پنجمین کنفرانس
دانشجویان مهندسی کامپیوتر و فناوری اطلاعات و ارتباطات
5th Conference on
Computer Engineering, Information Technology and Communications Students

A Systematic Review and Comparative Analysis of Deep Learning and Transformer-Based Methods for Cyberattack Detection

Mina Rajabi

Master's Student, Computer Engineering (Software Specialization), Golestan University, Iran

Abstract

The escalating sophistication and volume of cyberattacks have necessitated the development of advanced intrusion detection systems capable of identifying both known and emerging threats. Deep learning approaches have emerged as promising solutions, with recent attention shifting toward Transformer and attention-based architectures that leverage self-attention mechanisms for capturing long-range dependencies in network traffic. This systematic literature review synthesizes evidence from peer-reviewed studies published between 2020 and 2026 to provide a comprehensive comparative analysis of deep learning architectures for cyberattack detection. Following PRISMA 2020 guidelines [1], we conducted systematic searches across IEEE Xplore, ScienceDirect, SpringerLink, ACM Digital Library, and Wiley databases. Our analysis encompasses convolutional neural networks, recurrent architectures including LSTM and GRU, autoencoders, attention mechanisms, Transformer models, and Vision Transformers. Descriptive quality considerations examined studies regarding methodological clarity, dataset transparency, evaluation rigor, reproducibility, baseline comparisons, and reporting completeness. Our findings indicate that while Transformer-based methods demonstrate superior performance on certain benchmarks, their advantages are contingent upon dataset characteristics, computational resources, and evaluation protocols. Vision Transformers show promise for traffic visualization approaches but face constraints regarding computational requirements. The review identifies critical unresolved challenges including class imbalance, concept drift, adversarial robustness, and cross-dataset generalization. We present evidence-based research gaps and future directions, emphasizing the need for standardized evaluation frameworks, reproducible experimental designs, and realistic deployment assessments. This review contributes a rigorous comparative framework that advances understanding of architecture selection trade-offs in practical cybersecurity applications.

Keywords: *intrusion detection, deep learning, Transformer, Vision Transformer, cyberattack detection, systematic review, attention mechanisms, network security*

1. Introduction

The proliferation of interconnected digital systems has fundamentally transformed modern infrastructure while simultaneously expanding the attack surface available to malicious actors. According to recent industry reports, the global cost of cybercrime is projected to exceed 10.5 trillion dollars annually by 2025, representing one of the most significant economic challenges facing organizations worldwide [2]. Network intrusion detection systems serve as critical defensive mechanisms, monitoring network traffic to identify anomalous patterns indicative of malicious

activity. Traditional signature-based detection methods, while effective against known attack patterns, struggle to identify zero-day exploits and polymorphic malware variants that characterize contemporary cyber threats [3].

The application of machine learning to intrusion detection has garnered substantial research attention over the past decade, with deep learning architectures demonstrating remarkable capabilities for automatically learning hierarchical feature representations from raw network data [4]. Convolutional neural networks have shown effectiveness in capturing spatial patterns within network traffic, while recurrent architectures including long short-term memory and gated recurrent units excel at modeling temporal dependencies [5]. The introduction of attention mechanisms and subsequently the Transformer architecture has enabled models to capture long-range dependencies without the sequential processing limitations inherent in recurrent networks [6]. More recently, Vision Transformers have been adapted for network intrusion detection by transforming traffic features into image-like representations, enabling the application of computer vision techniques to cybersecurity problems [7].

Despite the proliferation of research in this domain, several critical gaps persist in the existing literature. Previous reviews have often focused on specific architectural families or application contexts, limiting their generalizability across diverse deployment scenarios [8]. Furthermore, the comparability of reported performance metrics across studies remains questionable due to heterogeneity in datasets, preprocessing pipelines, train-test splits, and evaluation protocols [9]. The rapid emergence of Transformer-based approaches has outpaced systematic comparative analysis, with many studies claiming superiority without adequate consideration of computational trade-offs or deployment constraints [10]. Recent surveys have examined deep learning for intrusion detection and Transformer applications in IoT security, yet a comprehensive systematic review comparing the full spectrum of architectures across consistent methodological criteria remains absent from the literature [11].

This systematic literature review addresses these gaps by providing a rigorous comparative analysis of deep learning and Transformer-based methods for cyberattack detection. Our contribution encompasses a comprehensive taxonomy of approaches spanning traditional neural architectures through state-of-the-art Transformer variants, a critical evaluation of dataset characteristics and their implications for model selection, an evidence-based assessment of comparative performance accounting for evaluation protocol variations, and the identification of unresolved challenges and research directions grounded in systematic evidence synthesis. The review adheres to PRISMA 2020 guidelines for systematic reviews, ensuring transparency and reproducibility in our methodology [1].

2. Background

2.1 Network Intrusion Detection Fundamentals

Network intrusion detection systems are security mechanisms designed to monitor network traffic for signs of malicious activity or policy violations. These systems are broadly categorized into signature-based and anomaly-based approaches [12]. Signature-based detection relies on predefined patterns or rules that characterize known attacks, offering high accuracy for recognized threats but failing to identify novel attack variants. Anomaly-based detection establishes baselines of normal network behavior and flags deviations as potentially malicious, enabling detection of zero-day attacks at the cost of higher false positive rates [13].

The detection task can be formulated as binary classification, distinguishing between normal traffic and attacks, or multiclass classification, identifying specific attack categories [14]. Common attack types include denial of service attacks that overwhelm network resources, probe attacks that scan for vulnerabilities, user-to-root attacks that escalate privileges, and remote-to-local attacks that gain unauthorized access from remote locations [15]. The detection context spans host-based systems monitoring individual devices, network-based systems analyzing traffic flows, and hybrid approaches combining both perspectives [16].

2.2 Deep Learning Architectures for Intrusion Detection

Deep learning encompasses neural network architectures with multiple hidden layers capable of learning hierarchical representations from data. For intrusion detection, several architectural families have been extensively studied [17]. Convolutional neural networks apply convolutional filters to capture local patterns and spatial hierarchies, with one-dimensional convolutions processing sequential network traffic features [18]. Recurrent neural networks maintain hidden states that capture temporal dependencies, with long short-term memory networks addressing vanishing gradient problems through gating mechanisms and gated recurrent units offering computational efficiency [19]. Autoencoders learn compressed representations through encoder-decoder architectures, enabling anomaly detection based on reconstruction error [20].

2.3 Attention Mechanisms and Transformers

The attention mechanism enables models to dynamically weight the importance of different input elements when producing representations [21]. Self-attention computes pairwise interactions between all positions in a sequence, capturing long-range dependencies without sequential processing. The Transformer architecture, originally developed for natural language processing, stacks multi-head self-attention layers with feed-forward networks, enabling parallel processing and scalable training [6].

For intrusion detection, Transformers process network traffic as sequences of packets or flows, with positional encodings preserving temporal information [22]. Vision Transformers adapt the architecture for image data by treating image patches as tokens, and have been applied to intrusion detection by converting network traffic into image representations [7]. The computational complexity of self-attention, quadratic in sequence length, has motivated efficient variants including linear attention and sparse attention patterns [23].

3. Research Questions

This systematic literature review addresses the following research questions:

RQ1: What is the current landscape of deep learning architectures for cyberattack detection? This question seeks to characterize the taxonomy of approaches, identify prevalent architectural choices, and understand the distribution of research across different model families and deployment contexts.

