International Journal of Information and Communication Sciences
2026, Vol. 11, No. 2, pp. 47–69
https://doi.org/10.11648/j.ijics.20261102.13

Review Article

Toward Adaptive Intelligent Intrusion Detection: A Review of Zero-day Attack Detection Methods and Future Cyber Defence Directions

Uchechukwu Samuel Nwankwo[1], Obi Chukwuemeka Nwokonkwo[1], Charles Ikerionwu[2], Adetokunbo MacGregor John-Otumu[1,*], Udoka Felista Eze[1]

[1]Department of Information Technology, Federal University of Technology, Owerri, Nigeria

[2]Department of Software Engineering, Federal University of Technology, Owerri, Nigeria

Abstract

This study presents a comprehensive synthesis of artificial intelligence driven approaches for zero-day attack detection, addressing the growing limitations of traditional signature-based intrusion detection systems in dynamic and evolving cyber environments. The review examines major detection paradigms, dataset utilization patterns, and performance trends across widely adopted cybersecurity benchmarks such as NSL-KDD, UNSW-NB15, CICIDS2017, BoT-IoT, and ToN-IoT. The analysis reveals a clear progression from classical machine learning techniques, which rely heavily on handcrafted features, to more advanced deep learning and hybrid frameworks that leverage automated feature representation and behavioural modelling. These modern approaches have significantly improved detection capability, particularly in identifying complex and previously unseen attack patterns. However, despite these advancements, several critical challenges remain. These include limited dataset realism, weak cross-domain generalization, high computational overhead, vulnerability to adversarial manipulation, and lack of model interpretability, all of which constrain real-world deployment. In response to these limitations, this study proposes a unified adaptive hybrid detection framework that integrates anomaly-based monitoring, supervised classification, deception-driven intelligence, and predictive threat forecasting within a single architecture. Furthermore, the paper outlines key future research directions aimed at enhancing dataset diversity, improving model robustness, and advancing explainable artificial intelligence techniques. These contributions provide a strategic foundation for developing scalable, interpretable, and resilient intrusion detection systems capable of operating effectively in modern heterogeneous cybersecurity environments.

Keywords

Artificial Intelligence, Zero-day Attack Detection, Intrusion Detection System, Anomaly Detection, Adaptive Cyber Defence

*Correspondence: Adetokunbo MacGregor John-Otumu (adetokunbo.johnotumu@futo.edu.ng)

Received: 26 May 2026; **Accepted:** 8 June 2026; **Published:** 26 June 2026

1. Introduction

The rapid expansion of digital technologies and interconnected computing environments has significantly reshaped the cybersecurity landscape. Modern organizations now rely heavily on cloud computing, industrial systems, mobile platforms, and Internet of Things infrastructures to support critical operations. While these advancements have improved efficiency and connectivity, they have also widened the attack surface, making systems more vulnerable to increasingly sophisticated cyber threats. As a result, traditional security mechanisms are often unable to cope with the dynamic and evolving nature of modern attacks [1, 2].

Among the various forms of cyber threats, zero-day attacks remain one of the most critical challenges. These attacks exploit unknown vulnerabilities before patches or signatures are developed, making them extremely difficult to detect using conventional signature-based intrusion detection systems. Several studies have shown that reliance on predefined attack patterns limits the ability of traditional systems to respond effectively to emerging threats [3, 4]. This has driven the need for more intelligent and adaptive detection approaches capable of identifying previously unseen attack behaviours.

Artificial intelligence has emerged as a promising solution for addressing these challenges. Machine learning techniques such as Random Forest, Support Vector Machines, and ensemble models have demonstrated strong performance in detecting cyber-attacks by learning patterns from network traffic data rather than relying on static rules [5, 6]. These approaches have improved detection accuracy and reduced reliance on manual feature engineering.

More recently, deep learning methods have gained attention due to their ability to automatically extract complex and hierarchical features from large-scale data. Techniques such as convolutional neural networks, recurrent neural networks, and attention-based models have been successfully applied to capture both spatial and temporal characteristics of cyber threats [3, 7]. These models have shown improved capability in identifying multi-stage and evolving attack patterns.

In addition to supervised learning, anomaly detection approaches have been widely explored for zero-day attack detection. These methods focus on learning normal system behaviour and identifying deviations that may indicate malicious activity. Techniques such as autoencoders and isolation forests have shown strong potential in detecting previously unseen attacks without prior labeling [8, 9]. Hybrid approaches that combine anomaly detection with supervised learning have further improved detection robustness and adaptability.

Despite these advancements, several limitations remain. Many studies depend heavily on benchmark datasets, which may not accurately represent real-world traffic conditions, thereby limiting model generalization. In addition, deep learning models often require high computational resources, making deployment difficult in resource-constrained environments such as IoT and edge systems [10, 11]. Other challenges include data imbalance, adversarial attacks, and the scarcity of high-quality labeled datasets for zero-day scenarios.

Given these challenges, there is a strong need for a comprehensive and structured synthesis of existing research efforts. This study therefore presents a systematic review of artificial intelligence-driven approaches for zero-day attack detection. The review focuses on detection techniques, datasets used, performance trends, and existing research gaps, with the aim of providing clear insights and guiding future research directions in this rapidly evolving field.

2. Literature Review

Research on zero-day attack detection has evolved significantly in recent years due to the increasing sophistication of cyber threats and the limitations of traditional signature-based intrusion detection systems, which are often ineffective in identifying previously unseen attacks [3, 12]. Existing studies can broadly be categorized into classical machine learning approaches, deep learning-based detection frameworks, anomaly-driven detection strategies, hybrid intelligent defence architectures, and deception-based cybersecurity analytics, all of which have been explored to enhance detection capability and adaptability in modern cybersecurity environments [2, 10].

2.1. Classical Machine Learning Approaches for Intrusion Detection

Early research on intrusion detection largely focused on classical machine learning techniques such as decision trees, support vector machines, K-nearest neighbour, and random forest models. These approaches were effective in identifying known attack patterns by learning relationships between network traffic features and predefined labels [1]. For instance, models such as XGBoost and CatBoost achieved very high detection accuracy when combined with feature reduction techniques, showing the strength of traditional machine learning in structured environments [1].

However, despite their strong performance, these methods depend heavily on labelled datasets and predefined attack signatures. As a result, they often struggle to detect zero-day attacks, which do not follow previously known patterns. This limitation has been widely recognized, as many studies have shown that classical models fail to generalize effectively to unseen or evolving cyber threats [2, 13].

2.2. Deep Learning-Based Detection Frameworks

With the increasing availability of large-scale cybersecurity datasets, deep learning techniques have gained significant attention for intrusion detection. Models such as convolutional neural networks, recurrent neural networks, and transformer-

based architectures have shown strong capability in automatically learning complex patterns from network traffic data [3].

Deep learning models can significantly reduce detection time by shifting from static signature-based methods to behavioral analysis [3]. Similarly, a transformer-based framework that achieved high detection accuracy and strong robustness against adversarial attacks was introduced [14]. Other studies have also shown that deep learning models can achieve accuracy levels above 97% in detecting zero-day attacks [14, 15]. Despite these advantages, deep learning models require large amounts of labelled data and high computational resources. This makes them difficult to deploy in real-time environments, especially in resource-constrained systems such as IoT networks. In addition, issues such as overfitting, model interpretability, and vulnerability to adversarial attacks remain significant challenges [10, 16].

2.3. Anomaly Detection Techniques for Zero-day Threat Identification

Anomaly detection techniques have become an important approach for identifying zero-day attacks. Unlike supervised learning methods, these techniques focus on learning normal system behavior and detecting deviations as potential threats. This makes them particularly suitable for detecting unknown attacks [5].

Isolation Forest and deep neural networks can effectively identify anomalous traffic patterns associated with zero-day attacks [5]. Similarly, Autoencoder-based models outperform traditional methods in detecting complex and previously unseen attack patterns [17]. However, anomaly detection methods often produce high false positive rates, especially in dynamic environments where normal behavior changes over time. This can reduce their reliability in real-world deployment. In addition, many studies rely on simulated datasets, which may not fully represent real network conditions [8, 18].

2.4. Hybrid Intelligent Intrusion Detection Systems

To overcome the limitations of individual approaches, hybrid intrusion detection systems have been proposed. These systems combine multiple techniques such as machine learning, deep learning, and anomaly detection to improve overall performance. A hybrid model that integrates autoencoders with supervised classifiers was developed, achieving near-perfect detection accuracy on unseen data [19]. Similarly, LSTM was combined with Isolation Forest to improve detection performance for zero-day attacks, achieving high accuracy and low false positive rates [20]. Hybrid models have been shown to provide better detection accuracy and robustness compared to single-model approaches. However, they are often complex and computationally expensive, which can limit their practical implementation in real-time systems [2, 10].

2.5. Deception-Based Cybersecurity Analytics

Deception-based approaches such as honeypots have also been explored as proactive strategies for zero-day attack detection. These techniques involve creating decoy systems to attract attackers and collect valuable data for analysis. Integrating machine learning with deception mechanisms significantly improves detection performance and provides proactive defense against cyber threats [6]. Similarly, honeypot-based data collection was used to generate real-world attack datasets, enabling more accurate analysis of attack behavior [21].

Although these approaches provide valuable insights, they may introduce bias due to the use of simulated or controlled environments. This can affect the generalization of models when applied to real-world scenarios [21].

2.6. Taxonomy of Artificial Intelligence Driven Zero-day Detection Approaches

To provide a clearer analytical comparison of existing approaches, Table 1 summarizes the strengths, limitations, and operational characteristics of major zero-day detection paradigms identified in literature.

Existing research on intelligent intrusion detection demonstrates significant methodological diversity in the development of artificial intelligence driven zero-day attack detection frameworks. Based on underlying learning paradigms, detection objectives, and operational deployment strategies, current approaches can be broadly categorized into five major groups. This taxonomy provides a structured lens for understanding the evolution of intelligent cyber defence mechanisms and highlights the progression from traditional classification-based techniques toward adaptive and intelligence driven threat detection architectures.

