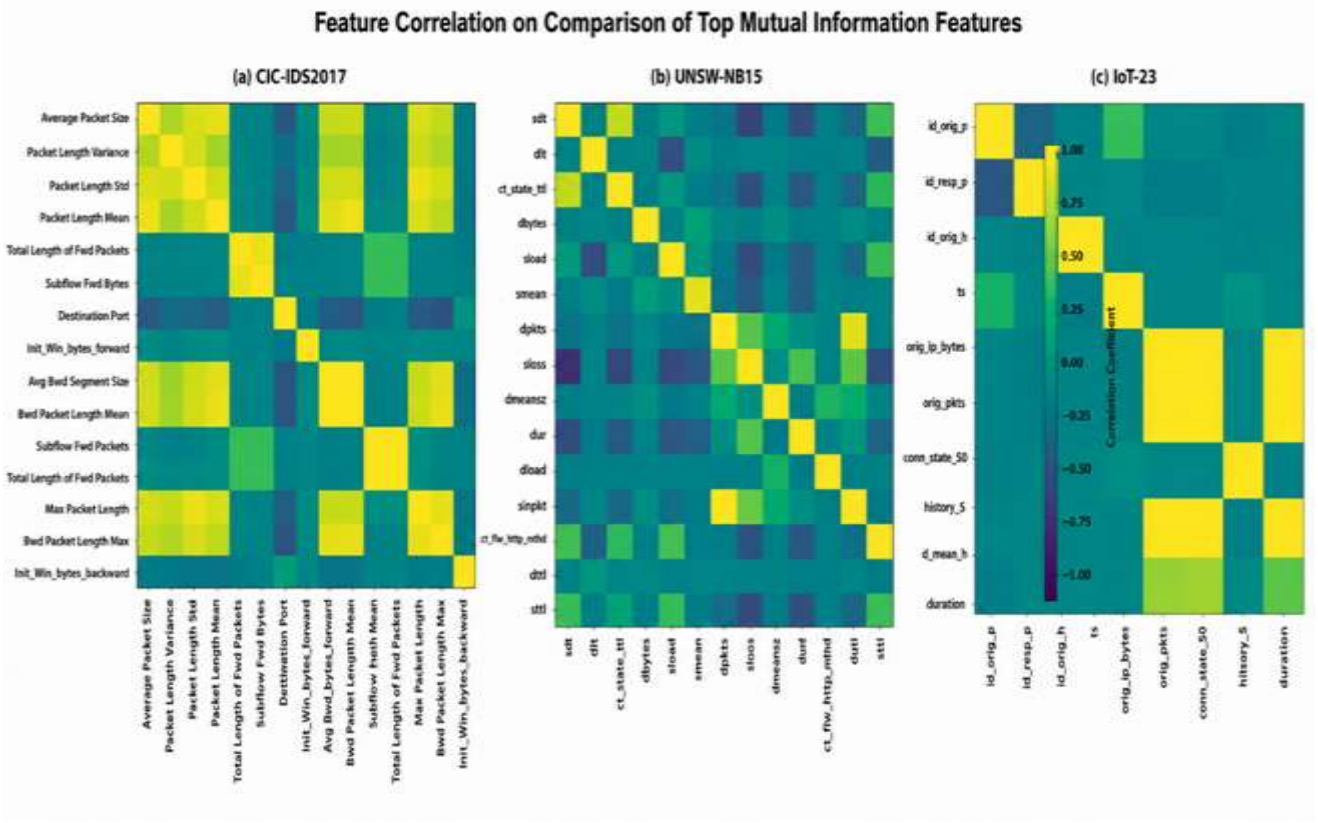

Figure 7. Feature Correlation of Top Information Features

The correlation matrices of the top Mutual Information features, for the three benchmark datasets, are shown in Figure 7. Different correlation patterns are observed, each associated with the various design philosophies and application areas of the datasets. In CIC-IDS2017, the packet length and packet size features are highly correlated, suggesting that a combination of statistical flow parameters captures similar network behaviour. These features are highly correlated and provide a lot of descriptive Information, but they also indicate feature redundancy, which could necessitate dimensionality reduction or feature selection when building a model.

UNSW-NB15 has a more balanced correlation structure, with most protocol- and traffic-related features showing weak to medium correlations. This means the dataset is complementary within feature groups, so AI models do not have to learn many similar features to make accurate predictions across different aspects of a network. This balanced representation of features makes UNSW-NB15 a good choice for testing traditional machine learning methods and state-of-the-art deep learning architectures.

The feature correlation of each of the three benchmark datasets is lowest for IoT-23. Most other communication attributes have low interdependency, except for a few highly related attributes, such as endpoint identifiers and traffic volume measurements. This implies that IoT-23 represents a wide range of communication activities produced by IoT devices and malware, and offers complementary Information for detecting abnormal communication flows in the network, while avoiding over-representation of features.