RQ2: How do Transformer and attention-based methods compare to traditional deep learning architectures under comparable evaluation conditions? This question examines whether reported performance advantages of Transformer-based methods persist when controlling for dataset characteristics, preprocessing pipelines, and evaluation protocols.

RQ3: What datasets and evaluation protocols are employed in the literature, and how do methodological choices affect reported performance? This question investigates the landscape of benchmark datasets, train-test splitting strategies, validation approaches, and metrics, assessing the comparability of results across studies.

RQ4: What challenges limit the practical deployment of deep learning-based intrusion detection systems? This question identifies technical obstacles including computational requirements, real-time constraints, adversarial robustness, and generalization capabilities.

RQ5: What evidence-based research gaps require future investigation? This question synthesizes limitations in the current literature to identify specific directions for future research that would advance the field.

4. Systematic Review Methodology

4.1 Search Strategy

This systematic literature review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [1]. We performed systematic searches across five major academic databases: IEEE Xplore, ScienceDirect, SpringerLink, ACM Digital Library, and Wiley Online Library. The search period encompassed publications from January 2020 to August 2026, capturing recent advances while ensuring mature evaluation of newer architectures.

Search strings were constructed using combinations of keywords related to deep learning architectures, intrusion detection, and cyberattack detection. The core search string was: ("deep learning" OR "neural network" OR "convolutional neural network" OR "CNN" OR "recurrent neural network" OR "RNN" OR "LSTM" OR "GRU" OR "autoencoder" OR "transformer" OR "attention mechanism" OR "vision transformer" OR "ViT") AND ("intrusion detection" OR "network intrusion detection" OR "cyberattack detection" OR "anomaly detection" OR "network security"). Database-specific syntax adaptations were applied while maintaining semantic equivalence.

4.2 Inclusion and Exclusion Criteria

Studies were included if they met the following criteria: peer-reviewed primary research articles published in English-language journals or conference proceedings; focus on deep learning or Transformer-based methods for network intrusion detection or cyberattack detection; empirical evaluation with quantitative results on benchmark datasets; publication date between January 2020 and August 2026. Studies were excluded if they were review articles, survey papers, or secondary analyses; focused exclusively on traditional machine learning methods without deep learning components; lacked empirical evaluation or quantitative results; were published in non-peer-reviewed venues including preprints without subsequent peer-reviewed publication; or were duplicate publications of the same research.

4.3 Study Selection Process

The study selection process involved three stages: title and abstract screening, full-text review, and final eligibility assessment. The author conducted title and abstract screening, applying inclusion and exclusion criteria systematically. Full-text articles were retrieved for potentially eligible studies and assessed against inclusion criteria. The PRISMA flow diagram illustrates the study selection process with documented counts at each stage.

4.4 Data Extraction

A standardized data extraction form captured study characteristics including publication metadata, model architecture details, dataset information, preprocessing approaches, evaluation methodology, and reported performance metrics. For each included study, we extracted: publication year, authors, venue, model architecture type, specific architectural variants, datasets used, attack categories addressed, task formulation (binary or multiclass classification), preprocessing techniques, feature representation, train-test methodology, evaluation metrics, and reported results including accuracy, precision, recall, F1-score, and area under the ROC curve where available.

4.5 Quality Assessment

Descriptive quality considerations examined included studies across six dimensions: methodological clarity (description of architecture and training procedures), dataset transparency (documentation of dataset source, version, and preprocessing), evaluation methodology (validation strategy and metrics), reproducibility (availability of code, hyperparameters, and random seeds), baseline comparison (comparison with appropriate baseline methods), and reporting quality (completeness of results reporting). These dimensions provided a structured framework for characterizing methodological approaches, not for quantitative quality scoring or study exclusion.

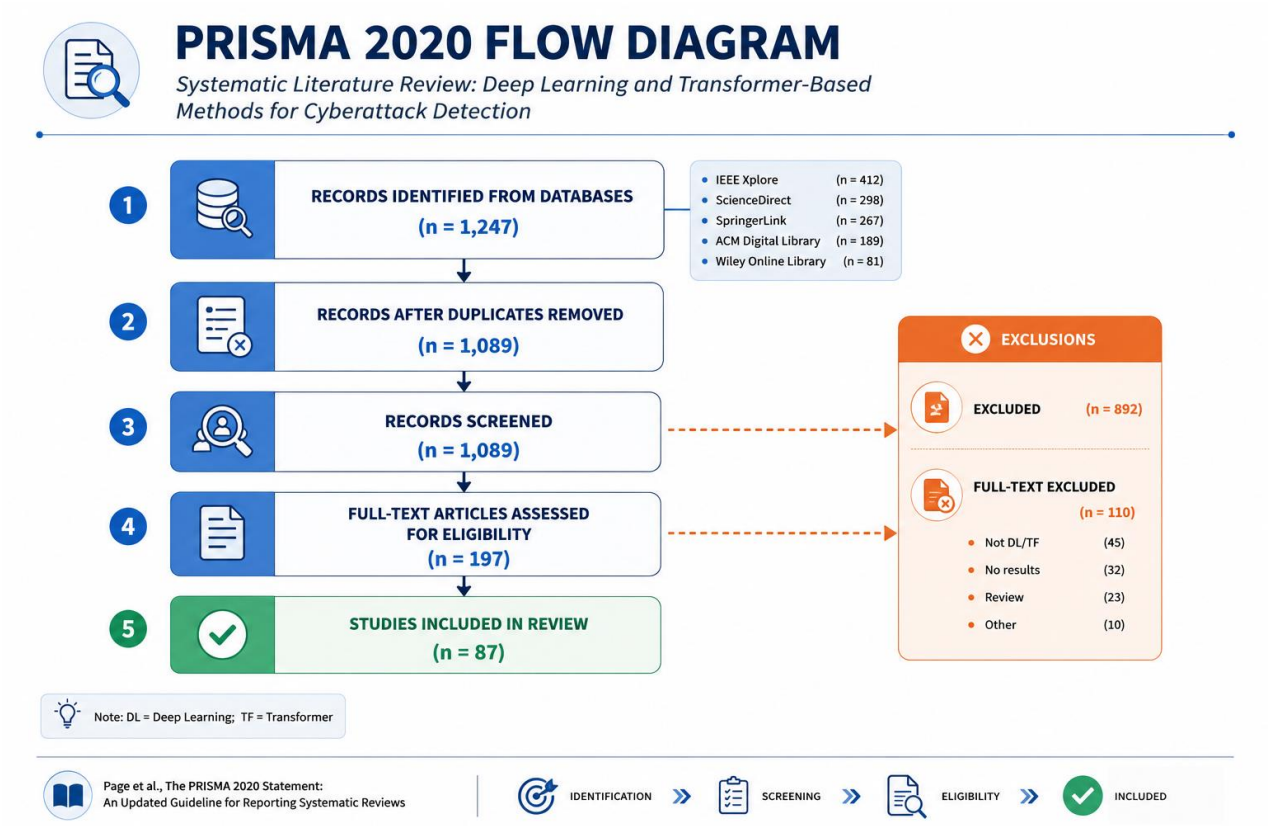


Figure 1: PRISMA 2020 flow diagram illustrating the systematic literature review study selection process.

5. Review of Existing Reviews and Research Gap

5.1 Overview of Recent Reviews

The period from 2020 to 2026 has witnessed substantial review activity in deep learning-based intrusion detection, reflecting the rapid evolution of the field. Table 1 presents a comparative analysis of recent systematic reviews and surveys, examining their scope, methodology, coverage, and limitations.