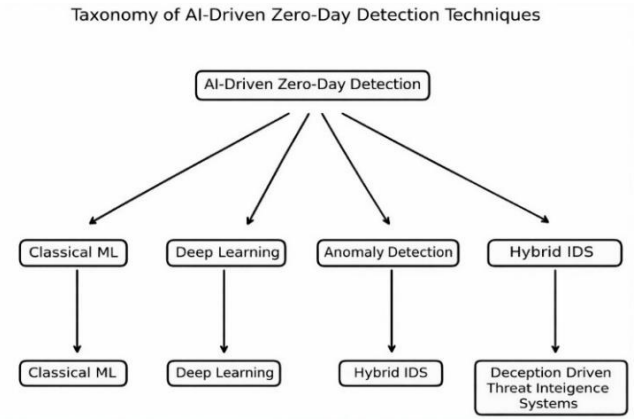

Figure 1. *Taxonomy of AI-Driven Zero-Day Detection Techniques.*

Figure 1 Taxonomy of AI-Driven Zero-Day Detection Techniques presents a structured classification of intelligent detection paradigms used in cybersecurity research. The diagram shows that AI-driven zero-day detection evolves from

classical machine learning and deep learning approaches toward anomaly detection and hybrid intrusion detection systems. It further highlights the integration of deception driven threat intelligence mechanisms as an advanced strategy for proactive cyber threat monitoring and adaptive defence.

Table 1 presents a comparative evaluation of major zero-day detection paradigms, highlighting their core techniques, strengths, and operational limitations. Classical machine learning methods offer interpretable and computationally efficient detection but struggle with unseen threats. Deep learning improves feature learning and detection accuracy at higher computational cost. Anomaly and hybrid approaches enhance zero-day detection capability, while deception-driven systems support proactive intelligence gathering, though concerns remain regarding false alarms, complexity, and dataset realism.

Table 1. *Comparative Evaluation of Zero-Day Detection Paradigms.*

Core Techniques	Key Strengths	Major Limitations	Zero-Day Detection Capability	Computational Cost
DT, SVM, KNN, RF	Interpretable models, lower computational requirements, effective for known attacks	Heavy dependence on handcrafted features and labelled datasets	Moderate	Low
CNN, RNN, LSTM, Transformer models	Automated feature extraction, ability to capture complex temporal and spatial patterns	Requires large training datasets and significant computational resources	High	High
Autoencoders, Isolation Forest, clustering algorithms	Detects previously unseen attacks by modelling normal behaviour	Higher false positive rates in dynamic network environments	Very High	Moderate
Combination of anomaly detection and supervised learning models	Improved robustness, better generalization across attack types	Architectural complexity and training overhead	Very High	High
Honeypots, Honeynets, deception platforms	Generates realistic attack intelligence, supports proactive threat detection	Synthetic or simulated traffic may introduce bias	High	Moderate

2.7. Synthesis of Literature

The reviewed studies show that zero-day attack detection has evolved from classical machine learning approaches to more advanced deep learning and hybrid frameworks. Classical methods provide simplicity and efficiency but lack the ability to detect unknown threats. Deep learning models improve detection accuracy by learning complex patterns, while anomaly detection methods enable the identification of previously unseen attacks [2, 3].

More recent studies emphasize hybrid frameworks that combine multiple techniques to enhance detection performance and robustness. These approaches have shown promising results in controlled environments, achieving high accuracy and improved detection rates [14, 19].

However, several challenges remain unresolved. These include limited availability of high-quality datasets, class imbalance, computational complexity, vulnerability to adversarial attacks, and lack of real-world validation. These limitations highlight the need for more adaptive, scalable, and intelligent frameworks capable of handling evolving zero-day threats in diverse environments [10, 16].

2.8. Challenges and Limitations of Existing Zero-day Detection Systems

The comparative analysis of the reviewed studies reveals several persistent challenges affecting the effectiveness and real-world deployment of artificial intelligence driven intrusion detection systems. One major limitation is the continued reliance on benchmark datasets such as NSL-KDD, UNSW-NB15, and CICIDS2017, which may not fully represent modern network traffic conditions or evolving cyber threats [2, 22].

In addition, dataset imbalance and limited representation of low-frequency and stealthy attack behaviours reduce detection reliability, especially in heterogeneous environments such as Internet of Things and industrial systems [10, 11].

Several studies also highlight the scarcity of high-quality labeled datasets for zero-day attacks, which limits the effectiveness of supervised learning models and reduces their ability to generalize across different environments [16, 26]. Although deep learning and hybrid models demonstrate strong detection capability, their computational complexity and scalability challenges hinder deployment in real-time and resource-constrained environments [8, 14].

Furthermore, many detection systems remain vulnerable to adversarial manipulation, where attackers craft inputs that mimic normal behaviour to evade detection [10, 20].

Another important limitation is the lack of explainability in complex models, which reduces transparency and affects trust in automated detection systems [7, 18].

Finally, the limited integration of proactive mechanisms such as threat forecasting, continual learning, and deception-based intelligence indicates that many current systems remain reactive rather than adaptive [6, 24].

2.9. Synthesized Research Gaps

Based on the reviewed studies, several key research gaps are identified. First, there is a lack of realistic and heterogeneous datasets that accurately capture evolving zero-day attack behaviours across diverse environments [10, 18]. Second, many studies report high detection performance using single datasets, raising concerns about overfitting and limited generalization across different network environments [19, 29].

Third, cross-domain validation remains insufficient, as many models fail to maintain consistent performance when applied to different datasets or operational contexts [13, 30]. Fourth, computational overhead associated with deep learning and hybrid models limits real-time deployment, particularly in large-scale and resource-constrained systems [8, 14].

Fifth, many detection systems remain vulnerable to adversarial attacks and evolving threat strategies, highlighting the need for more robust and adaptive frameworks [10, 16]. Sixth, there is limited integration of predictive and proactive defence mechanisms capable of identifying threats before exploitation occurs [6]. Finally, the lack of explainable and transparent decision-making processes in many models limits their adoption in practical cybersecurity operations [7]. These gaps highlight the need for adaptive hybrid intrusion detection frameworks that support cross-domain learning, real-time deployment, explainability, and proactive threat intelligence integration.

3. Methodology

This study adopts a systematic review methodology to identify, analyze, and synthesize existing research contributions on artificial intelligence driven zero-day attack detection. A structured literature search strategy was employed to ensure comprehensive coverage of relevant scientific publications across multiple cybersecurity research domains.

3.1. Justification for PRISMA-Based Study Selection

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework was adopted in this study to ensure transparency, reproducibility, and methodological rigor in the literature selection process. The PRISMA approach provides a structured mechanism for identifying, screening, and selecting relevant studies while minimizing selection bias and improving the reliability of synthesized findings (See Figure 2).

In the context of zero-day attack detection research, where studies vary significantly in datasets, methodologies, and evaluation strategies, the use of PRISMA supports consistent filtering and ensures that only relevant and high-quality studies are included. Furthermore, it enhances the traceability of the selection process, thereby strengthening the credibility and scientific validity of the review.

3.2. Literature Search Strategy

Relevant studies were retrieved from major scientific databases, as summarized in Table 2. The search process involved the use of predefined keyword combinations presented in Table 3. In addition, Boolean operators and keyword variations were systematically applied to enhance search sensitivity and broaden coverage, ensuring the inclusion of interdisciplinary studies related to behavioural intrusion detection and intelligent cyber defence mechanisms. The formulated search strings are detailed in Table 4.

Table 2. *Database sources for thematic analysis.*

S/N	Database Source	Purpose of Inclusion
1	Scopus	Comprehensive indexing of high-quality journal publications in cybersecurity and artificial intelligence
2	IEEE Xplore	Access to conference proceedings and technical papers related to intrusion detection and intelligent security systems.
3	SpringerLink	Retrieval of interdisciplinary cybersecurity research and machine learning applications
4	ScienceDirect	Inclusion of engineering and computer science journal articles relevant to intelligent threat detection
5	ACM Digital Library	Access to cutting-edge research in network security, data analytics, and cyber defence frameworks
6	Google Scholar	Supplementary search to capture recently published or cross-disciplinary studies

Table 3. Literature Search Strategy.

S/N	Keyword / Search Phrase	Purpose of Use
1	Zero-day attack detection	To retrieve studies specifically addressing detection of unknown or previously unseen cyber threats
2	Intrusion detection system	To capture general research on network intrusion detection frameworks and security monitoring mechanisms
3	Machine learning cybersecurity	To identify studies applying classical machine learning techniques to cybersecurity problems
4	Deep learning intrusion detection	To retrieve research involving neural network-based detection models and automated feature learning approaches
5	Anomaly-based cyber-attack detection	To include studies focusing on behavioural deviation modelling and unsupervised threat identification
6	Intelligent network security	To capture broader artificial intelligence-driven cyber defence architectures and adaptive security systems

Table 4. Boolean Search Strings Used for Literature Retrieval.

S/N	Search String Formulation	Purpose
1	("zero-day attack" OR "unknown attack") AND ("intrusion detection" OR "IDS")	To retrieve studies focusing on detection of previously unseen cyber threats within intrusion detection frameworks.
2	("machine learning" OR "classification") AND ("cybersecurity" OR "network security")	To identify research applying classical machine learning techniques to cyber threat detection.
3	("deep learning" OR "neural network") AND ("intrusion detection" OR "network anomaly")	To capture studies using deep neural architectures for behavioural traffic analysis and attack classification
4	("anomaly detection" OR "behavioural analysis") AND ("cyber-attack" OR "network intrusion")	To include research addressing unsupervised or semi-supervised detection of abnormal network activities
5	("hybrid intrusion detection" OR "ensemble IDS") AND ("cyber defence" OR "threat detection")	To retrieve studies proposing integrated detection architectures combining multiple analytical paradigms.
6	("honeypot" OR "deception system") AND ("threat intelligence" OR "attack detection")	To identify research on deception-based cybersecurity analytics and proactive threat monitoring.

3.3. Inclusion Criteria

Studies considered relevant to the objectives of this review were selected based on clearly defined inclusion conditions. These criteria were applied to ensure that only high-quality and methodologically appropriate research contributions related to artificial intelligence driven intrusion detection were retained for detailed analysis. A summary of the inclusion requirements used during the study selection process is presented in Table 5.

Table 5. Summary of inclusion Criteria.