In general, the comparative correlation analysis shows that each benchmark dataset has a different feature dependency structure. The statistical feature space of CIC-IDS2017 is rich, with some redundancy; the network behaviour of UNSW-NB15 is balanced across modern networks, and the features of IoT-23 are independent and communication-centric. Such variations are crucial in feature engineering, dimensionality reduction, explainable AI, and model complexity. Therefore, feature correlation should be considered alongside feature importance and attack diversity when selecting benchmark datasets for AI-based intrusion detection studies.

DISCUSSIONS

As shown in Section 4, there are no universal best intrusion detection benchmark datasets for all research on Artificial Intelligence (AI) based cybersecurity. Rather, each dataset has different features in its scale, representation, attack diversity, and feature dependencies, and therefore is appropriate for different research goals. The remark emphasizes the need to choose benchmark datasets for the analysis from among the many available options, not just because they are trendy in the research community, but because they are analytically relevant. Table 5 demonstrates the details of the Comparative Benchmark Evaluation Matrix.

Table 5. Comparative Benchmark Evaluation Matrix

Evaluation Criterion	CIC-IDS2017	UNSW-NB15	IoT-23	Justification
Data Volume	★★★☆☆	★★★☆☆	★★★★★	Based on the total number of records.
Feature Richness	★★★★★	★★★★☆	★★★☆☆	Based on the number and diversity of available features.
Attack Diversity	★★★★★	★★★★☆	★★★★☆	Based on the number and variety of attack categories.
Class Balance	★★★☆☆	★★★☆☆	★★☆☆☆	Based on normal/attack distribution and minority classes.
Explainable AI Suitability	★★★★★	★★★★☆	★★★☆☆	Based on the interpretability of features and feature attribution suitability.
Deep Learning Suitability	★★★★☆	★★★★☆	★★★★★	Based on dataset size and suitability for high-capacity models.
IoT Security Suitability	★☆☆☆☆	★★☆☆☆	★★★★★	Based on relevance to IoT environments and malware traffic.
Feature Selection Suitability	★★★★★	★★★★☆	★★★☆☆	Based on feature dimensionality and ranking potential.
Cross-Dataset Research Suitability	★★★☆☆	★★★★★	★★★★☆	Based on suitability for transferability and generalization studies.
Computational Efficiency	★★★★☆	★★★★☆	★★☆☆☆	Based on expected processing cost and dataset complexity.

Comparative evaluation of the three benchmark datasets is summarised in Table 5, with respect to the major analytical criteria studied in this paper. CIC-IDS2017 stands out for its high feature richness, attack diversity, suitability for AI explainability, and feature selection, achieved through a rich statistical flow representation and extended coverage of enterprise attack scenarios. The aforementioned attributes particularly make it well-suited for explainable machine learning, feature engineering, and multi-class intrusion detection studies.

The most balanced dataset among the three is UNSW-NB15. It is the moderate dimensionality of the features, modern attack methods, and the complementary feature relationships that make it applicable to transfer learning, cross-dataset assessment, and creation of general-purpose intrusion detection models. It provides a more up-to-date dataset of network attack data and more efficient computation than CIC-IDS2017.

IoT-23 is unique in providing the massive amount of data needed and in the realistic Internet of Things traffic. Due to its large data volume and the presence of malicious behaviours, the dataset is well-suited for deep learning and large-scale intrusion detection and IoT security applications. While IoT-23 is not as comprehensive as the other benchmark datasets, its communication-focused attributes are good at capturing the behaviour of IoT networks and offer useful Information for anomaly detection in resource-constrained environments.

Overall, the benchmark scorecard validates the approach to selecting benchmarks for a research scenario, rather than merely serving as an evaluation criterion. For studies on explainability and feature engineering, CIC-IDS2017 is likely the most suitable dataset; for research on modern network environments or transfer learning, UNSW-NB15 is the better choice. Similarly, IoT-23 is the recommended test set for IoT-specific cybersecurity studies and heavy-duty deep learning operations. The results of these comparisons will provide the basis for the benchmark selection process described next, which turns the comparative analytical results into a guide for selecting datasets for use.

Benchmark Selection Framework

The previous sections have comparatively examined the characteristics of benchmark intrusion datasets; in this section, the results of this analysis are presented as a useful decision-support tool for researchers. The framework does not recommend a single benchmark dataset for all studies on Artificial Intelligence (AI)- based Intrusion Detection; rather, it helps researchers choose the right benchmark dataset for their research goals. This is the main contribution of the present study and introduces a methodology to make comparative dataset analysis an evidence-based benchmarking process. Table 6 presents detailed Information on the developed Benchmark Selection Matrix.