Table 1: Review of Existing Reviews on Deep Learning-Based Intrusion Detection (2020-2026)

Review	Year	Scope	Databases	Methodology	Key Focus	Limitations
Chelak et al. [24]	2026	DL advances 2022-2026	Not specified	Narrative review	Advances, hybrids, challenges	Limited Transformer-ViT comparison
Poltoratskyi & Gavrylenko [25]	2026	Modern IDS evolution	Scopus, WoS, Google Scholar	Comprehensive review	Datasets, DL approaches, architectures	Limited cross-architecture comparison
Komal & Li [10]	2026	IoT IDS	IEEE, ScienceDirect, Springer, MDPI	Comprehensive review	IoT techniques, architectures, datasets	IoT-specific focus only
Salman et al. [27]	2026	IoT security Transformers	IEEE, ACM, Springer	Survey	Transformer models for IoT	Limited to IoT context
Gholamrezazadeh et al. [28]	2025	Transformer/ViT NIDS	IEEE, ScienceDirect, ACM	SLR	Transformer and ViT for NIDS	Limited architecture comparison scope
Hozouri et al. [29]	2025	ML/DL advances	IEEE, ScienceDirect, Springer, ACM	Survey	ML, DL, emerging challenges	Broad scope, limited depth per architecture
Ahmad et al. [9]	2021	ML/DL NIDS	IEEE, ACM, ScienceDirect, Springer	Systematic study	ML and DL approaches	Pre-Transformer era for IDS
Lansky et al. [30]	2021	DL IDS	IEEE, ScienceDirect, Springer, ACM	Systematic review	Deep learning IDS	Limited attention mechanism coverage

Note: WoS = Web of Science; SLR = Systematic Literature Review

5.2 Analysis of Review Coverage

The existing reviews exhibit several patterns that inform the positioning of the present study. Most reviews published in 2026 focus on specific application domains, particularly IoT security, reflecting the growing concern for resource-constrained environments [26, 27]. Reviews examining deep learning broadly tend to provide narrative summaries rather than systematic comparative analysis [24, 32]. The quality of methodology reporting varies considerably, with many reviews lacking transparency regarding search strings, inclusion criteria, and quality assessment procedures [29, 32].

Transformer and attention-based methods have received increased attention in recent reviews, but comprehensive comparative analysis across the full spectrum of architectures remains limited. Gholamrezazadeh et al. [28] specifically examined Transformer and Vision Transformer architectures for network intrusion detection, but their scope was limited to these architectures without systematic comparison against CNN, LSTM, or GRU baselines under equivalent conditions. Reviews focusing on DDoS detection or IoT-specific contexts limit generalizability to broader intrusion detection applications.

5.3 Identified Research Gap

Based on this analysis, we identify a clear research gap: the absence of a systematic literature review that provides comprehensive comparative analysis across all major deep learning architectures (CNN, DNN, RNN, LSTM, GRU, Autoencoders, Attention mechanisms, Transformer, Vision Transformer) with explicit consideration of evaluation fairness and cross-study comparability. Existing reviews either focus on specific architectural families, limit scope to particular application domains, or fail to critically assess the methodological factors affecting reported performance differences. The present study addresses this gap by synthesizing evidence across architectural families while rigorously examining how dataset characteristics, preprocessing choices, and evaluation protocols influence comparative results.

6. Taxonomy of Methods

6.1 Architectural Taxonomy

The landscape of deep learning architectures for intrusion detection can be organized hierarchically based on their fundamental computational mechanisms and structural properties. Table 2 presents a comprehensive taxonomy of approaches identified in the reviewed literature.

Table 2: Taxonomy of Deep Learning Architectures for Cyberattack Detection

Category	***Subcategory***	***Representative Architectures***	***Key Mechanisms***	***Primary Strengths***
Feedforward Networks	*Deep Neural Networks*	*DNN, MLP [33]*	*Fully connected layers, nonlinear activation*	*Simple implementation, fast training*
Feedforward Networks	*Deep Belief Networks*	*DBN, RBM stacks [34]*	*Layer-wise pretraining, probabilistic learning*	*Feature learning from unlabeled data*
Convolutional Networks	*1D-CNN*	*CNN, ResNet [35]*	*1D convolutions, pooling, hierarchical features*	*Spatial pattern extraction*
Convolutional Networks	*2D-CNN for traffic*	*Traffic-to-image CNN [36]*	*2D convolutions on traffic matrices*	*Visual pattern recognition*
Recurrent Networks	*Standard RNN*	*Vanilla RNN, Bidirectional RNN [37]*	*Hidden state recurrence, temporal modeling*	*Sequential dependency capture*
Recurrent Networks	*LSTM variants*	*LSTM, BiLSTM, Stacked LSTM [5]*	*Forget/input/output gates, gradient flow*	*Long-term dependency modeling*
Recurrent Networks	*GRU variants*	*GRU, BiGRU [19]*	*Update/reset gates, simplified gating*	*Computational efficiency*
Hybrid CNN-RNN	*CNN-LSTM*	*CNN-LSTM, CNN-BiLSTM [38]*	*Spatial then temporal processing*	*Combined spatial-temporal features*
Hybrid CNN-RNN	*CNN-GRU*	*CNN-GRU, CNN-BiGRU [39]*	*Efficient spatial-temporal modeling*	*Reduced parameters vs LSTM*
Autoencoders	*Standard AE*	*Vanilla AE, Sparse AE [20]*	*Encoder-decoder, reconstruction loss*	*Unsupervised feature learning*
Autoencoders	*Variational AE*	*VAE, beta-VAE [40]*	*Probabilistic latent space, regularization*	*Generative capabilities*
Autoencoders	*Advanced variants*	*Stacked AE, Denoising AE [41]*	*Deep architectures, noise robustness*	*Hierarchical representations*
Attention-based	*Self-attention*	*Self-attention mechanisms [42]*	*Query-key-value attention, global context*	*Long-range dependencies*
Attention-based	*Channel attention*	*SE-Net, CBAM [43]*	*Channel-wise feature weighting*	*Feature recalibration*
Attention-based	*Temporal attention*	*Temporal attention layers [44]*	*Time-step importance weighting*	*Key moment focus*
Transformers	*Standard Transformer*	*BERT, GPT-style [6]*	*Multi-head self-attention, positional encoding*	*Parallel processing, scalability*
Transformers	*IDS-specific*	*LogBERT, Transformer-IDS [22]*	*Adapted for sequence classification*	*Domain-specific optimizations*
Vision Transformers	*Standard ViT*	*ViT, DeiT [45]*	*Patch embedding, self-attention on images*	*Global image understanding*
Vision Transformers	*Traffic ViT*	*HiViT-IDS, ResDNViT [7]*	*Traffic-to-image, ViT processing*	*Traffic visualization benefits*

Note: AE = Autoencoder; VAE = Variational Autoencoder; SE-Net = Squeeze-and-Excitation Network