S/N	Inclusion Criterion	Description
1	Peer-reviewed publications	Only journal articles and conference papers that underwent peer review were considered to ensure scientific quality and credibility

S/N	Inclusion Criterion	Description
2	Artificial intelligence-based intrusion detection focus	Studies investigating machine learning, deep learning, anomaly detection, or hybrid intelligent techniques for intrusion detection were included
3	Empirical experimental validation	Selected papers were required to present experimental evaluation results using benchmark or real cybersecurity datasets
4	Relevance to unknown or evolving cyber threats	Research addressing detection of zero-day attacks, anomaly-based intrusion detection, or adaptive cyber defence mechanisms was prioritized
5	Publication time window	Only studies published between 2009 and 2026 were included to capture contemporary developments in intelligent cybersecurity research

3.4. Exclusion Criteria

Studies that did not meet the defined relevance and quality requirements were excluded from this review. The exclusion process was guided by specific criteria aimed at ensuring methodological rigor and thematic alignment with artificial intelligence driven intrusion detection research. A summary of the exclusion conditions applied during the screening process is presented in Table 6.

Table 6. *Exclusion Criteria for Study Selection.*

S/N	Exclusion Criterion	Description
1	Non-peer-reviewed publications	Editorials, opinion papers, blog articles, white papers, and unpublished manuscripts were excluded to maintain scientific rigor
2	Lack of experimental evaluation	Studies that did not provide empirical validation using cybersecurity datasets or real network traffic were excluded
3	Irrelevant research focus	Papers focusing solely on cryptography, access control, malware analysis without intrusion detection context, or general network optimization were excluded
4	Duplicate publications	Duplicate records retrieved from multiple databases were removed during the screening process
5	Insufficient methodological description	Studies lacking clear description of detection techniques, dataset usage, or evaluation metrics were excluded

3.5. Study Selection Process

The study selection process followed a structured and systematic approach to ensure the inclusion of relevant and methodologically sound studies. An initial total of 95 publications were identified through comprehensive searches across selected scientific databases. Following the removal of duplicate records, 80 unique studies remained for preliminary screening. Subsequent title and abstract screening was conducted to assess relevance to artificial intelligence-driven intrusion detection and zero-day attack analysis. Studies that did not directly address the research focus or lacked empirical validation were excluded, resulting in a reduced set of 45 articles for full-text assessment. During the eligibility stage, each study was critically evaluated based on methodological rigor, dataset utilization, and the clarity and completeness of reported performance metrics. Articles with insufficient experimental detail, ambiguous evaluation procedures, or limited relevance to zero-day attack detection were excluded. As a result, a final set of 30 studies was selected for detailed synthesis and comparative analysis. This systematic filtering process ensured that the selected studies represent diverse methodological approaches while maintaining consistency, relevance, and scientific rigor.

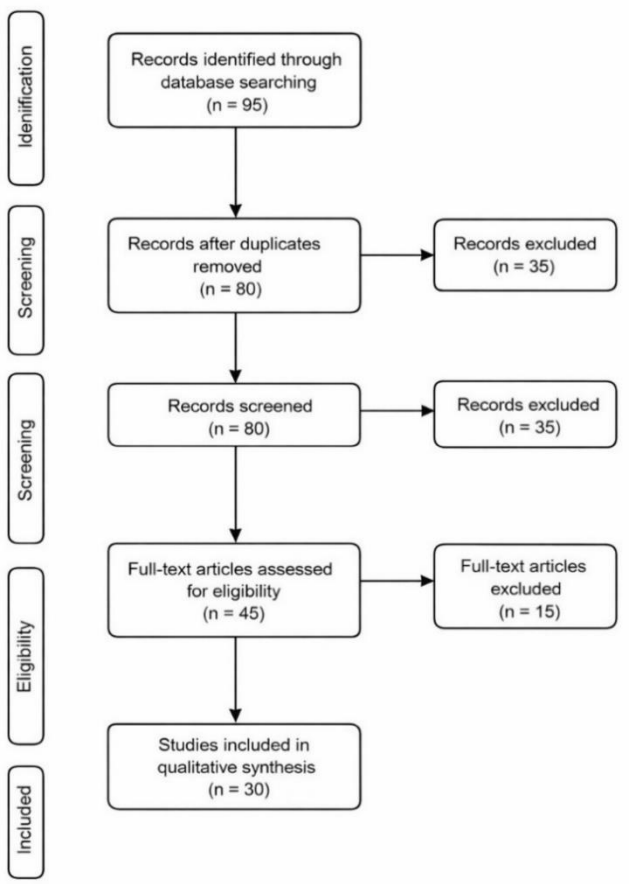

Figure 2. *PRISMA Flow Diagram of Study Selection Process for Zero-Day Attack Detection Review.*

Figure 2 illustrates the structured study selection process employed in this review using the PRISMA framework. A total of 95 records were initially identified through database searching. After duplicate removal, 80 records remained and were subjected to title and abstract screening. During this stage, 35 records were excluded based on predefined relevance criteria. A total of 45 full-text articles were subsequently assessed for eligibility, out of which 15 were excluded due to insufficient methodological rigor or lack of focus on artificial intelligence-based zero-day attack detection. Ultimately, 30 studies were included in the qualitative synthesis for detailed analysis.

3.6. Study Quality Assessment

To ensure methodological rigor and enhance the reliability of the review findings, each selected study was evaluated using a structured quality assessment framework. The assessment was designed to capture key dimensions of research quality, including dataset relevance, methodological soundness, evaluation completeness, and practical applicability to real-world cybersecurity environments. Each study was assessed using a standardized scoring scheme ranging from 1 to 5 across four core criteria:

(a) Dataset quality and realism, which considers whether the study makes use of data that is diverse and closely reflects real cybersecurity environments

(b) Methodological soundness, which examines how well the proposed detection approach is designed, including its clarity, structure, and overall reliability

(c) Evaluation completeness, which looks at whether the study uses appropriate performance measures and proper validation methods to assess the model

(d) Relevance to zero-day attack detection, which evaluates how well the study contributes to identifying new or previously unknown cyber threats

The cumulative score obtained across these criteria was used to classify studies into high-quality, moderate-quality, and low-quality categories. This classification enabled a more structured synthesis of findings and ensured that the analysis prioritizes studies with strong methodological foundations and practical relevance.

Table 7 outlines the criteria used to assess the quality of the selected studies. The framework ensures that the review is based on research that is both methodologically sound and practically relevant. It considers key aspects such as the use of realistic datasets, the clarity and reliability of the proposed methods, and the completeness of performance evaluation. It also examines how well each study reflects real-world application. This approach allows for fair comparison and helps identify studies with stronger contributions.

Table 7. *Study Quality Assessment Criteria.*

Criterion	Description	Score Range
Dataset Quality	Use of realistic and diverse datasets	1–5
Methodological Rigor	Clarity and robustness of model design	1–5
Evaluation Metrics	Use of comprehensive metrics (Accuracy, Recall, AUC, etc.)	1–5
Real-world Relevance	Applicability to practical deployment scenarios	1–5

3.7. Bias and Limitations in Study Selection

Despite the structured selection process, potential sources of bias may exist in the reviewed literature. First, publication bias may influence the inclusion of studies reporting higher performance results, as studies with lower accuracy or negative outcomes are less frequently published.

Second, dataset bias remains a significant concern, as many studies rely on a limited number of benchmark datasets, which may not accurately represent real-world network environments. This can lead to inflated performance estimates and reduced generalization capability.

Third, methodological bias may arise from variations in experimental design, including differences in preprocessing techniques, feature selection methods, and validation strategies.

To mitigate these effects, this study incorporates comparative analysis across multiple studies, emphasizes dataset diversity, and critically evaluates reported performance results to ensure balanced and objective interpretation.

4. Results and Discussion

This section presents the key findings obtained from the analysis of zero-day attack detection studies. It discusses how different datasets, detection methods, and evaluation approaches influence performance outcomes. The results are examined to highlight trends, strengths, and limitations of existing models, while also providing insights into their practical applicability in real-world cybersecurity environments.

4.1. Dataset Analysis

The effectiveness of artificial intelligence driven zero-day attack detection frameworks is strongly influenced by the nature and quality of datasets used during model development and evaluation. Evidence from the reviewed studies indicates that detection accuracy and model robustness are closely linked to dataset diversity, realism of network traffic, and the ability to represent evolving cyber-attack patterns [2, 10]. For instance, several studies reported high detection accuracy when models were trained and tested on structured benchmark datasets; however, such performance often declines when exposed to more complex or real-world environments.

Although benchmark intrusion detection datasets provide a controlled environment for evaluating algorithm performance, they may not fully reflect real operational network conditions. Many studies rely on simulated or pre-labelled datasets, which can limit the ability of models to generalize emerging and previously unseen threats. This challenge has been widely recognized, as models trained on static datasets often struggle with concept drift and dynamic attack behaviours in real-time systems [3, 16].

4.2. Dataset Utilization Patterns in Zero-day Detection Research

Several datasets have emerged as dominant benchmarks in zero-day attack detection research due to their accessibility and standardized feature representation. The NSL-KDD dataset, for example, has been widely used as a foundational benchmark for evaluating machine learning based intrusion detection systems. However, it has been criticized for its outdated traffic characteristics and limited representation of modern cyber threats [22].

More recent datasets such as UNSW-NB15 and CICIDS2017 have gained significant popularity because they provide more realistic and diverse attack scenarios. These datasets include a wider range of modern attack types and richer feature sets, which contribute to improved model evaluation and performance analysis [1, 8]. As a result, many recent studies have adopted these datasets to enhance detection accuracy and better simulate real-world network environments.

In the context of industrial cybersecurity and Internet of Things environments, datasets such as ToN-IoT and Edge-IIoT have been introduced to support the evaluation of intrusion detection systems in heterogeneous and resource-constrained settings. These datasets incorporate multi-layered network traffic data and diverse attack patterns, making them suitable for assessing adaptive detection frameworks in distributed cyber-physical systems [11, 23].

Furthermore, some studies have attempted to address dataset limitations by generating synthetic attack data or combining multiple data sources. While this approach can improve dataset diversity, it may also introduce bias and reduce the reliability of performance evaluation in real-world scenarios [20, 24].