Table 6. Benchmark Selection Matrix

Research Objective	Recommended Dataset	Reason
Classical Machine Learning	CIC-IDS2017	Rich numerical flow features and manageable size.
Deep Learning	IoT-23 / UNSW-NB15	Large-scale records and diverse traffic patterns.
Explainable AI	CIC-IDS2017	Feature-rich structure supports transparent feature attribution.
IoT Security	IoT-23	Contains realistic IoT malware, botnet, C&C, and scanning Traffic.
Feature Selection	CIC-IDS2017	A large number of statistical features support feature ranking.
Transfer Learning	UNSW-NB15	Balanced modern network representation supports transfer experiments.
Cross-Dataset Evaluation	All three datasets	Combining enterprise, modern, and IoT traffic improves robustness evaluation.
Large-Scale IDS	IoT-23	Very large record volume supports scalable model development.
Multi-Class IDS	CIC-IDS2017 / UNSW-NB15	Broad attack coverage supports multi-class classification.

In Table 6, some practical suggestions are offered for choosing benchmarks. These datasets are for various intrusion detection research scenarios. The selection matrix indicates that CIC-IDS2017 is the best benchmark for classical machine learning, explainable AI, feature selection, and multi-class intrusion detection, owing to its diverse attack types and rich statistical flow features. With its well-balanced feature representation and contemporary network attacks, UNSW-NB15 is recommended for transfer learning and cross-dataset generalization studies. However, IoT-23 is best suited for large-scale intrusion detection research, Internet of Things security, and deep learning applications because of the large amount of data it contains and the realistic IoT malware traffic it includes. It is recommended to use all three benchmark datasets to assess the model's robustness and generalization across heterogeneous network environments. The proposed Benchmark Selection is shown in Figure 8.

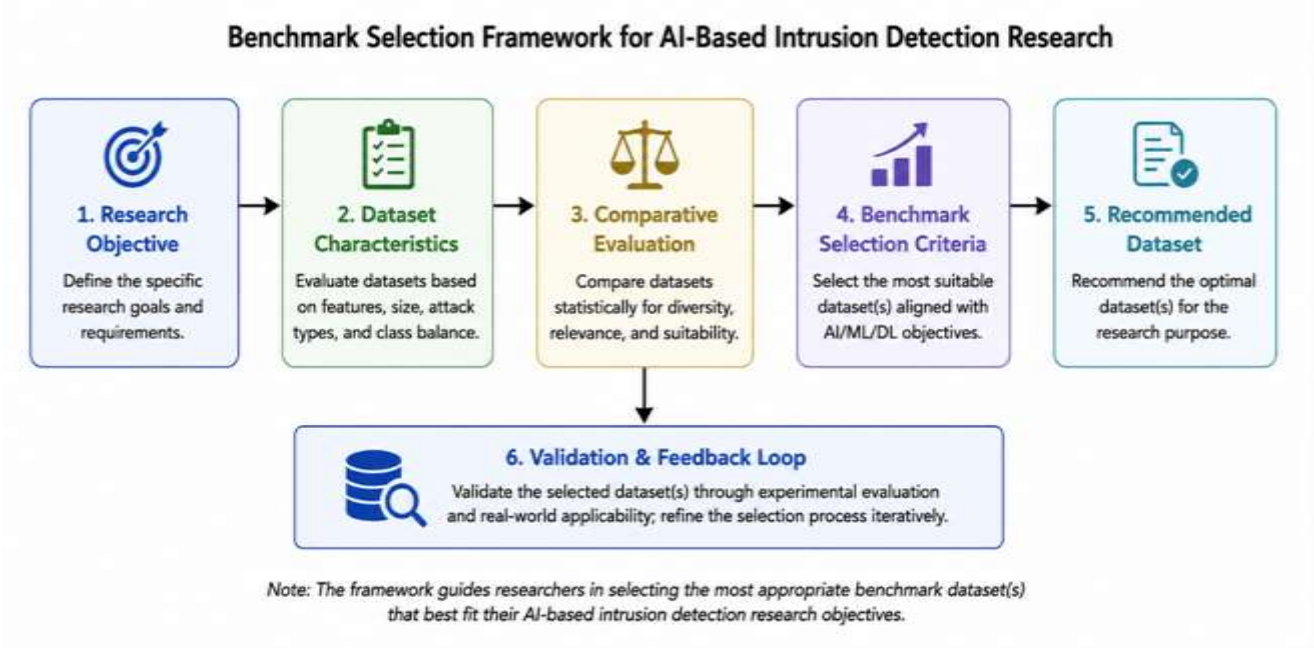

Figure 8. The proposed Benchmark Selection

This study suggests an overall Benchmark Selection Framework as shown in Figure 8. The proposed framework differs from current benchmark comparisons that primarily report descriptive statistics on datasets by incorporating a decision process to select datasets aligned with research goals. The framework starts with the research scenario definition and ends with analyses of statistical properties, attack diversity, and feature importance and correlation in the datasets. Using this evidence, researchers can systematically select the benchmark data most appropriate for their experiments, rather than following convention or the popularity of datasets.