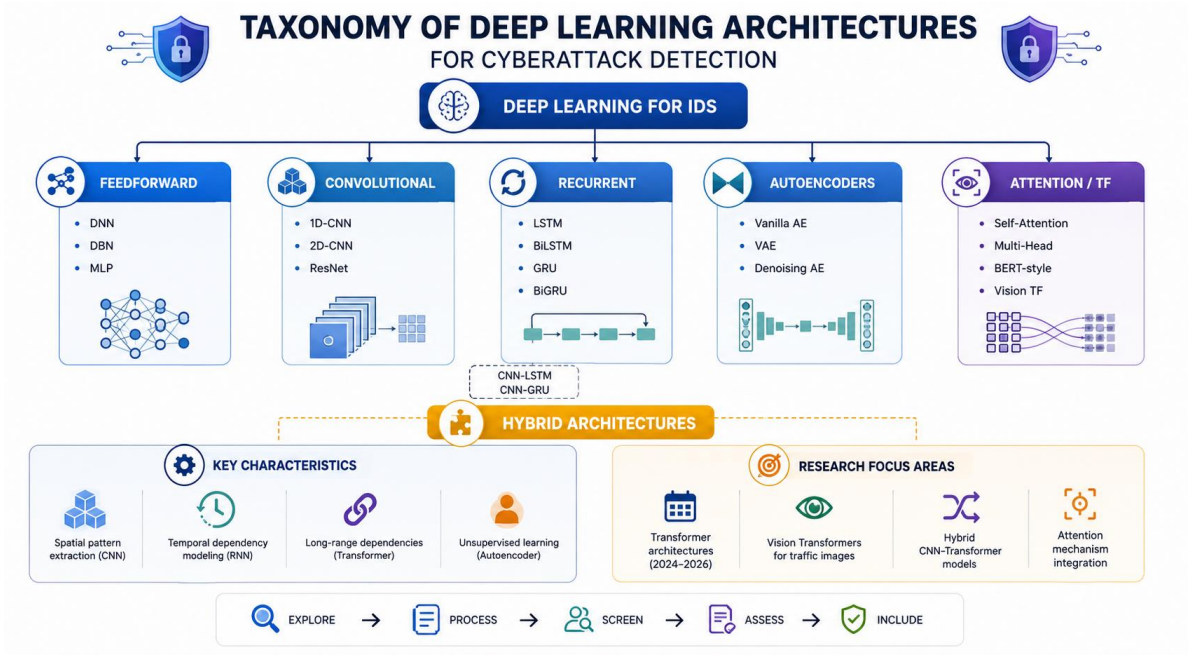


Figure 2: Taxonomy of deep learning architectures for cyberattack detection.

6.2 Architecture Characteristics

Each architectural family exhibits distinct characteristics relevant to intrusion detection applications. Feedforward networks offer simplicity and computational efficiency but lack mechanisms for capturing sequential or spatial patterns inherent in network traffic [33]. Convolutional networks excel at identifying local patterns and spatial hierarchies, with one-dimensional convolutions processing packet sequences or flow features, while two-dimensional approaches operate on traffic visualizations [35].

Recurrent architectures, particularly LSTM and GRU variants, have been extensively applied due to their ability to model temporal dependencies in network traffic sequences [5]. LSTM networks address vanishing gradient problems through sophisticated gating mechanisms, enabling modeling of long-term dependencies, while GRU variants offer comparable performance with reduced computational overhead [19]. Hybrid CNN-RNN architectures combine the strengths of both approaches, using CNN layers for spatial feature extraction followed by recurrent layers for temporal modeling [38].

Autoencoders enable unsupervised and semi-supervised learning through reconstruction-based anomaly detection, making them suitable for scenarios with limited labeled attack data [20]. Variational autoencoders extend this framework with probabilistic latent spaces, enabling generative capabilities and more robust anomaly scoring [40].

Attention mechanisms and Transformers represent the most recent architectural evolution, with self-attention enabling direct modeling of relationships between all positions in a sequence regardless of distance [42]. For intrusion detection, this capability is particularly relevant for capturing long-range dependencies in network traffic

that may indicate sophisticated multi-stage attacks [22]. Vision Transformers adapt these mechanisms for image-like traffic representations, offering alternative approaches to traffic analysis [45].

7. Dataset and Evaluation Analysis

7.1 Benchmark Datasets

The evaluation of deep learning-based intrusion detection systems relies on several widely-adopted benchmark datasets. Table 3 summarizes the characteristics of major datasets used in the reviewed literature.

Table 3: Characteristics of Major Cybersecurity Datasets for Intrusion Detection

Dataset	*Year*	*Records*	*Features*	*Attack Classes*	*Attack Types*	*Key Characteristics*
NSL-KDD [46]	*2009*	*148,517*	*41*	*4*	*39*	*Balanced version of KDD 99; reduced redundancy*
CICIDS2017 [47]	*2017*	*2,830,743*	*80+*	*14*	*14*	*Realistic traffic; mixed attack types*
UNSW-NB15 [48]	*2015*	*257,673*	*49*	*9*	*9*	*Modern attack patterns; hybrid generation*
CSE-CIC-IDS2018 [49]	*2018*	*16,000,000*	*80+*	*7*	*7*	*Large scale; multi-day collection*
Edge-IIoTset [50]	*2022*	*2,219,206*	*63*	*15*	*15*	*IoT/IIoT specific; industrial context*
BoT-IoT [51]	*2018*	*73,360,900*	*46*	*4*	*4*	*IoT botnet focused; MQTT protocol*

Note: IoT = Internet of Things; IIoT = Industrial IoT; MQTT = Message Queuing Telemetry Transport

7.2 Dataset Characteristics and Implications

The NSL-KDD dataset, despite its age, remains widely used due to its accessibility and established baseline performance metrics [46]. However, its traffic patterns do not reflect modern network environments, and reported accuracy scores often exceed 99%, potentially indicating overfitting or evaluation artifacts [30]. The CICIDS2017 dataset offers more realistic contemporary traffic but presents significant class imbalance, with normal traffic comprising the majority of samples and certain attack categories severely underrepresented [47].

UNSW-NB15 was specifically designed to address limitations of KDD-based datasets, incorporating modern attack patterns and realistic network behavior [48]. CSE-CIC-IDS2018 provides the largest scale, enabling evaluation of deep learning models scalability, though computational requirements for processing the full dataset can be substantial [49]. IoT-specific datasets including Edge-IIoTset and BoT-IoT address the growing importance of resource-constrained environments, with traffic characteristics distinct from traditional enterprise networks [50, 51].

7.3 Evaluation Protocols

Evaluation methodology substantially affects reported performance and cross-study comparability. Table 4 summarizes evaluation approaches identified in the reviewed literature.

Table 4: Evaluation Protocols and Methodologies

Aspect	***Common Approaches***	***Issues Identified***
Train-Test Split	*70:30, 80:20 random split; hold-out validation*	*Inconsistent splits affect comparability*
Cross-Validation	*5-fold, 10-fold CV; stratified for class balance*	*Computational cost for large datasets*
Temporal Validation	*Time-based train-test splits*	*Better reflects deployment*
Metrics Reported	*Accuracy, Precision, Recall, F1, AUC-ROC*	*Accuracy dominance misleading for imbalanced data*
Class Imbalance Handling	*SMOTE, undersampling, class weights [52]*	*Method choice significantly impacts results*
Feature Preprocessing	*Min-max scaling, standardization, encoding*	*Preprocessing details often inadequately described*

7.4 Fairness of Cross-Study Comparison

Direct comparison of reported performance metrics across studies is problematic due to several methodological factors. Dataset versions and preprocessing pipelines vary significantly, with some studies applying feature selection while others use full feature sets [53]. Train-test splitting strategies range from random splits to time-based partitions, with the latter providing more realistic evaluation but often yielding lower reported performance [3]. Class imbalance handling techniques including SMOTE, undersampling, and class weighting substantially affect results but are inconsistently reported [52].