To provide a clearer analytical understanding of how dataset characteristics influence methodological design and performance outcomes, Table 8 highlights the progressive evolution of cybersecurity datasets from traditional simulated network environments toward more realistic, heterogeneous, and domain-specific traffic representations. Earlier datasets such as NSL-KDD and CSIC 2012 remain valuable for benchmarking and comparative evaluation; however, their limited realism and outdated attack patterns reduce their effectiveness for assessing modern zero-day attack detection systems. More recent datasets, including UNSW-NB15 and CICIDS2017, provide improved attack diversity and richer behavioural characteristics, thereby supporting the development of more robust machine learning and deep learning-based intrusion detection models.

The emergence of datasets such as ToN-IoT and Edge-IIoT reflects the growing importance of IoT and Industrial IoT security, providing realistic telemetry data for evaluating intelligent detection systems in heterogeneous and resource-constrained environments. Similarly, specialized datasets such as CIC-MalMem-2022, UGRansome, and MaleX enable focused

investigation of malware, ransomware, and memory-based attack behaviours, thereby supporting the development of domain-specific detection frameworks.

Despite these advances, the comparative analysis reveals persistent challenges, including dataset imbalance, limited attack diversity, insufficient labelled samples, and restricted cross-domain representation. These limitations may affect model generalization and contribute to performance inflation when evaluations are conducted on a single benchmark dataset. Consequently, future research should emphasize the development of realistic, diverse, and continuously updated datasets that better reflect evolving cyber threat landscapes and support more reliable evaluation of adaptive zero-day attack detection systems.

Table 8. *Comparative Characteristics of Cybersecurity Datasets Used in Reviewed Zero-Day Detection Studies.*

Dataset	Year Introduced	Traffic Environment	Attack Diversity	IoT / IIoT Support	Realism Level	Major Strength	Key Limitation
NSL-KDD	2009	Simulated enterprise network traffic	Moderate	No	Low	Reduced redundancy compared to KDD Cup dataset and widely benchmarked	Outdated attack patterns and limited representation of modern traffic behaviour
UNSW-NB15	2016	Hybrid real and synthetic enterprise traffic	High	Partial	Moderate	Diverse attack scenarios and balanced feature distribution	Limited industrial and IoT behavioural representation
CICIDS2017	2017	Realistic enterprise network flows	Very High	No	High	Rich behavioural features and comprehensive attack coverage	Large dataset complexity and preprocessing overhead
ToN-IoT	2020	IoT and IIoT telemetry environment	High	Yes	High	Captures heterogeneous device communication patterns	Limited labelled samples for some attack categories
Edge-IIoT	2022	Industrial edge computing traffic	High	Yes	High	Supports evaluation of edge-based intrusion detection models	Class imbalance and evolving attack dynamics
CIC-MalMem-2022	2022	Malware memory behaviour traces	Moderate	No	Moderate	Enables evaluation of memory-based detection techniques	Narrow focus on malware behaviour rather than full network context
UGRansome	2021	Ransomware network traffic	Moderate	No	Moderate	Useful for ransomware behaviour modelling	Limited diversity of intrusion categories
CSIC 2012	2012	Web application traffic simulation	Low	No	Low	Suitable for anomaly detection benchmarking	Synthetic behaviour patterns reduce generalization capability
MaleX	2024	Malware execution telemetry	Moderate	No	Moderate	Supports transformer-based malware detection evaluation	Binary classification dominance limits broader threat modelling

4.3. Comparative Performance Evidence Across Reviewed Studies

Empirical findings across the reviewed literature reveal noticeable variation in detection performance, largely influenced by dataset characteristics and model complexity. Classical machine learning models have demonstrated strong performance on structured benchmark datasets, particularly when feature engineering and dimensionality reduction techniques are applied [1].

However, their effectiveness tends to decline in dynamic

and real-world environments, where evolving attack patterns and unseen threats are more prevalent [2, 13].

Deep learning architectures have shown improved detection capability due to their ability to automatically learn complex feature representations and capture temporal behavioural patterns in network traffic. For instance, hybrid deep learning models integrating sequential and spatial learning mechanisms have achieved high accuracy and reduced false positive rates in zero-day attack detection [7]. Similarly, transformer-based and vision-oriented models have demonstrated strong adaptability and robustness in detecting complex and evolving cyber threats [25].

Despite these promising results, the high-performance values reported in many studies should be interpreted with caution. Several works highlight the risk of overfitting, particularly when models are evaluated using single datasets or controlled experimental conditions. For example, studies employing hybrid and ensemble models have reported near-perfect accuracy, yet acknowledge limitations related to dataset dependency and lack of cross-validation across diverse environments [19, 20].

In addition, the computational complexity of advanced deep learning and hybrid frameworks presents a significant challenge for real-time deployment. Many high-performing models require substantial processing power and memory, making them difficult to implement in large-scale or resource-constrained environments such as IoT networks [8, 10].

To provide a comprehensive synthesis of these empirical observations, Table 9 presents a comparative evaluation of representative zero-day attack detection studies, highlighting their datasets, techniques, performance outcomes, and identified limitations.

Table 9 provides a comprehensive comparison of representative studies on artificial intelligence-driven zero-day attack detection across different cybersecurity domains. A noticeable observation is that many studies report exceptionally high performance values, with several models achieving accuracy levels above 99%. For example, PCA-enhanced XGBoost and CatBoost models achieved 99.99% accuracy on the UNSW-NB15 dataset, while Autoencoder-RF/XGBoost hybrid models reported 99.98% accuracy on the CIC-MalMem-2022 dataset. Similarly, several deep learning and hybrid frameworks consistently achieved detection rates exceeding 97%. These findings demonstrate the significant progress that intelligent detection techniques have made in identifying both known and previously unseen cyber threats.

Despite these promising results, the reported performance values should be interpreted cautiously. Most of the reviewed studies rely heavily on benchmark datasets such as NSL-KDD, UNSW-NB15, CICIDS2017, ToN-IoT, and CIC-MalMem-2022. While these datasets provide standardized evaluation environments, they may not fully represent the complexity, diversity, and dynamic behaviour of real-world network traffic. Consequently, models trained and tested on the same benchmark datasets may achieve artificially inflated performance due to dataset-specific characteristics rather than true generalization capability.

Another important concern relates to external validation and reproducibility. Several studies report near-perfect accuracy using a single dataset without conducting cross-dataset validation or testing across heterogeneous network environments. This raises questions regarding the robustness of these models when deployed in practical settings characterized by evolving attack patterns, concept drift, encrypted traffic, and changing user behaviours. Models that perform exceptionally well in laboratory conditions may experience substantial performance degradation when exposed to operational network environments.

The comparative evidence also reveals that increasing model complexity generally leads to improved detection performance. Classical machine learning techniques such as Random Forest, Support Vector Machines, and K-Nearest Neighbour achieved strong performance when combined with feature engineering techniques. However, deep learning architectures including CNNs, RNNs, LSTMs, transformer-based models, and attention mechanisms demonstrated superior capability in learning complex behavioural patterns directly from raw network traffic. More recently, hybrid intelligent frameworks integrating anomaly detection, supervised learning, reinforcement learning, and ensemble methods have emerged as the dominant research direction because they combine the strengths of multiple detection paradigms.

Nevertheless, higher detection accuracy often comes at the cost of increased computational complexity. Several studies highlighted challenges associated with training overhead, memory consumption, scalability, and deployment feasibility in resource-constrained environments such as IoT and Industrial IoT systems. This trade-off suggests that future research should not focus solely on maximizing accuracy but should also consider computational efficiency, scalability, explainability, and real-time deployment requirements.

Another recurring limitation observed across the reviewed studies is dataset imbalance and limited attack diversity. Several datasets contain disproportionately represented attack categories, which may bias learning algorithms toward dominant classes while reducing their effectiveness in detecting rare or stealthy attacks. This issue is particularly important in zero-day attack detection because previously unseen attacks often occur infrequently and may not be adequately represented in existing datasets.

Furthermore, explainability remains an important challenge. Although deep learning and hybrid frameworks achieve superior performance, many operate as black-box systems, limiting transparency and reducing trust among cybersecurity analysts. Only a small number of studies incorporated explainable artificial intelligence techniques such as SHAP or other model interpretation mechanisms. Consequently, there is a growing need for intelligent intrusion detection systems that combine high detection performance with transparent and interpretable decision-making processes.

In all, the comparative analysis presented in Table 9 demonstrates that the field is progressing from traditional machine learning approaches toward adaptive, hybrid, and intelligence-driven detection architectures. However, persistent challenges including dataset realism, limited cross-domain validation, reproducibility concerns, adversarial robustness, computational overhead, and lack of explainability indicate that substantial research opportunities remain. Future studies should prioritize realistic datasets, cross-environment evaluation, continual learning mechanisms, explainable artificial intelligence integration, and proactive threat intelligence capabilities to enhance the practical effectiveness of zero-day attack detection systems.