The proposed framework also has several practical benefits. First, it provides for an objective, transparent selection of benchmarks by using multiple characteristics in the data set. Secondly, it enhances the repeatability and comparability of future intrusion detection research by encouraging researchers to justify their dataset choice through analysis. Lastly, the framework offers a portable guideline that can be extended to incorporate new benchmark datasets and cybersecurity domains, enabling the creation of more accurate, trustworthy, and widely applicable AI intrusion detection systems.

Future Research Directions and Limitations

While the proposed benchmark framework will bring many benefits, there are also several points to note. The first is that this study has been conducted on three widely used intrusion detection benchmarks: CIC-IDS2017, UNSW-NB15, and IoT-23. These are enterprise networks, modern network environments, and Internet of Things (IoT) ecosystems, but do not include the complete range of public cybersecurity benchmarks. The framework suggested, therefore, is to be interpreted as an example guideline for a representative dataset and not as a complete assessment of all intrusion detection benchmarks.

Second, the evaluation framework is based on analyses of attack diversity and feature importance, as well as feature correlations and the characterization of the comprehensive datasets, rather than on the performance of specific Artificial Intelligence (AI) algorithms. This design was purposeful, as the scope of this study is to aid evidence-based benchmark selection prior to the development of a model. However, a potential direction for future development of this framework is to include additional algorithmic performance metrics, such as computational efficiency, model robustness, and explainability, to enhance the recommendation of these datasets further.

Third, the framework for selecting benchmarks provides qualitative recommendations grounded in analytical evidence. Automated ranking of benchmark datasets based on user-defined research priorities and evaluation weights is a possible topic for future research that may be pursued using quantitative multi-criteria decision-making approaches.

There are several points of interest for future research that are raised by this work. The proposed framework can be expanded to other benchmark datasets as Bot-IoT, TON_IoT, CSE-CIC-IDS2018, NF-UQ-NIDS to improve its applicability in different cybersecurity scenarios. Furthermore, the framework can be adapted to evaluate novel paradigms, like graph neural networks, transfer learning, foundation models, and large language model (LLM)-based intrusion detection

systems.

Additionally, standardised cross-dataset evaluations and improved replicability of AI-driven cybersecurity research can be achieved by establishing a single repository for a standardised benchmark with a common feature representation across various intrusion datasets a promising avenue to explore.

The proposed benchmark framework overall is a foundation for a more systematic and objective-based approach to benchmark selection. The framework can be extended to provide a more complete decision support system for fair benchmarking, reproducible experiments and more powerful and generalizable IDSs, using added data sets, attack scenarios, and AI techniques.

In addition to the selection of benchmarks, the proposed framework will help to build more transparent, reproducible, and trustworthy intrusion detection systems, which will foster sustainable cybersecurity. Choosing benchmark datasets with analytical evidence and not appeal to scientific reproducibility and responsible AI deployment. The characteristics directly support SDG 9 (strengthening resilient digital infrastructure), SDG 16 (trustworthy digital governance and cyber resilience) and SDG 4 (high-quality cybersecurity education and research with the standardized selection of benchmark).

CONCLUSIONS

This study describes an extensive benchmarking framework for the comparative characterization and selection of intrusion detection data sets for use in Artificial Intelligence (AI)-based cybersecurity studies. In contrast to most studies that focus on developing or comparing intrusion detection algorithms, this paper examined the intrinsic properties of three popular benchmark datasets (CIC-IDS2017, UNSW-NB15, and IoT-23) to assist in selecting suitable datasets for AI model development. The proposed approach includes dataset characterization, attack diversity analysis, and evaluation of each feature's importance through Mutual Information and feature correlation analyses to provide a systematic understanding of the strengths and limitations of each benchmark dataset.

The comparative analyses showed that the three reference sample datasets are used for different types of research. CIC-IDS2017 offers the most comprehensive statistical feature representation and the most diverse set of attacks, making it suitable for explainable AI, feature engineering, and multi-class intrusion detection research. UNSW-NB15 offers a well-balanced mix of contemporary network data and complementary feature relationships to facilitate transfer learning and cross-dataset assessment. In comparison, IoT-23 features the highest data density and realistic IoT malware traffic, which are ideal for research in deep learning, large-scale intrusion detection, and Internet of Things security.