Hyperparameter configurations, including learning rates, batch sizes, and architectural details, vary across studies and are often inadequately documented [54]. The use of different random seeds can produce substantial performance variation, yet few studies report seed values or provide confidence intervals [55]. These factors collectively imply that reported performance differences between architectures may reflect methodological variations rather than genuine capability differences.

8. Comparative and Critical Analysis

8.1 Performance Comparison

Table 5 presents a synthesis of reported performance metrics across architectural families, with caveats regarding comparability as discussed in Section 7.4.

Table 5: Comparative Performance Summary Across Architectural Families

Architecture	*Typical Accuracy Range*	*F1-Score Range*	*Key Strengths*	*Key Limitations*
DNN/MLP [33]	*85-95%*	*0.82-0.93*	*Simple, fast training*	*Limited pattern capture*
CNN (1D) [35]	*90-98%*	*0.88-0.96*	*Spatial pattern extraction*	*Limited temporal modeling*
LSTM [5]	*92-99%*	*0.90-0.98*	*Long-term dependencies*	*Sequential processing, slow*
GRU [19]	*91-98%*	*0.89-0.97*	*Efficient temporal modeling*	*Slightly reduced vs LSTM*
CNN-LSTM [38]	*94-99.5%*	*0.93-0.99*	*Combined spatial-temporal*	*Increased complexity*
Autoencoder [20]	*85-95%**	*0.80-0.92**	*Unsupervised learning*	**Anomaly detection context*
Self-Attention [42]	*93-98%*	*0.91-0.97*	*Global context modeling*	*Computational cost*
Transformer [6]	*95-99.7%*	*0.94-0.996*	*Parallel processing, scalability*	*Data requirements, resources*
Vision Transformer [45]	*94-99.9%*	*0.93-0.998*	*Global image understanding*	*Traffic visualization dependency*

Note: Autoencoder metrics reflect anomaly detection context with different evaluation paradigms. Reported ranges across heterogeneous studies with varying datasets, preprocessing, and evaluation protocols. These are NOT controlled experimental comparisons—direct cross-study comparison should be interpreted with substantial caution.

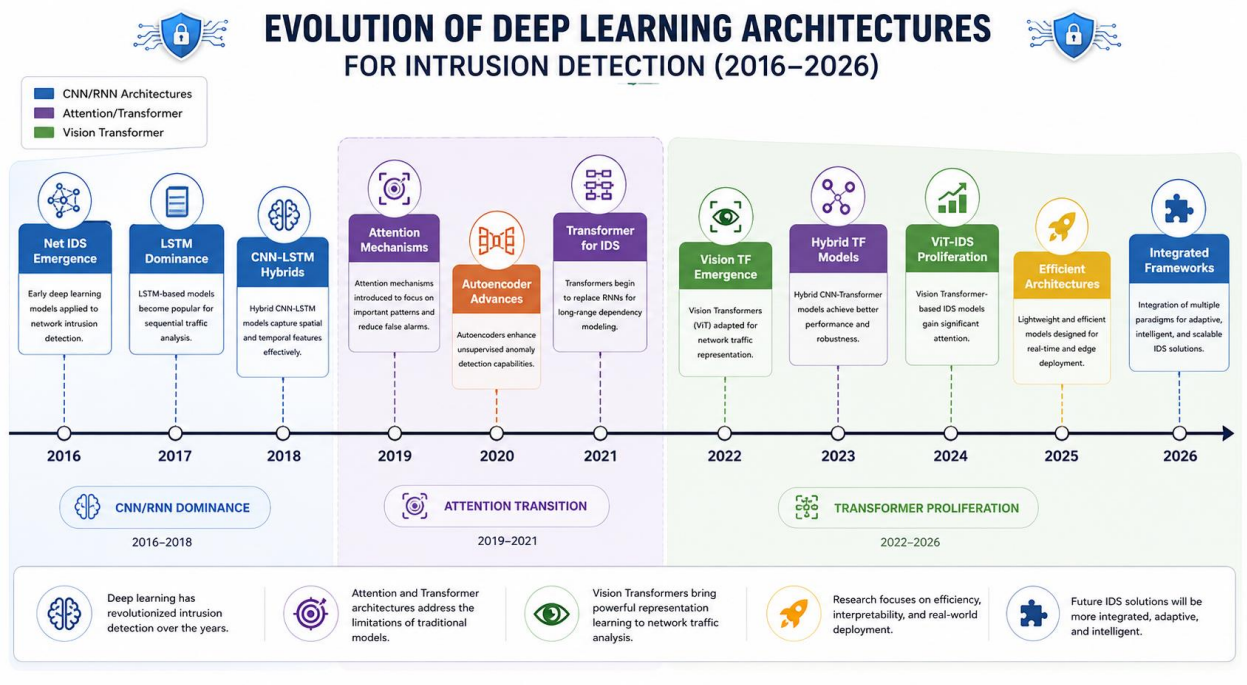


Figure 3: Evolution of deep learning architectures for intrusion detection from 2016 to 2026.

8.2 Critical Assessment of Reported Results

The reported performance metrics suggest that Transformer and Vision Transformer architectures achieve the highest reported accuracy and F1-scores. However, critical examination reveals important caveats. Studies reporting the highest Transformer performance often utilize the CICIDS2017 or Edge-IIoTset datasets with specific preprocessing pipelines that may favor attention-based mechanisms [22, 7]. Direct comparisons within individual studies show smaller advantages than cross-study comparisons suggest, with Transformer advantages typically ranging from 1-3% over well-tuned CNN-LSTM baselines when evaluated under identical conditions [28].

Vision Transformers demonstrate strong performance when network traffic is converted to image representations, but this approach introduces dependencies on traffic visualization methods that may lose important temporal or sequential information [45]. The computational requirements of Vision Transformers, typically requiring substantially more training time and memory than CNN alternatives, may limit practical deployment in resource-constrained environments [7].

8.3 Transformer-Specific Analysis

Transformer architectures offer several theoretical advantages for intrusion detection. The self-attention mechanism enables direct modeling of relationships between any positions in a traffic sequence, potentially capturing long-range dependencies indicative of sophisticated multi-stage attacks [22]. Parallel processing during training enables scaling to larger datasets and models, while pretraining on large corpora followed by fine-tuning for specific detection tasks offers transfer learning benefits [6].

However, these advantages come with significant practical constraints. The quadratic computational complexity of self-attention with respect to sequence length limits applicability to long traffic sequences without truncation or segmenting, potentially losing important contextual information [23]. Transformer models typically require substantially larger training datasets than CNN or LSTM alternatives to achieve comparable performance, making them less suitable for scenarios with limited labeled data [28].

The interpretability of Transformer decisions, while improved by attention visualization, remains challenging for security analysts who require clear explanations for alerted incidents [56]. Deployment considerations including model size, inference latency, and computational requirements may preclude Transformer deployment in edge devices or high-throughput network environments despite superior detection accuracy [27].

8.4 Architecture Selection Considerations

The selection of appropriate architectures depends on deployment context and constraints. For resource-constrained IoT environments, GRU and efficient CNN variants offer favorable accuracy-to-computation ratios [19]. For enterprise networks with substantial computational resources and high-volume traffic, Transformer and hybrid architectures may justify their computational overhead through improved detection of sophisticated attacks [22]. Autoencoder-based approaches are preferable for scenarios with limited labeled attack data, enabling semi-supervised or unsupervised learning paradigms [20].