Table 9. *Comprehensive Comparative Evaluation of Reviewed Zero-Day Detection Studies.*

S/N	Domain & Studies	Dataset Source & Description	Methods / Techniques Used	Experimental Results	Research Gap
1	IoT Security [1]	UNSW-NB15 dataset with modern attack categories	PCA + XGBoost, CatBoost, KNN, SVM	99.99% accuracy, 99.97% MCC	Lack of real IoT validation, need for newer datasets
2	Cybersecurity [3]	Heterogeneous telemetry data (logs, packets)	CNN, RNN, Attention, GNN	Reduced detection time by over 30%	Data scarcity, concept drift, high cost, interpretability issues
3	Cybersecurity [26]	NVD, CVE, logs, simulated attacks	RF, SVM, K-Means, Deep Learning	Up to 0.95 AUC, 0.92 TPR	Lack of labelled data, generalization issues
4	Cybersecurity [5]	IoT23 dataset with simulated attacks	RF, Isolation Forest, DNN	~95% accuracy (RF), >90% (DNN)	Dataset imbalance, scalability challenges
5	Smart Systems [27]	NSL-KDD, UNSW-NB15, ToN-IoT	ML + XAI (SHAP), IDPS	94.89% accuracy	Limited real-world datasets, computational complexity
6	Cybersecurity [14]	CICIDS2017, NSL-KDD, UNSW-NB15	BERT, LoRA, Reinforcement Learning	97.8% accuracy, 95.7% detection rate	High computational overhead, adversarial risks
7	Cybersecurity [19]	CIC-MalMem-2022 dataset	Autoencoder + RF/XGBoost	99.98% accuracy, near-perfect scores	Overfitting, limited dataset validation
8	Cybersecurity [2]	Multiple datasets (NSL-KDD, CICIDS2017, etc.)	ML, DL, Hybrid Models	Accuracy above 99% in some models	High false positives, dataset imbalance
9	Cybersecurity [12]	UNSW-NB15 dataset	RF, MLP, Zero-Shot Learning	>98% accuracy	Inconsistent class detection, dataset dependency
10	Cybersecurity [13]	UNSW-NB15, NF-UNSW-NB15-v2	RF, MLP	Up to 92.45% Z-DR	Poor detection of some attack classes
11	IoT Security [17]	CICIDS2017, NSL-KDD	Autoencoder, One-Class SVM	>90% detection accuracy	Limited dataset diversity
12	Cybersecurity [18]	Network logs, event logs, threat intelligence feeds	RF, SVM, K-Means, Isolation Forest, RNN, LSTM, RL	Up to 92% accuracy (LSTM), 85% zero-day detection	Limited labelled data, adversarial risks, high computational cost
13	Cloud Security [20]	CICIDS2017 + synthetic attacks	LSTM + Isolation Forest	98.9% accuracy	Use of synthetic zero-day data
14	IIoT Security [10]	Multiple IDS datasets	ML, DL, Reinforcement Learning	>99% accuracy in many models	Dataset imbalance, poor generalization
15	Cybersecurity [8]	CICIDS2017, UNSW-NB15	Isolation Forest, AE, GNN	LSTM-AE: 92.5% detection rate	High computational cost
16	IoT Security [11]	ToN-IoT dataset	CNN (L1, L2 regularization)	Up to 97.94% accuracy	Limited attack diversity
17	Cybersecurity [6]	UGRansome dataset	Random Forest + Deception	99% accuracy	Limited integration with adaptive models

S/N	Domain & Studies	Dataset Source & Description	Methods / Techniques Used	Experimental Results	Research Gap
18	Cybersecurity [24]	CICIDS, UNSW-NB15 + simulations	Hybrid AI (meta-learning, anomaly detection)	AUC = 0.92, F1 = 84%	Reliance on simulated data
19	Software Security [28]	Binary vulnerability dataset	BiLSTM + GNN + Attention	91.3% accuracy	Limited vulnerability types
20	Cybersecurity [16]	Logs, malware datasets	ML, DL, Clustering	Improved detection over traditional IDS	Lack of transparency, adversarial risks

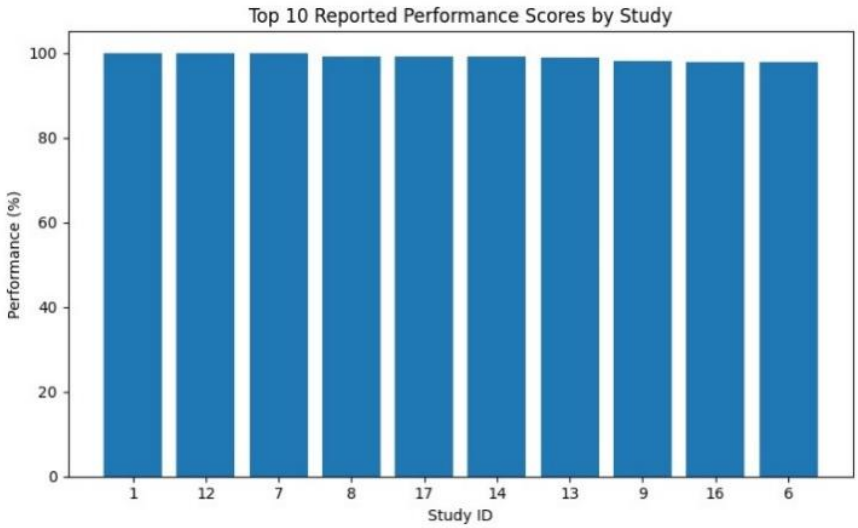

Figure 3. *Top 10 Reported Performance Scores.*

Figure 3 highlights the top ten reported performance scores across the reviewed studies, showing that all selected studies achieved results close to 100%. This strong clustering at the upper range suggests that recent intelligent zero-day detection models, particularly hybrid and deep learning approaches, have reached very high effectiveness under benchmark evaluation settings.

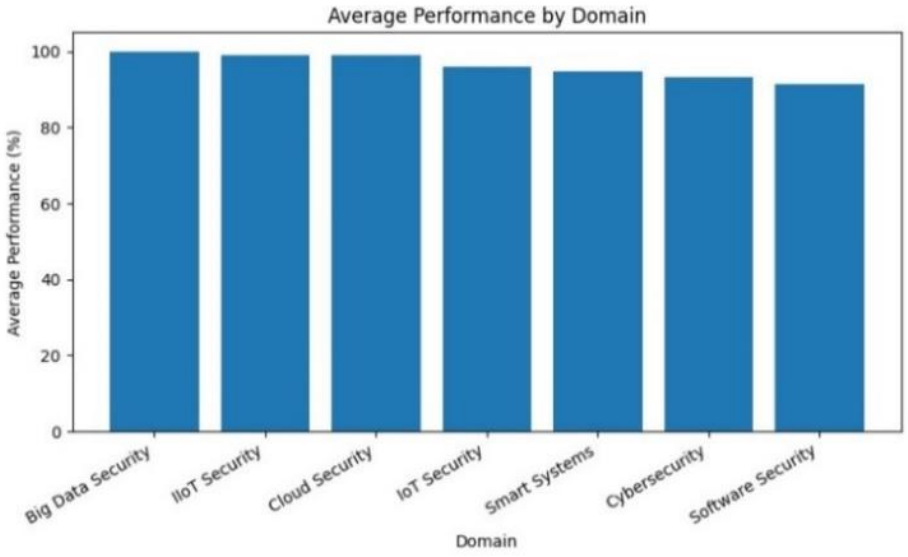

Figure 4. *Average Performance by Domain.*

Figure 4 compares the average performance across application domains and shows consistently high results, with all domains recording values above 90%. Big Data Security, IoT Security, and Cloud Security appear slightly higher, while Cybersecurity and Software Security are marginally lower. The narrow spread suggests that intelligent zero-day detection models maintain strong effectiveness across diverse operational domains.

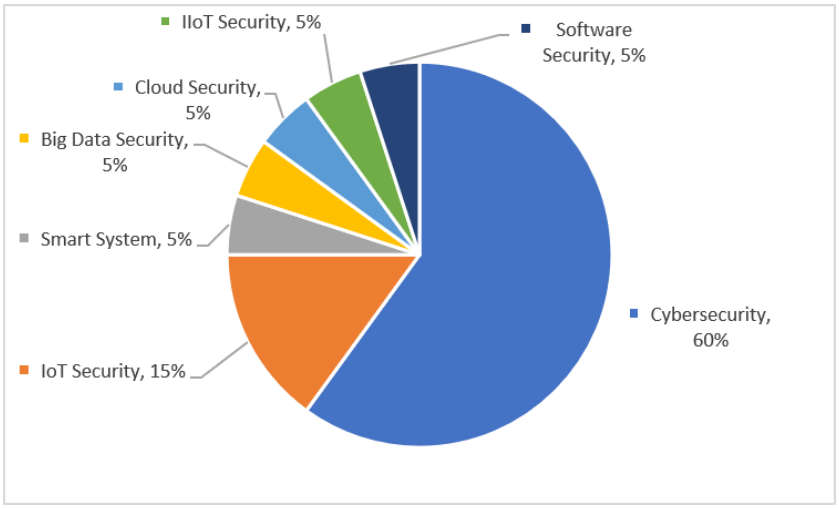

Figure 5. *Distribution of Studies by Domain.*

Figure 5 presents the domain distribution of the reviewed studies and shows that Cybersecurity dominates the literature with 60% of the total studies, indicating that most zero-day attack detection research is still concentrated in general cybersecurity environments. IoT Security accounts for 15%, reflecting growing interest in connected and resource-constrained systems. The remaining domains, including Smart Systems, Big Data Security, Cloud Security, IIoT Security, and Software Security, each contribute 5%, showing emerging but still limited research attention. This distribution suggests that while the field is expanding into specialized domains, broader cross-domain investigation is still needed to improve the generalizability and real-world applicability of intelligent detection frameworks.

4.4. Performance Trends Across Detection Paradigms

Performance outcomes reported across the reviewed studies reveal a clear methodological progression in artificial intelligence driven intrusion detection research. Early studies based on classical machine learning techniques provided the foundation for automated threat detection, with performance largely influenced by feature engineering methods and dataset characteristics [16]. While these approaches achieved strong accuracy in structured benchmarking environments, their ability to model complex behavioural patterns associated with evolving cyber threats remained limited, particularly in detecting previously unseen attacks [13].

The increasing adoption of deep learning architectures has significantly enhanced detection capability through automated hierarchical feature representation learning. Neural models such as convolutional and recurrent networks have demonstrated strong performance by capturing both spatial and temporal dependencies in network traffic data [7]. In addition, advanced architectures, including transformer-based and attention-driven models, have further improved adaptability and robustness in detecting complex and evolving attack patterns [14, 25]. These models reduce reliance on manual feature extraction and improve the ability of detection systems to identify multi-stage intrusions and sophisticated zero-day threats.

Anomaly-based detection approaches have also contributed significantly to performance improvement by modelling normal system behaviour and identifying deviations that may indicate malicious activity. Techniques such as Isolation Forest and autoencoder-based frameworks have shown strong capability in detecting unknown attack patterns without requiring labelled data [5, 17]. These methods are particularly suitable for zero-day detection scenarios, where prior knowledge of attack signatures is unavailable.

More recently, hybrid intelligent detection frameworks that combine supervised learning, anomaly detection, and adaptive decision mechanisms have emerged as dominant research directions. These approaches leverage the strengths of multiple techniques to improve detection accuracy and robustness. For example, hybrid models integrating deep learning with anomaly detection have reported high performance in identifying unseen attacks [19, 20].