The main contribution of this work is a proposed Benchmark Selection Framework that turns comparative dataset analysis into a practical decision-making process for selecting benchmark datasets based on specific research objectives. The framework can be complemented by the Benchmark Scorecard and Benchmark Selection Matrix, enabling researchers to shift from subjective dataset selection to a transparent, evidence-based process that enhances experimental design, benchmarking consistency, and research reproducibility.

In conclusion, the selection of the benchmark dataset is a crucial step that significantly affects the development and evaluation of AI-based intrusion detection systems, as illustrated in this study. The proposed methodology and practical selection guidelines enhance the reliability, explanation, and generalizability of cybersecurity research. The proposed benchmark framework can serve as a model for future studies that expand to include more benchmark datasets, quantitative decision support/automated techniques, and new AI paradigms such as graph neural networks, transfer learning, foundation models, and large language model-assisted intrusion detection. From a broader perspective, the proposed benchmark framework goes beyond technical dataset comparisons by supporting sustainable AI adoption in cybersecurity. Through evidence-based benchmark selection, the framework promotes reproducible research, trustworthy AI, and resilient digital infrastructures aligned with SDG 9 (Industry, Innovation and Infrastructure), SDG 16 (Peace, Justice and Strong Institutions), and the strategic objectives of Oman Vision 2040 for digital transformation and innovation. These contributions position the framework as a practical foundation for future AI-enabled cybersecurity research supporting sustainable digital societies.

Author Contributions: Conceptualization, B.S.; Methodology, B.S.; Software, B.S.; Validation, B.S.; Formal Analysis, B.S.; Investigation, B.S.; Resources, B.S.; Data Curation, B.S.; Writing – Original Draft Preparation, B.S.; Writing – Review & Editing, B.S.; Visualization, B.S.; Supervision, B.S.; Project Administration, B.S.; Funding Acquisition, B.S. Authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Ethical review and approval were waived for this study because the research does not involve vulnerable groups or sensitive issues.

Funding: The University of Buraimi is funding this research work under the internally funded project "UoB\RIEU\IRG\2025-26\CoB\001", titled "AI-Driven Framework for Privacy, Safety, and Security Awareness (PSSA): User Maturity Modeling and Strategic Alignment with SDGs and Oman Vision 2040".

Acknowledgments: The authors recognize that support and funding from the University of Buraimi have made the successful implementation of this research feasible.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to restrictions.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process: During the preparation of this work, the author(s) used Grammarly for proofreading and spell checking since the Authors are not native speakers. All intellectual content, analysis, and interpretations were produced solely by the authors. After using this AI tool/service, the author(s) reviewed and edited the content as needed, taking full responsibility for the publication's content.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

Arul, M., Kishore Kumar, R., Santhosh, S., Boomika, M., Bhavana Shree, J., & Shree Durga, K. (2026). A multi-model deep learning approach for advanced cybersecurity threat detection. In Proceedings of the 2026 Second International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI). https://doi.org/10.1109/ICMSCI67830.2026.11469498

Abdo, S., Brits, J., & Akpinar, K. O. (2025). *A comparative study on IoT attack detection*. In *Proceedings of the 2025 2nd International Conference on Artificial Intelligence, Metaverse, and Cybersecurity (ICAMAC)*. https://doi.org/10.1109/ICAMAC67779.2025.11398518

Alhassan, S., Abdul-Salaam, G., Michael, A., Missah, Y., Ganaa, E., & Shirazu, A. S. (2024). CFS-AE: Correlation-based feature selection and autoencoder for improved intrusion detection system performance. *Journal of Internet Services and Information Security*, *14*(1), 104-120. https://doi.org/10.58346/jisis.2024.i1.007

AlMohamad, J. A. (2026). *Synergizing explainable AI and federated learning for proactive information security: A novel framework for zero-day threat detection*. *ECOSOCIAL Studies: Banking, Finance and Cybersecurity*, *1*(1), 1–13. https://doi.org/10.56334/ecosbankfincyber/8.1.1

Assudani, P., Kumar, N., Mohanambal, K., & Chitra, R. (2025). *Explainable artificial intelligence-driven intrusion detection system for enhancing reliability and interpretability in IoT-based network security solutions*. *Journal of Intelligent Systems and Internet of Things*, *17*(1), 219-238. https://doi.org/10.54216/jisiot.170116

Balega, M., Farag, W., Wu, X.-W., Ezekiel, S., & Good, Z. (2024). *Enhancing IoT security: Optimizing anomaly detection through machine learning*. Electronics, *13*(11), 2148. https://doi.org/10.3390/electronics13112148