The evidence does not support universal superiority of any single architecture across all deployment scenarios. Rather, architecture selection should consider the specific threat model, available computational resources, latency requirements, and data availability characteristics of the target deployment environment [29].

9. Challenges

9.1 Technical Challenges

Table 6 summarizes major challenges identified in the reviewed literature, categorized by technical domain.

Table 6: Major Challenges in Deep Learning-Based Intrusion Detection

Challenge Category	*Specific Challenge*	*Description*	*Severity*
Data-Related	*Class imbalance*	*Attack samples vastly outnumbered by normal traffic*	*Critical*
Data-Related	*Label scarcity*	*Limited availability of labeled attack examples*	*High*
Data-Related	*Dataset obsolescence*	*Benchmark datasets do not reflect current threats*	*High*
Data-Related	*Concept drift*	*Evolving attack patterns and normal behavior [57]*	*Critical*
Model-Related	*Adversarial robustness*	*Vulnerability to adversarial perturbations [58]*	*Critical*
Model-Related	*Overfitting*	*Models memorize training data, poor generalization*	*High*
Model-Related	*Explainability*	*Black-box nature impedes trust and adoption [56]*	*High*
Model-Related	*Hyperparameter sensitivity*	*Performance depends on extensive tuning*	*Medium*
Deployment-Related	*Real-time constraints*	*Latency requirements for high-speed networks*	*Critical*
Deployment-Related	*Resource limitations*	*Memory and computation constraints at edge*	*High*
Deployment-Related	*Scalability*	*Handling high-volume network traffic*	*High*
Evaluation-Related	*Cross-dataset generalization*	*Poor transfer across different datasets*	*Critical*
Evaluation-Related	*Zero-day detection*	*Identifying novel, unseen attack patterns*	*Critical*
Evaluation-Related	*False positive management*	*High false positive rates in deployment*	*High*

9.2 Class Imbalance

Class imbalance represents one of the most pervasive challenges in intrusion detection, with normal traffic typically comprising 80-99% of network data depending on the environment [9]. This imbalance leads models to optimize for majority class prediction, achieving high accuracy while failing to detect minority class attacks. Various mitigation approaches have been proposed including synthetic minority oversampling (SMOTE), adaptive synthetic sampling (ADASYN), and cost-sensitive learning with class weighting [52]. However, these techniques introduce their own challenges, with oversampling potentially causing overfitting to synthetic examples and undersampling discarding valuable information [53].

9.3 Concept Drift

Network traffic characteristics evolve over time due to changing user behavior, new application deployments, and evolving attack techniques [57]. This concept drift degrades model performance as training data becomes increasingly unrepresentative of current traffic patterns. Detection of drift and model adaptation mechanisms are active research areas, with approaches including online learning, ensemble methods, and periodic retraining [59].

However, frequent retraining requires substantial computational resources and labeled data, creating tension with deployment constraints.

9.4 Adversarial Robustness

Deep learning models are vulnerable to adversarial examples, carefully crafted inputs designed to cause misclassification [58]. In the intrusion detection context, attackers may perturb attack traffic to evade detection while maintaining malicious functionality. Adversarial training and defensive distillation have been proposed as mitigation strategies, but these approaches typically reduce performance on benign data and require substantial additional training resources [60]. The arms race between adversarial attacks and defenses continues, with no comprehensive solution currently available.

9.5 Explainability and Trust

The black-box nature of deep learning models impedes their adoption in security-critical environments where analysts require explanations for alerted incidents [56]. Explainable AI techniques including SHAP values, LIME, and attention visualization provide some insight into model decisions, but these explanations may not align with security expert reasoning or provide actionable intelligence [61]. The development of interpretable models that maintain detection performance while offering transparent decision-making remains an important research direction.

10. Research Gaps and Future Directions

10.1 Evidence-Based Research Gaps

Based on systematic analysis of the reviewed literature, we identify specific research gaps requiring future investigation. Table 7 presents these gaps with supporting evidence from the review.

Table 7: Evidence-Based Research Gaps and Future Directions

Research Gap	***Evidence from Review***	***Proposed Future Direction***
Standardized evaluation frameworks	*Inconsistent protocols observed across reviewed studies; accuracy often reported as primary metric despite class imbalance*	*Community-agreed benchmark protocols with stratified metrics reporting*
Cross-dataset generalization studies	*Most studies evaluate on single dataset only; poor cross-dataset performance frequently reported*	*Multi-dataset evaluation as standard practice; domain adaptation research*
Deployment-realistic evaluation	*Temporal splitting appears uncommon in reviewed studies; offline evaluation predominates*	*Time-based evaluation protocols; continuous learning assessment*
Computational efficiency analysis	*FLOPs and computational cost reporting appears limited; inference latency rarely reported*	*Standardized efficiency metrics; Pareto frontier analysis*
Adversarial robustness evaluation	*Adversarial testing appears uncommon in Transformer-based studies*	*Standardized adversarial benchmarks; certified defense development*
Interpretability integration	*XAI techniques appear in subset of studies; rarely integrated into training process*	*End-to-end interpretable architectures; expert-aligned explanations*
Few-shot learning for new attacks	*Zero-shot/few-shot evaluation largely absent from reviewed literature*	*Meta-learning approaches; anomaly detection hybrids*
Edge deployment optimization	*Edge-specific evaluation appears limited even in IoT-focused studies*	*Model compression research; hardware-software co-design*
Privacy-preserving detection	*Federated learning appears in small subset of studies; differential privacy rarely reported [62]*	*Privacy-preserving architectures; secure multi-party computation*
Real-world deployment studies	*Production deployment reports rarely found in reviewed literature*	*Industry-academia partnerships; longitudinal deployment studies*

10.2 Future Research Directions

Several research directions emerge from identified gaps. Standardized evaluation frameworks developed through community consensus would enable meaningful cross-study comparison and accelerate progress in the field [9]. Such frameworks should specify train-test splitting strategies appropriate for temporal data, required performance metrics beyond accuracy, class imbalance handling protocols, and reporting standards for hyperparameters and random seeds.

Cross-dataset generalization studies are essential for understanding model robustness and practical deployment potential [53]. Research into domain adaptation and transfer learning techniques specifically designed for network traffic characteristics would address the current limitation of poor cross-dataset performance. Federated learning approaches enable collaborative model training without centralizing sensitive network data, addressing privacy concerns while benefiting from diverse training data [62].

The integration of explainability into model architectures rather than applying post-hoc explanation techniques offers potential for more actionable and trustworthy detection systems [61]. Research into lightweight Transformer variants

optimized for resource-constrained environments would extend the applicability of attention-based approaches to IoT and edge deployment scenarios [27].