However, these performance improvements must be interpreted with caution. Several studies highlight the risk of overfitting, especially when models are evaluated using single datasets or controlled experimental settings. In addition, limited cross-dataset validation reduces confidence in the generalizability of reported results. Furthermore, the increasing complexity of hybrid and deep learning models introduces challenges related to computational cost and real-time deployment [10, 20].

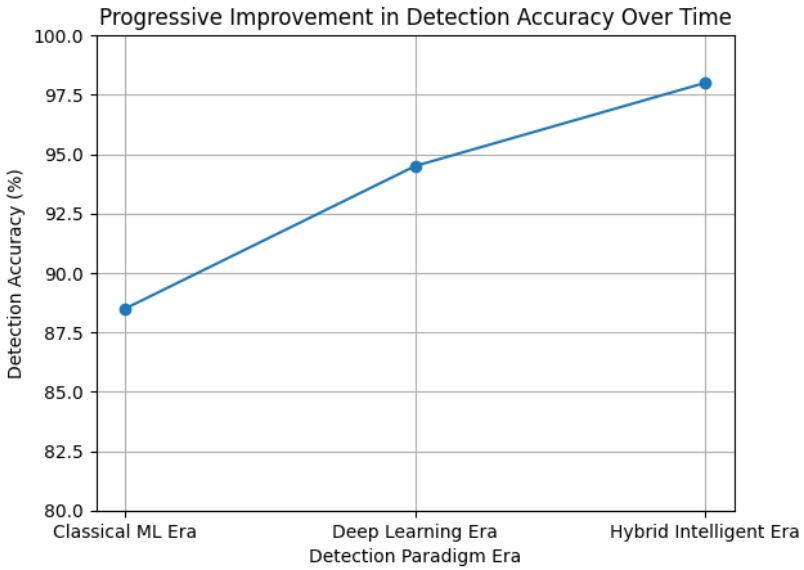

Figure 6. *Progressive Improvement in Detection Accuracy Across Detection Paradigms.*

Figure 6 clearly illustrates how intrusion detection performance has progressed as the underlying methodologies became increasingly sophisticated. Early classical machine learning approaches generally achieved detection accuracy within the range of approximately 85% to 92%, reflecting the effectiveness of feature-driven statistical learning under structured benchmark conditions. With the emergence of deep learning architectures, reported performance improved further to roughly 92% to 97%, largely driven by automated feature representation learning and improved modelling of temporal and behavioural dependencies within network traffic. More recent hybrid intelligent detection frameworks, particularly those integrating anomaly detection with ensemble learning strategies, have pushed performance even further, frequently reporting detection accuracy between 97% and 99%. This progression reflects the growing effectiveness of adaptive, multi-paradigm cybersecurity analytics in addressing the complexity of zero-day attack detection, particularly in relation to evolving and previously unseen threat behaviours.

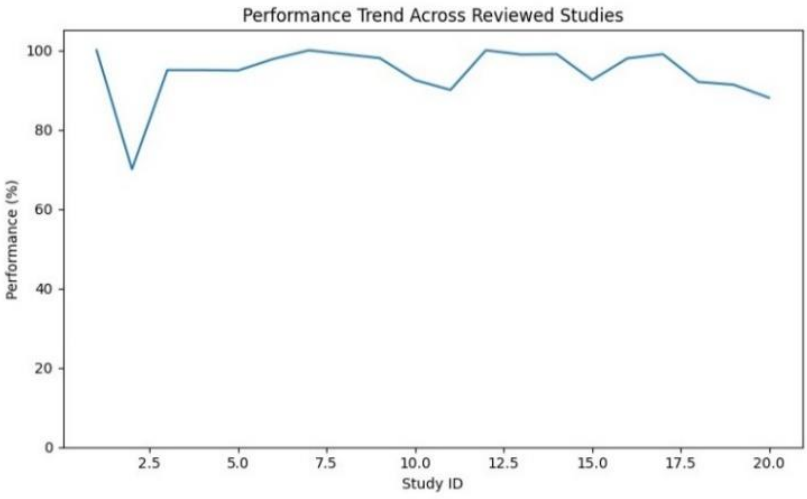

Figure 7. *Performance Trend Across Reviewed Studies.*

Figure 7 illustrates the performance trend observed across the reviewed studies, showing that most reported results remain consistently high within the 90% to 100% range. Although there is a noticeable decline at the early stage of the curve, the trend quickly recovers and maintains a stable high-performance pattern across most of the subsequent studies, with only minor fluctuations at a few points. This behaviour reflects the steady advancement of zero-day attack detection approaches from traditional machine learning methods to more robust deep learning and hybrid intelligent frameworks. The slight dips along the curve may be associated with differences in dataset complexity, model generalization, or variations in evaluation conditions.

4.5. Cross-Domain Generalization Challenges

A recurring limitation identified across the reviewed intrusion detection studies is the limited emphasis on cross-domain validation during model development and evaluation. Many detection frameworks are trained and evaluated using a single benchmark dataset, which often results in overly optimistic performance outcomes that may not accurately reflect real-world deployment conditions [2]. This reliance on single-source datasets reduces the ability of models to generalize effectively across diverse network environments.

Variations in traffic distribution, protocol behaviour, and attack patterns across different domains significantly influence detection reliability. For example, models trained on enterprise network datasets may experience performance degradation when applied to industrial or Internet of Things environments, where communication patterns are more heterogeneous and system constraints are more pronounced [11]. These differences highlight the limitations of dataset-specific optimization and the need for more generalized detection frameworks.

Furthermore, the increasing integration of cyber-physical systems, cloud infrastructures, and edge computing environments introduces additional complexity in model evaluation. These environments generate diverse and dynamic traffic patterns, making it difficult for models trained on static datasets to maintain consistent performance. As a result, multi-dataset validation and domain adaptation strategies are essential for improving the robustness and scalability of intelligent intrusion detection systems [10].

4.6. Adversarial Vulnerability and Model Robustness

Despite significant improvements in detection accuracy, artificial intelligence driven intrusion detection systems remain vulnerable to adversarial manipulation. Attackers can deliberately craft malicious traffic that closely resembles normal behaviour, thereby bypassing detection mechanisms and reducing system effectiveness in dynamic threat environments [20].

Deep learning-based detection models, although powerful in capturing complex behavioural patterns, are particularly susceptible to performance degradation when exposed to adversarial inputs or distributional shifts. Studies have shown that even high-performing models may struggle to maintain reliability under evolving attack conditions, especially when trained on limited or static datasets [10]. This challenge emphasizes the importance of developing robust detection frameworks that incorporate adaptive learning, continuous retraining, and ensemble decision strategies.

Another critical concern is the lack of interpretability in many advanced detection models. Complex neural architectures often function as black-box systems, making it difficult for cybersecurity analysts to understand how decisions are made. This lack of transparency can reduce trust in automated systems and delay incident response processes, ultimately affecting the operational effectiveness of intelligent security platforms [16].

4.7. Overall Interpretation of Performance Evolution

The cumulative evidence from the reviewed literature indicates that intrusion detection performance has improved steadily with increasing methodological sophistication. Classical machine learning approaches provided the initial foundation for automated threat detection, with performance largely dependent on feature engineering and dataset characteristics [1, 2].

The introduction of deep learning techniques significantly enhanced detection capability by enabling automated feature extraction and improved modelling of complex behavioural patterns. These models have demonstrated strong performance across various benchmark datasets, particularly in detecting sophisticated and multi-stage attack scenarios [7].

More recently, hybrid intelligent detection frameworks that combine supervised learning, anomaly detection, and adaptive decision mechanisms have shown greater effectiveness in identifying previously unseen attacks while reducing false alarm rates [10]. These approaches leverage the strengths of multiple techniques to improve detection robustness and adaptability.

However, this performance improvements remain closely tied to dataset characteristics, evaluation methodologies, and deployment environments. Differences in attack diversity, traffic dynamics, and system architecture continue to influence detection outcomes. As a result, there is a growing need for scalable, adaptive, and generalizable intrusion detection systems capable of maintaining consistent performance across heterogeneous cybersecurity environments [11]. Addressing these challenges is essential for transitioning intrusion detection research from controlled experimental evaluation toward practical, real-world cybersecurity deployment.

4.8. Proposed Framework Positioning

The reviewed studies demonstrate that artificial intelligence techniques have significantly improved zero-day attack detection through machine learning, deep learning, and hybrid approaches. Machine learning models such as Random Forest, Support Vector Machines, and ensemble methods provide strong baseline performance, while deep learning architectures including convolutional, recurrent, and transformer-based models improve detection of complex attack patterns [1, 14]. However, despite these advancements, several limitations persist. Many models rely heavily on benchmark datasets, limiting their ability to generalize across real-world environments. In addition, high computational cost, dataset imbalance, and vulnerability to adversarial attacks continue to affect detection performance and deployment feasibility [10, 20].

To address these challenges, this study positions a hybrid and adaptive detection framework that integrates multiple complementary techniques within a unified system. The framework combines supervised learning for classification, anomaly detection for identifying unknown threats, and continuous learning mechanisms to adapt to evolving attack patterns [5, 17]. The framework also emphasizes the use of heterogeneous datasets, including IoT, enterprise, and cloud-based traffic, to improve generalization and robustness across different environments [7, 11].

In addition, anomaly-based techniques such as autoencoders and Isolation Forest are incorporated to detect previously unseen attacks by modelling normal behaviour and identifying deviations [8, 19]. The proposed framework further integrates deception-based intelligence and threat monitoring strategies to enhance proactive detection capability and improve system adaptability to emerging threats [6].

To improve transparency and trust, the framework supports explainable artificial intelligence techniques, enabling security analysts to understand model decisions and respond effectively to detected threats [27]. Finally, the framework is designed to support scalability and efficient deployment in real-world environments, including resource-constrained systems such as IoT networks, by balancing detection accuracy with computational efficiency [10, 14]. As a way forward, the proposed conceptual framework is depicted in Figure 8.