Chinnasamy, R., & Subramanian, M. (2025). An explainable intrusion detection system using novel Indian millipede optimization and WGAN-GP with a dynamic attention-based ensemble model. *PeerJ Computer Science*, *11*, e3278.https://doi.org/10.7717/peerj-cs.3278

Chizari, M., Alam, A., Ali Mirza, Q. K., & Chizari, H. (2026). A Tri-Axis Systematic Literature Review of AI-Powered Cyber Defense: ATT&CK-Aligned Analysis of Cyberattacks, Machine Learning Methods, and Datasets. *Electronics*, *15*(13), 2804. https://doi.org/10.3390/electronics15132804

Darwish, R., & Roy, K. (2025). *Comparative analysis of federated learning, deep learning, and traditional machine learning techniques for IoT malware detection*. In *Proceedings of the International Conference on Applied Informatics and Communication (ICAIC)*. https://doi.org/10.1109/ICAIC63015.2025.10849203

Dash, N., Chakravarty, S., & Rath, A. (2024). *Deep learning model for elevating Internet of Things intrusion detection*. *International Journal of Electrical and Computer Engineering (IJECE), 14(5),* 5874-5883. https://doi.org/10.11591/ijece.v14i5.pp5874-5883

Dhote, S., & Agrawal, D. D. (2026). Machine Learning and Deep Learning-Based Intrusion Detection Systems: A Comprehensive Review of Datasets, Algorithms, Challenges, Explainability, and Future Research Directions. *Interdisciplinary Journal of AI, Machine Learning & Data Science*, *1*(3), e004. https://doi.org/10.66261/tg4asf94

Dubey, A., Pandey, V. K., Shukla, A., Sahu, A., & Prakash, S. (2025). *Deep learning-based intrusion detection framework for healthcare IoT networks*. In *Proceedings of the 2025 International Conference on Decision Aid Sciences and Applications (DASA)*. https://doi.org/10.1109/DASA68193.2025.11499071

García, P., Curtò, J., & Zarzà, I. D. (2025). *Foundation models for tabular intrusion detection: Evaluating TabPFN and LLM few-shot classification on IoT network security*. In *Proceedings of the 2025 3rd International Conference on Foundation and Large Language Models (FLLM)*. https://doi.org/10.1109/FLLM67465.2025.11391169

He, M., Wang, X., Wei, P., Yang, L., Teng, Y., & Lyu, R. (2024). Reinforcement learning meets network intrusion detection: A transferable and adaptable framework for anomaly behavior identification. *IEEE Transactions on Network and Service Management*, *21*(2), 2477-2492. https://doi.org/10.1109/TNSM.2024.3352586

Hejazi, S. M., Alshalabi, A. Y., Hatamleh, M., & Albaroudi, E. (2025). *A lightweight hybrid deep learning-based intrusion detection system for detecting botnet attacks in IoT networks*. *Journal of Scientific Research and Reports, 31*(11), 97-120. https://doi.org/10.9734/jsrr/2025/v31i113654

Hleha, K., & Hol, V. (2025). *XAI optimization for low-latency neural-based intrusion detection systems in network environments*. *Bulletin of V. N. Karazin Kharkiv National University, Series "Mathematical Modeling. Information Technology. Automated Control Systems"*, *66*, 19–36. https://doi.org/10.26565/2304-6201-2025-66-02

Hoa, N. T. (2025). *Intrusion detection in network systems using transformer-based approach*. *Vinh University Journal of Science ,54(4A)*, 59-72. https://doi.org/10.56824/vujs.2025a0117a

Huang, W., Tian, H., Wang, S., Zhang, C., & Zhang, X. (2024). Integration of simulated annealing into pigeon inspired optimizer algorithm for feature selection in network intrusion detection systems. *PeerJ Computer Science*, *10*, e2176.. https://doi.org/10.7717/peerj-cs.2176

Hussein, Z. N., Hammood, D., & Al-Abbasi, Z. (2025). *DeepCyber-IDS: A deep learning-based intrusion detection system*. In *Proceedings of the 2025 VI International Conference on Neural Networks and Neurotechnologies (NeuroNT)*. https://doi.org/10.1109/NeuroNT66873.2025.11049980

Kaliyaperumal, P., Periyasamy, S., Manikandan, T., Balusamy, B., & Benedetto, F. (2024). *A novel hybrid unsupervised learning approach for enhanced cybersecurity in the IoT*. *Future Internet*, *16*(7), 253. https://doi.org/10.3390/fi16070253

Kalyana, K. K. (2026). *AI-powered real-time CNN-LSTM intrusion detection: From streaming traffic to actionable alerts*. *International Journal of Scientific Research in Engineering & Technology*, *6(2)*, 123-128. https://doi.org/10.59256/ijsreat.20260602018