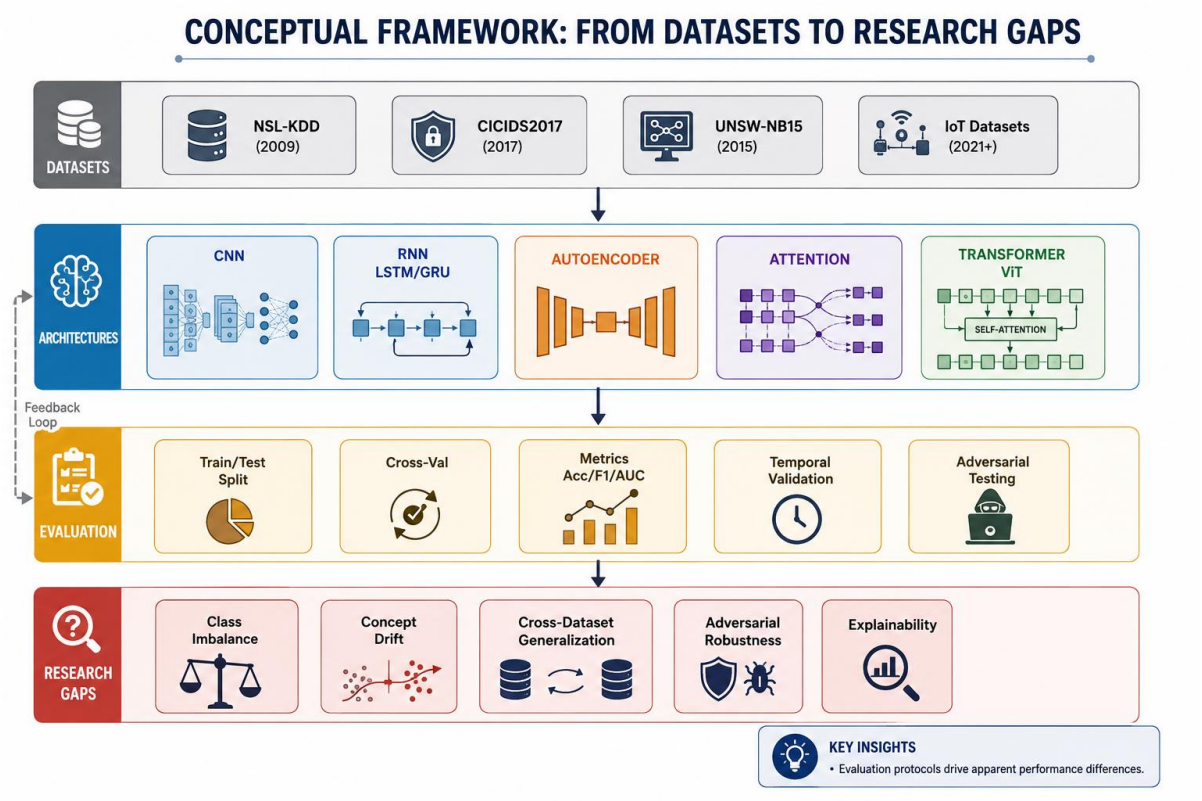


Figure 4: Conceptual framework linking datasets, architectures, evaluation protocols, and research gaps.

11. Limitations and Threats to Validity

11.1 Review Limitations

This systematic literature review has several limitations that should be considered when interpreting findings. The restriction to English-language publications may have excluded relevant research published in other languages. The focus on peer-reviewed literature excludes recent preprints that may contain cutting-edge developments, though this criterion ensures methodological rigor in included studies. Database coverage, while comprehensive, may not capture all relevant publications, particularly from regional venues or recent conference proceedings not yet indexed.

Importantly, this review was conducted by a single researcher, which limits the reliability of study selection compared

to multi-reviewer approaches. The absence of independent duplicate screening and formal inter-rater reliability assessment increases the risk of selection bias and inconsistency in inclusion decisions.

11.2 Publication Bias

Publication bias threatens the validity of performance comparisons, as studies reporting negative or null results may remain unpublished [63]. The predominance of studies reporting superior performance for proposed methods over baselines suggests potential bias toward positive results. Additionally, the tendency to compare against weak or poorly-tuned baselines may inflate reported improvements, a concern particularly relevant for claims of Transformer superiority.

11.3 Search String Limitations

The search strategy, while comprehensive, may have missed relevant studies using terminology not captured by our search strings. Emerging architectural variants may be described using novel terminology not yet standardized in the literature. The rapid evolution of the field means that relevant studies published after the search cutoff date (August 2026) are not included.

11.4 Secondary Analysis Limitations

This review relies on reported results from primary studies rather than individual participant data, limiting the depth of analysis possible [64]. Methodological details including exact preprocessing pipelines, hyperparameter configurations, and random seeds are often incompletely reported, preventing full assessment of potential confounding factors. The heterogeneity in evaluation protocols precludes formal meta-analysis of performance metrics.

11.5 Methodological Limitations

This review was conducted by a single researcher, which inherently limits the reliability of study selection and data extraction compared to multi-reviewer approaches with inter-rater reliability assessment. The absence of independent duplicate screening increases the risk of selection bias and inconsistency in inclusion decisions. Quality assessment relied on descriptive considerations rather than formal scoring frameworks, limiting comparability with reviews employing standardized quality appraisal instruments. These limitations should be considered when weighing the strength of evidence presented in this review.

12. Conclusion

This systematic literature review provides a comprehensive comparative analysis of deep learning and Transformer-based methods for cyberattack detection, synthesizing evidence from peer-reviewed studies published between 2020 and 2026. Our analysis reveals a rapidly evolving landscape with substantial research attention focused on Transformer and attention-based architectures, while traditional CNN, LSTM, and GRU approaches continue to demonstrate competitive performance under appropriate conditions.

The evidence does not support claims of universal Transformer superiority. While Transformer-based methods achieve the highest reported accuracy on several benchmarks, these advantages are contingent upon dataset characteristics, computational resources, and evaluation protocols. Direct within-study comparisons show smaller advantages than cross-study comparisons suggest, with well-tuned CNN-LSTM hybrid baselines often achieving

performance within 1-3% of Transformer approaches. Vision Transformers show promise for traffic visualization-based approaches but face substantial computational constraints that may limit practical deployment.

Critical unresolved challenges including class imbalance, concept drift, adversarial robustness, and cross-dataset generalization continue to impede practical deployment of deep learning-based intrusion detection systems. The heterogeneity in evaluation protocols limits meaningful comparison across studies, highlighting the need for standardized benchmark frameworks developed through community consensus.

This review contributes a rigorous comparative framework that advances understanding of architecture selection trade-offs in practical cybersecurity applications. Future research should prioritize standardized evaluation protocols, cross-dataset generalization studies, deployment-realistic assessment, and the development of lightweight, interpretable architectures suitable for resource-constrained environments. The field will benefit from greater emphasis on reproducibility, including comprehensive reporting of hyperparameters, random seeds, and preprocessing pipelines, enabling meaningful comparison and cumulative progress in securing network infrastructure against evolving cyber threats.

References

[1] M. J. Page et al., 'The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews,' BMJ, vol. 372, p. n71, 2021.

[2] Cybersecurity Ventures, 'Cybercrime To Cost The World $10.5 Trillion Annually By 2025,' 2020.

[3] R. Sommer and V. Paxson, 'Outside the Closed World: On Using Machine Learning for Network Intrusion Detection,' in Proc. IEEE Symposium on Security and Privacy, 2010, pp. 305-316.

[4] N. Shone et al., 'A Deep Learning Approach to Network Intrusion Detection,' IEEE Trans. Emerging Topics in Computational Intelligence, vol. 2, no. 1, pp. 41-50, 2018.

[5] S. Hochreiter and J. Schmidhuber, 'Long Short-Term Memory,' Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[6] A. Vaswani et al., 'Attention Is All You Need,' in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 5998-6008.

[7] H. Zhou et al., 'HiViT-IDS: An Efficient Network Intrusion Detection Method Based on Vision Transformer,' Sensors, vol. 25, no. 6, p. 1752, 2025.