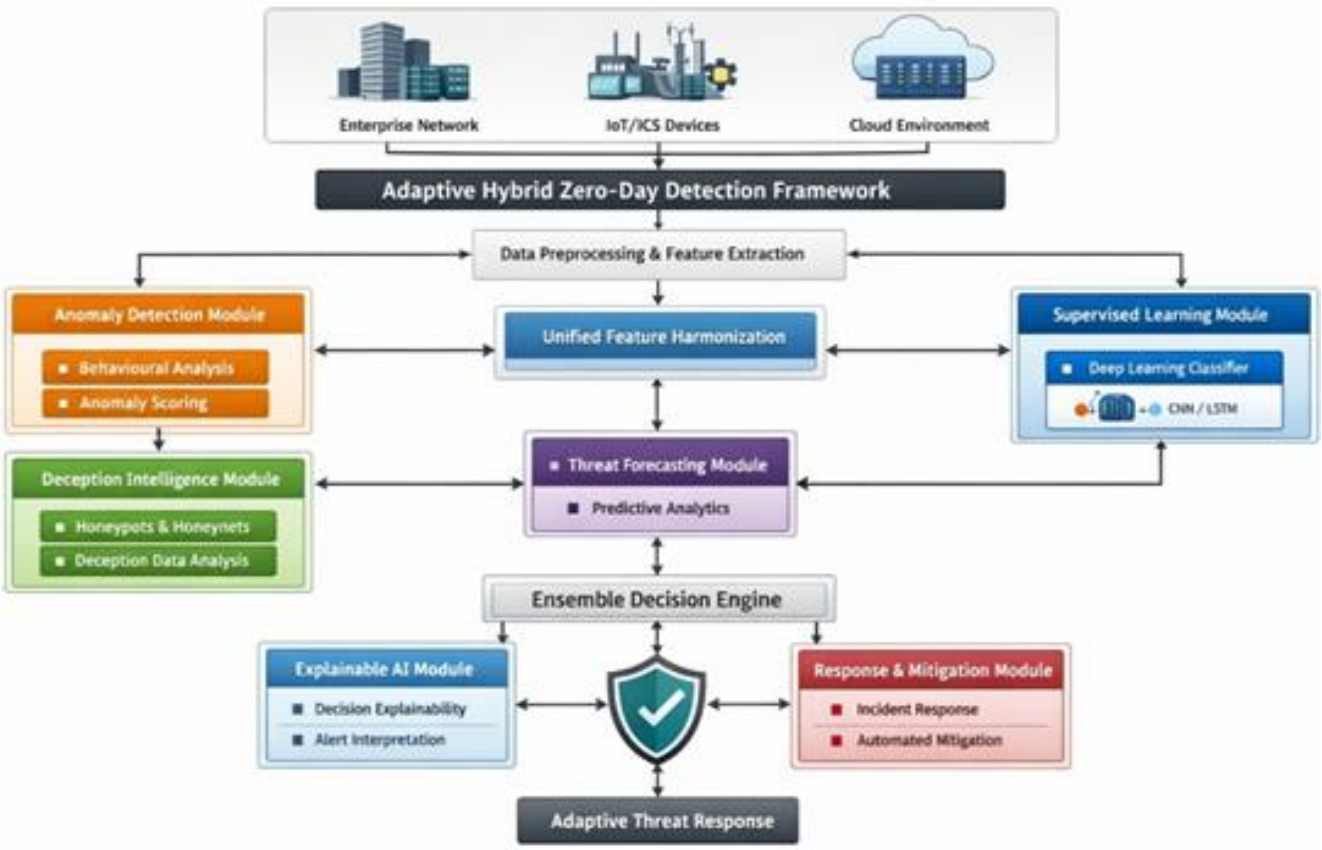

Figure 8. *Conceptual Framework of the Adaptive Hybrid Zero Day Detection Framework.*

Figure 8 illustrates the conceptual framework of the proposed adaptive hybrid zero-day detection framework developed to enhance intelligent cyber defence across heterogeneous network environments. The framework begins by acquiring multi-source traffic data from enterprise networks, industrial Internet of Things systems, and cloud computing infrastructures. This diverse input enables the detection system to

learn behavioural representations that reflect real operational threat dynamics rather than relying solely on benchmark datasets.

Following data acquisition, preprocessing and unified feature harmonization modules transform heterogeneous traffic streams into standardized behavioural representations suitable for intelligent analysis. This stage supports cross domain learning and reduces performance degradation commonly associated with dataset distribution shifts.

The analytical core of the framework integrates anomaly detection mechanisms, supervised deep learning classifiers, deception driven intelligence acquisition, and predictive threat forecasting capability. Anomaly detection modules continuously monitor deviations from normal behavioural patterns, enabling early identification of unknown attack activities. Supervised classifiers enhance decision precision by distinguishing malicious traffic categories using hierarchical feature learning strategies.

Deception intelligence components enrich the detection pipeline by capturing realistic attacker interaction traces from controlled honeypot environments. These behavioural insights improve adaptive learning capability and strengthen resilience against evolving cyber threats. The inclusion of temporal forecasting mechanisms further supports proactive threat anticipation by identifying early warning signals associated with multi-stage intrusion campaigns.

Outputs from these analytical modules are combined through an ensemble decision fusion engine designed to improve detection robustness and reduce false alarm rates. Explainable analytics modules provide interpretable insights into detected threat behaviours, thereby supporting analyst trust and facilitating effective incident response decision making.

Finally, the framework incorporates automated mitigation and adaptive learning feedback loops that enable continuous model updating and operational scalability. This integrated design supports the transition of intelligent intrusion detection research from experimental benchmarking toward real world cyber defence deployment.

4.9. Comparative Evaluation of the Proposed Adaptive Hybrid Framework Against Existing Hybrid IDS Architectures

The comparative analysis of the reviewed studies reveals a progressive evolution in intrusion detection methodologies, ranging from traditional machine learning approaches to more sophisticated deep learning, anomaly detection, and hybrid intelligent intrusion detection systems. Traditional machine learning-based detection models rely heavily on handcrafted feature engineering and labelled datasets, which limits their ability to adapt to rapidly evolving cyber threat behaviours. Although these approaches offer interpretable outputs and relatively low computational requirements, their effectiveness in detecting previously unseen attacks remains constrained, particularly in highly dynamic cybersecurity environments [1, 2].

Deep learning-based intrusion detection architectures have significantly improved detection capability by enabling automatic feature extraction and the modelling of complex behavioural patterns within network traffic. These models have demonstrated strong performance across widely adopted benchmark datasets such as UNSW-NB15 and CICIDS2017, particularly in identifying sophisticated, multi-stage, and evolving attack scenarios [8, 14]. However, their deployment is often challenged by high computational requirements, reduced transparency, and limited interpretability, which may restrict their practical applicability in real-time cybersecurity operations [7, 10].

Anomaly-based detection approaches have further enhanced zero-day attack detection by focusing on deviations from normal system behaviour rather than relying on predefined attack signatures. This enables the identification of previously unknown threats and makes such approaches particularly valuable in dynamic network environments [5, 17]. Nevertheless, anomaly detection systems frequently suffer from elevated false positive rates and often require complementary classification mechanisms to distinguish malicious activities from legitimate behavioural variations accurately [8, 18].

To overcome the limitations of individual approaches, hybrid intelligent intrusion detection systems have emerged as a dominant research direction. By combining supervised learning and anomaly detection techniques, these systems achieve improved detection accuracy, robustness, and adaptability in identifying both known and unknown threats. Despite these advantages, many existing hybrid IDS architectures remain dependent on benchmark datasets, face scalability challenges, and frequently exhibit limited capability to adapt continuously to evolving threat landscapes [10, 19].

In response to these limitations, the proposed adaptive hybrid framework introduces a more comprehensive and intelligence-driven approach to zero-day attack detection. The framework integrates multiple complementary components within a unified architecture, including multi-source data acquisition, anomaly detection, supervised classification, adaptive learning, and intelligent threat analysis. The integration of these components enhances detection reliability across heterogeneous network environments and improves the framework's ability to respond to emerging cyber threats [7, 11].

Table 10 presents a comparative evaluation of the proposed framework against conventional hybrid IDS architectures reported in the literature. While existing hybrid systems primarily focus on improving detection performance through combinations of machine learning, deep learning, and anomaly detection techniques, most remain largely reactive, identifying malicious activities only after they have occurred. In contrast, the proposed framework incorporates several advanced capabilities designed to support proactive, adaptive, and explainable cyber defence.

A key distinguishing feature is the integration of deception-based threat intelligence through honeypot-generated attack data. This capability enables the collection of realistic attacker

behaviours and provides valuable threat intelligence for identifying emerging attack strategies. Such functionality is absent or only partially represented in many existing hybrid IDS architectures, thereby limiting their ability to perform proactive threat analysis.

Another important advancement is the incorporation of threat forecasting capabilities. Unlike traditional detection-centric systems, the proposed framework leverages behavioural analytics and predictive intelligence to identify emerging attack patterns before exploitation occurs. This proactive capability strengthens cyber defence preparedness and supports earlier intervention against potential zero-day threats.

The framework also incorporates continual learning mechanisms that enable adaptation to evolving attack behaviours, changing network conditions, and concept drift. This addresses one of the most significant limitations of static intrusion detection models, whose performance often deteriorates when confronted with previously unseen attack patterns or changing operational environments [6]. Through continuous model refinement, the framework is designed to maintain detection effectiveness over extended deployment periods.

Furthermore, explainable artificial intelligence (XAI) is integrated into the framework to improve transparency and enhance analyst confidence in automated detection decisions. While many existing hybrid IDS models operate as black-box systems, the proposed framework provides interpretable insights into model outputs and detection outcomes, thereby improving accountability, operational usability, and decision support.

An additional distinguishing capability is the inclusion of an adaptive feedback loop that continuously updates detection knowledge using newly observed threat intelligence and operational outcomes. This feedback-driven adaptation mechanism enables sustained effectiveness in dynamic cybersecurity environments and enhances resilience against sophisticated and evolving zero-day attacks.

The comparative evaluation presented in Table 10 demonstrates that the proposed adaptive hybrid framework extends beyond conventional hybrid intrusion detection architectures through the integration of deception-based intelligence, predictive threat forecasting, continual learning, explainable artificial intelligence, and adaptive feedback mechanisms. By combining reactive detection capabilities with proactive threat anticipation and continuous adaptation, the framework provides a more scalable, interpretable, and resilient cyber defence strategy. These characteristics position the framework as a comprehensive solution for addressing the growing complexity of modern zero-day attack detection and enhancing the practical deployment of intelligent intrusion detection systems in real-world cybersecurity environments.