Kamal, H., & Mashaly, M. (2025). Enhanced Hybrid Deep Learning Models-Based Anomaly Detection Method for Two-Stage Binary and Multi-Class Classification of Attacks in Intrusion Detection Systems. *Algorithms*, *18*(2), 69. https://doi.org/10.3390/a18020069

Kodete, C. S., Raju, K. B., Karmakonda, K., Sikindar, S., Ramesh, J. V. N., & Tirumanadham, N. K. M. K. (2025). *Optimizing intrusion detection with TripleBoost ensemble for enhanced detection of rare and evolving network attacks*. *International Journal of Electrical and Electronic Engineering & Telecommunications*, *14*(3), 115-129. https://doi.org/10.18178/ijeetc.14.3.115-129

Kwubeghari, A., & Ezeji, N. G. (2025). Designing an explainable intrusion detection system (X-Ids) using machine learning: a framework for transparency and trust. *ABUAD Journal of Engineering Research and Development (AJERD)*, *8*(2), 319-328. https://doi.org/10.53982/ajerd.2025.0802.32-j

Li, L., Zhang, Y., Wang, J., & Ke, X. (2024). Deep learning-based network traffic anomaly detection: A study in IoT environments. *World Journal of Innovative Modern Technology*, *7*(6), 13-26. https://doi.org/10.53469/WJIMT.2024.07(06).03

Mada, Y. M., Bello-Salau, H., Yusuf, S. M., Ahmad, B., Adekale, A., & Dauda, A. (2024). *Deep learning-based network intrusion detection system using predator optimization algorithm for feature selection*. In *Proceedings of the 2024 IEEE 5th International Conference on Electro-Computing Technologies for Humanity (NIGERCON)*. https://doi.org/10.1109/NIGERCON62786.2024.10927361

Mittal, S., & Rajvanshi, P. (2025). *Towards a lightweight hybrid deep learning approach for malware detection enhancement in IoT-based systems*. In *Proceedings of the 2025 8th International Conference on Electronics, Materials Engineering and Nano-Technology (IEMENTech)*. https://doi.org/10.1109/IEMENTech65115.2025.10959623

Memmesheimer, P., Machmeier, S., & Heuveline, V. (2024). *Increasing detection rate for imbalanced malicious traffic using generative adversarial networks*. In *Proceedings of the European Interdisciplinary Cybersecurity Conference*. https://doi.org/10.1145/3655693.3655703

Mohamed, M. A., Emary, E., & Attalla, M. A. (2025). *Improved IoT anomaly detection through hybrid machine learning and deep learning approaches using the IoT-23 dataset*. In *Proceedings of the International Conference on Computing Advancements (ICCA)*. https://doi.org/10.1109/ICCA66035.2025.11430814

Prajwalasimha, S. N., Shelke, N., Saini, D. K. J. B., Pimpalkar, A., Tadkal, S., & Balla, R. (2025). Hybrid transformer-CNN neuro-symbolic explainable AI for cyber threat intelligence: Advancing transparency and adversarial robustness. In Proceedings of the 2025 3rd International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI). https://doi.org/10.1109/ICoICI65217.2025.11254796

Nugroho, K. A., Hariguna, T., & Barkah, A. S. (2025). Optimizing Early Network Intrusion Detection: A Comparison of LSTM and LinearSVC with SMOTE on Imbalanced Data. *Jurnal Teknik Informatika (Jutif)*, *6*(6), 5349-5370. https://doi.org/10.52436/1.jutif.2025.6.6.4672

Nzuva, S. M., Nder, L., & Mwalili, T. (2024). *A novel bagging-XGBoost ensemble model for attaining high accuracy and computational efficiency in network intrusion detection*. *E3S Web of Conferences* . https://doi.org/10.1051/e3sconf/202450101007

Oziegbe, T. E., Edje, A. E., & Akazue, M. (2026). DEEP LEARNING–BASED INTRUSION DETECTION IN VEHICULAR NETWORKS: A REVIEW OF GATED RECURRENT UNIT APPROACHES. *Science World Journal*, *21*(1), 264-272. https://doi.org/10.4314/swj.v21i1.37

Pinto, D., Vitorino, J., Maia, E., Amorim, I., & Praça, I. (2024). *Flow exporter impact on intelligent intrusion detection systems*. In *Proceedings of the International Conference on Information Systems Security and Privacy*. https://doi.org/10.48550/arXiv.2412.14021

Rajasa, M. C., Rahma, F., Rachmadi, R. F., Pratomo, B., & Purnomo, M. (2023). *A review of imbalanced datasets and resampling techniques in network intrusion detection systems*. In *Proceedings of the 2023 8th International Conference on Information Technology and Digital Applications (ICITDA)*. https://doi.org/10.1109/ICITDA60835.2023.10427217