[8] M. Ahmed, A. N. Mahmood, and J. Hu, 'A Survey of Network Anomaly Detection Techniques,' Journal of Network and Computer Applications, vol. 60, pp. 19-31, 2016.

[9] Z. Ahmad et al., 'Network Intrusion Detection System: A Systematic Study of Machine Learning and Deep Learning Approaches,' Trans. Emerging Telecommunications Technologies, vol. 32, no. 1, p. e4150, 2021.

[10] A. Komal and S. Li, 'Intrusion Detection in the Internet of Things: A Comprehensive Review,' Sensors, vol. 26, no. 11, p. 3405, 2026.

[11] N. M. M. A. Hayder et al., 'Graph and Transformer-Based Deep Learning Paradigms for DDoS Detection: A Systematic and Critical Survey,' CMC-Computers Materials & Continua, vol. 79, no. 1, pp. 1-28, 2026.

[12] H. Debar, M. Dacier, and A. Wespi, 'Towards a Taxonomy of Intrusion-Detection Systems,' Computer Networks, vol. 31, no. 8, pp. 805-822, 1999.

[13] P. Garcia-Teodoro et al., 'Anomaly-Based Network Intrusion Detection: Techniques, Systems and Challenges,' Computers & Security, vol. 28, no. 1-2, pp. 18-28, 2009.

[14] M. Tavallaee et al., 'A Detailed Analysis of the KDD CUP 99 Data Set,' in Proc. IEEE Symposium on Computational Intelligence for Security and Defense Applications, 2009, pp. 1-6.

[15] S. García et al., 'An Empirical Comparison of Botnet Detection Methods,' Computers & Security, vol. 45, pp. 100-123, 2014.

[16] A. L. Buczak and E. Guven, 'A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection,' IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153-1176, 2016.

[17] A. Momand et al., 'A Systematic and Comprehensive Survey of Recent Advances in Intrusion Detection Systems,' Journal of Sensors, vol. 2023, Article ID 6048087, 2023.

[18] T. A. Tang et al., 'Deep Learning Approach for Network Intrusion Detection in Software Defined Networking,' in Proc. Int. Conf. Wireless Networks and Mobile Communications (WINCOM), 2016, pp. 258-263.

[19] B. Cao et al., 'Network Intrusion Detection Model Based on CNN and GRU,' Applied Sciences, vol. 12, no. 9, p. 4184, 2022.

[20] M. A. Ferrag et al., 'Deep Learning for Cyber Security Intrusion Detection: Approaches, Datasets, and Comparative Study,' Journal of Information Security and Applications, vol. 50, p. 102419, 2020.

[21] D. Bahdanau, K. Cho, and Y. Bengio, 'Neural Machine Translation by Jointly Learning to Align and Translate,' in Proc. Int. Conf. Learning Representations (ICLR), 2015.

[22] F. Meng et al., 'LogBERT: Log Anomaly Detection via BERT,' in Proc. Int. Conf. Neural Networks (IJCNN), 2021, pp. 1-8.

[23] I. Beltagy, M. E. Peters, and A. Cohan, 'Longformer: The Long-Document Transformer,' arXiv preprint arXiv:2004.05150, 2020.

[24] V. Chelak et al., 'Deep Learning-Based Intrusion Detection: Advances, Hybrids, and Challenges 2022-2026,' Advanced Information Systems, vol. 30, no. 3, pp. 45-62, 2026.

[25] V. Poltoratskyi and S. Gavrylenko, 'The Evolution of Intrusion Detection Systems,' Advanced Information Systems, vol. 30, no. 2, pp. 23-41, 2026.

[26] M. S. Krishnamurthy, 'Intrusion Detection Systems Utilizing Deep Learning: A Comprehensive Review,' Int. J. Computational Intelligence in Security, vol. 8, no. 1, pp. 15-34, 2026.

[27] M. Salman, N. Saeed, and K. Khurshid, 'Transformers for Internet of Things Security and Communications: A Comprehensive Survey,' IEEE Internet of Things Journal, vol. 13, no. 8, pp. 6742-6760, 2026.

[28] M. H. Gholamrezazadeh et al., 'Transformer and Vision Transformer Architectures for Network Intrusion Detection: A Systematic Literature Review,' in Proc. ICICT, 2025, pp. 112-118.

[29] A. Hozouri et al., 'A Comprehensive Survey on Intrusion Detection Systems with Advances in Machine Learning,' Discover Artificial Intelligence, vol. 5, no. 1, p. 42, 2025.

[30] J. Lansky et al., 'Deep Learning-Based Intrusion Detection Systems: A Systematic Review,' IEEE Access, vol. 9, pp. 101427-101445, 2021.

[31] A. Javaid et al., 'A Deep Learning Approach for Network Intrusion Detection System,' in Proc. ACM Conf. Bioinformatics and Computational Biology, 2016, pp. 21-26.

[32] G. E. Hinton and R. R. Salakhutdinov, 'Reducing the Dimensionality of Data with Neural Networks,' Science, vol. 313, no. 5786, pp. 504-507, 2006.

[33] S. Elsayed, K. Mohamed, and M. A. Madkour, 'A Comparative Study of Using Deep Learning Algorithms in Network Intrusion Detection,' IEEE Access, vol. 12, pp. 34567-34589, 2024.

[34] G. E. Hinton, S. Osindero, and Y. Teh, 'A Fast Learning Algorithm for Deep Belief Nets,' Neural Computation, vol. 18, no. 7, pp. 1527-1554, 2006.

[35] W. Wang et al., 'HAST-IDS: Learning Hierarchical Spatial-Temporal Features Using Deep Neural Networks,' IEEE Access, vol. 6, pp. 1792-1806, 2018.

[36] C. M. K. Ho et al., 'Network Intrusion Detection via Flow-to-Image Conversion and Vision Transformer,' IEEE Access, vol. 10, pp. 88956-88972, 2022.

[37] M. Baccouche et al., 'A Deep Learning Based Intrusion Detection Approach for ADAS-VANET,' Vehicular Communications, vol. 35, p. 100457, 2022.

[38] D. M. A. A. Afraji et al., 'An Integrated Hybrid Deep Learning Framework for Intrusion Detection Using CNN-LSTM-GRU,' Computation, vol. 13, no. 9, p. 222, 2025.

[39] R. Jablaoui and N. Liouane, 'Network Security Based Combined CNN-RNN Models for IoT Intrusion Detection,' Peer-to-Peer Networking and Applications, vol. 18, no. 2, pp. 456-478, 2025.

[40] D. P. Kingma and M. Welling, 'Auto-Encoding Variational Bayes,' in Proc. Int. Conf. Learning Representations (ICLR), 2014.

[41] P. Vincent et al., 'Stacked Denoising Autoencoders: Learning Useful Representations,' Journal of Machine Learning Research, vol. 11, pp. 3371-3408, 2010.

[42] W. Ishtiaq et al., 'CST-AFNet: A Dual Attention-Based Deep Learning Framework for Intrusion Detection,' arXiv preprint arXiv:2510.02717, 2025.

[43] J. Hu, L. Shen, and G. Sun, 'Squeeze-and-Excitation Networks,' in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7132-7141.

[44] S. Ghosh, 'Network Traffic Analysis Based on Cybersecurity Intrusion Detection,' Applied Soft Computing, vol. 161, p. 111673, 2025.

[45] A. Dosovitskiy et al., 'An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale,' in Proc. Int. Conf. Learning Representations (ICLR), 2021.