Ren, Y. (2026). Data science and machine learning for cyber intrusion detection: a systematic review. *Mach Learn Res*, *11*(1), 8-21. https://doi.org/10.11648/j.mlr.20261101.12

Rai, I. N. A. S., Heryadi, D., Yani, Y. M., & Nashir, A. K. (2025). RETHINKING EUROPEAN CYBERSECURITY INTEGRATION THROUGH LIBERAL INTERGOVERNMENTAL POLITICS IN GENERAL DATA PROTECTION REGULATION ENFORCEMENT STUDY. *Bangladesh Journal of Multidisciplinary Scientific Research*, *11*(1), 23-34. https://doi.org/10.46281/bjmsr.v11i1.2638

Sagaran, E., Spier, J., & Hasan, R. (2025). *SpiCAE: Spiking contrastive learning and autoencoder for network intrusion detection*. In *Proceedings of the IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. https://doi.org/10.1109/IEMCON67450.2025.11381151

Sah, G., Singh, S., & Banerjee, S. (2024). Intrusion detection system using classification algorithms with feature selection mechanism over real-time data traffic. *China Communications*, *21*(9), 292-320. https://doi.org/10.23919/JCC.fa.2022-0076.202409

Saini, N., Bhat Kasaragod, V., Prakasha, K., & Das, A. K. (2023). A hybrid ensemble machine learning model for detecting APT attacks based on network behavior anomaly detection. *Concurrency and Computation: Practice and Experience*, *35*(28), e7865. https://doi.org/10.1002/cpe.7865

Sharma, V., & Kumar, M. (2025). Improving intrusion detection with hybrid deep learning models: A study on CIC-IDS2017, UNSW-NB15, and KDD CUP 99. *Journal of Information Systems Engineering and Management*, *10*(11S), 633-650. https://doi.org/10.52783/jisem.v10i11s.1665

Sravani, A., Sri, M. R., & Sanjana, C. (2026). *Evaluating single-model and ensemble-based intrusion detection on the CIC-IDS2017 dataset*. *International Scientific Journal of Engineering & Management*, *5(3)*, 1-6. https://doi.org/10.55041/isjem05901

Sreerenjith, P., & Benitta, D. (2026). *Cloud-native intrusion detection and DDoS attack mitigation using federated deep learning on AWS Free Tier*. In *Proceedings of the 2026 3rd International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)*. https://doi.org/10.1109/RMKMATE69073.2026.11519014

Tamuka, N., Mathonsi, T., Olwal, T., Maswikaneng, S., Muchenje, T., & Tshilongamulenzhe, T. (2026). *Intrusion detection in fog computing: A systematic review of security advances and challenges*. *Computers*, *15*(3), 169. https://doi.org/10.3390/computers15030169

Thomas, R., Chaturvedi, A., & Goswami, D. N. (2025). *The role of AI-enabled optimization in network traffic management*. *International Journal for Sciences and Technology* ,*16*(2), 1-15. https://doi.org/10.71097/ijsat.v16.i2.3428

Tian, W., Shen, Y., Guo, N., Yuan, J., & Yang, Y. (2024). VAE-WACGAN: An Improved Data Augmentation Method Based on VAEGAN for Intrusion Detection. *Sensors*, *24*(18), 6035. https://doi.org/10.3390/s24186035

Xie, H., Shao, Y., Li, Z., Alomari, Z., & Makanju, A. (2025). *Optimization of class imbalance techniques in machine learning models for network intrusion detection*. In *Proceedings of the International Conference on Cryptography, Security and Privacy*. https://doi.org/10.1109/CSP66295.2025.00025

Xu, L., Wang, L., & Jiang, Y. (2026). Optimized Deep Learning-Based Intrusion Detection System Using SMOTE and Genetic Algorithms. *International Journal of Pattern Recognition and Artificial Intelligence*, *40*(3), 2552032. https://doi.org/10.1142/S0218001425520329

Zhang, C., Li, J., Wang, N., & Zhang, D. (2025). *Research on intrusion detection methods based on transformer and CNN-BiLSTM in the Internet of Things*. *Sensors*, 25(9), 2725. https://doi.org/10.3390/s25092725

Zhang, X., & Tuo, J. (2025). *Research on network intrusion detection of automation systems based on machine learning*. In *Proceedings of the 2025 10th International Conference on Electronic Technology and Information Science (ICETIS)*. https://doi.org/10.1109/ICETIS66286.2025.11144041

Publisher's Note: CRIBFB stays neutral about jurisdictional claims in published maps and institutional affiliations.