Table 10. *Comparative Analysis of Conventional Hybrid IDS Architectures and the Proposed Adaptive Intelligent Intrusion Detection Framework.*

Feature	Traditional Hybrid IDS	Proposed Framework
Supervised Learning	Yes	Yes
Anomaly Detection	Yes	Yes
Honeypot Intelligence	Limited	Yes
Threat Forecasting	No	Yes
Continual Learning	Limited	Yes
Explainable AI	Limited	Yes
Adaptive Feedback Loop	Rare	Yes

5. Conclusions

This systematic review has examined the evolution of artificial intelligence driven approaches for detecting zero-day cyber-attacks, highlighting methodological advancements, dataset utilization patterns, performance trends, and persistent research challenges. The findings indicate that intrusion detection capability has improved progressively from classical machine learning models toward deep learning and hybrid intelligent detection frameworks capable of modelling complex behavioural patterns within network traffic. However, the review also reveals that significant limitations remain related to dataset realism, cross domain generalization, computational scalability, adversarial robustness, and explainability. These challenges continue to influence the practical deployment readiness of intelligent intrusion detection systems across modern digital infrastructures characterized by increasing heterogeneity and threat sophistication.

To address these issues, the study positions an adaptive hybrid detection framework that integrates anomaly monitoring, supervised learning, deception driven intelligence acquisition,

predictive threat forecasting, and explainable decision support within a unified cyber defence architecture. Such integration represents a strategic direction for advancing intelligent cybersecurity analytics from experimental benchmarking toward resilient real world operational deployment.

Future research must therefore prioritize the development of dynamic cybersecurity datasets, scalable learning pipelines, interpretable detection mechanisms, and proactive threat anticipation strategies capable of sustaining performance in evolving cyber threat environments. By synthesizing existing knowledge and identifying critical research pathways, this review contributes toward guiding the next generation of intelligent cyber defence innovations.

6. Future Research Directions

The increasing complexity of cyber threats and the limitations identified across existing intelligent intrusion detection studies highlight several important directions for future research. As cybersecurity systems continue to expand across cloud computing environments, Industrial Internet of Things infrastructures, and distributed networks, detection frameworks must emphasize adaptability, scalability, and contextual awareness to sustain effective performance. One critical research direction involves the development of realistic and continuously updated cybersecurity datasets capable of capturing emerging attack behaviours and diverse traffic patterns. Existing studies have shown that widely used datasets such as UNSW-NB15 and CICIDS2017 support model development and evaluation, but their static nature limits long-term generalization capability [2, 10].

Future research should therefore focus on incorporating more heterogeneous and continuously evolving data sources, including real-world network traffic and updated threat intelligence. Another important direction is the advancement of cross-domain detection approaches that can maintain performance across different operational environments. Several studies highlight that models trained on specific datasets often struggle to generalize across new network conditions due to differences in traffic distribution and attack characteristics [13, 29]. Research efforts should therefore explore domain adaptation techniques, federated learning approaches, and multi-source training frameworks to improve generalization capability.

Improving adversarial robustness also remains a major research challenge. Existing studies have shown that artificial intelligence-based intrusion detection systems can be vulnerable to adversarial manipulation, where attackers exploit model weaknesses to evade detection [16, 26]. Future frameworks should incorporate robust learning strategies, adaptive model updating, and hybrid detection mechanisms to reduce susceptibility to adversarial attacks.

Explainable artificial intelligence represents another important research priority. Although deep learning models have demonstrated strong detection performance, their lack of interpretability can reduce trust and limit their use in practical cybersecurity operations [7, 18]. Future work should therefore focus on developing transparent detection models that provide clear explanations for detected threats and support effective decision-making. Furthermore, integrating predictive threat detection capabilities into intrusion detection systems presents an important opportunity for advancing proactive cybersecurity. Studies have shown that analysing behavioural patterns and anomaly trends can improve early detection of emerging threats and support timely response strategies [3, 6].

Finally, future research should explore the integration of artificial intelligence with automated mitigation and adaptive defence mechanisms. Several studies emphasize the need for hybrid frameworks that combine detection, prediction, and response capabilities to enhance overall system resilience [4, 24]. Continuous learning, modular system design, and resource-efficient implementation will be essential for supporting scalable deployment in real-world cybersecurity environments.

7. Limitations of the Review

Despite providing a comprehensive synthesis of artificial intelligence driven zero-day attack detection research, several limitations should be acknowledged. First, the review primarily focuses on studies evaluated using publicly available benchmark cybersecurity datasets. Although such datasets support comparative analysis, their controlled experimental conditions may not fully reflect the complexity and variability of real-world network environments. This limitation has been highlighted in several studies, which emphasize the need for more realistic and heterogeneous datasets to improve generalization and real-world applicability [10, 18].

Second, variations in evaluation metrics and experimental design across reviewed studies present challenges for direct performance comparison. Differences in preprocessing techniques, feature engineering approaches, model configurations, and validation strategies can contribute to performance variability that is not solely attributable to the underlying detection methodologies. Such inconsistencies have been observed in comparative studies, limiting the ability to establish standardized performance benchmarks [2, 8].

Third, the review emphasizes algorithmic and architectural developments in artificial intelligence driven cybersecurity research but provides limited analysis of real-world deployment scenarios. Several studies note that practical challenges such as computational overhead, scalability constraints, and integration with operational environments remain significant barriers to implementation [10, 14]. In addition, the rapidly evolving nature of cyber threats means that some emerging detection techniques and datasets may not yet be fully represented in current literature. Many studies highlight the continuous evolution of attack patterns and the need for adaptive and continuously updated detection systems, indicating inherent

temporal limitations in capturing a dynamic research landscape [3, 16].

Finally, although this review proposes a conceptual adaptive hybrid detection framework based on identified research gaps, the framework has not been empirically validated within the scope of the study. Several studies emphasize the importance of cross-domain validation and real-world testing to ensure robustness and generalization, which remain areas for future investigation [13, 24].

Abbreviations

AI	Artificial Intelligence
IDS	Intrusion Detection System
IDPS	Intrusion Detection and Prevention System
ML	Machine Learning
DL	Deep Learning
RF	Random Forest
SVM	Support Vector Machine
KNN	K-Nearest Neighbour
PCA	Principal Component Analysis
DNN	Deep Neural Network
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
GNN	Graph Neural Network
AE	Autoencoder
SHAP	SHapley Additive exPlanations
XAI	Explainable Artificial Intelligence
GAN	Generative Adversarial Network
MLP	Multilayer Perceptron
LoRA	Low-Rank Adaptation
BERT	Bidirectional Encoder Representations from Transformers
AUC	Area Under the Curve
MCC	Matthews Correlation Coefficient
TPR	True Positive Rate
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
CVE	Common Vulnerabilities and Exposures
NVD	National Vulnerability Database
IoT	Internet of Things
IIoT	Industrial Internet of Things
ToN-IoT	Telemetry and Network Traffic for Internet of Things
NSL-KDD	Network Security Laboratory Knowledge Discovery in Databases
UNSW-NB15	University of New South Wales Network Benchmark Dataset 2015
CICIDS2017	Canadian Institute for Cybersecurity Intrusion Detection System Dataset 2017
Edge-IIoT	Edge Industrial Internet of Things
CIC-MalMem-2022	Canadian Institute for Cybersecurity Malware Memory Dataset 2022
UGRansome	University of Granada Ransomware Dataset
CSIC	Consejo Superior De Investigaciones Cient ficas
DT	Decision Tree
SQL	Structured Query Language
XGBoost	Extreme Gradient Boosting
CatBoost	Categorical Boosting
CVSS	Common Vulnerability Scoring System
DoS	Denial of Service
DDoS	Distributed Denial of Service
SIEM	Security Information and Event Management
SOC	Security Operations Center
NLP	Natural Language Processing

ANN	Artificial Neural Network
RFE	Recursive Feature Elimination
ROC	Receiver Operating Characteristic
F1-Score	Harmonic Mean of Precision and Recall

Author Contributions

Uchechukwu Samuel Nwankwo: Conceptualization, Resources, Writing – review & editing, Data curation, Writing – original draft

Obi Chukwuemeka Nwokonkwo: Investigation, Methodology, Supervision, Validation

Charles Ikerionwu: Formal Analysis, Investigation, Methodology, Supervision, Validation

Adetokunbo MacGregor John-Otumu: Conceptualization, Data curation, Investigation, Methodology, Validation

Udoka Felista Eze: Formal Analysis, Supervision

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Saheed, Y. K., Abiodun, A. I., Misra, S., Holone, M. K., & Colomo-Palacios, R. "A machine learning-based intrusion detection for detecting internet of things network attacks", *Alexandria Engineering Journal*, vol. 61, pp. 9395–9409, 2022. https://doi.org/10.1016/j.aej.2022.02.063

[2] Ali, S., Rehman, S. U., Imran, A., Adeem, G., Iqbal, Z., & Kim, K. I. "Comparative evaluation of AI-based techniques for zero-day attacks detection", *Electronics*, vol. 11, no. 23, 2022. https://doi.org/10.3390/electronics11233934

[3] Kang, G. "Artificial intelligence for threat detection: Leveraging deep learning to identify zero-day attacks in real time", *Universal Library of Engineering Technology*, vol. 2, no. 4, pp. 74–79, 2025. https://doi.org/10.70315/uloap.ulete.2025.0204013

[4] Hamid, K., Iqbal, M. W., Aqeel, M., Liu, X., & Arif, M. "Analysis of techniques for detection and removal of zero-day attacks", in *Ubiquitous Security (UbiSec 2022)*, CCIS 1768, pp. 248–262, 2023. https://doi.org/10.1007/978-981-99-0272-9_17

[5] Mathew, A. M. "ML-based zero-day attack detection", MSc Research Project, National College of Ireland, 2024.

[6] Cosmos, N. O., & Aimufua, G. "Application of diverse techniques for zero-day management: A review approach", *International Journal of Research and Innovation in Applied Science*, vol. 10, no. 5, pp. 1436–1452, 2025. https://doi.org/10.51584/IJRIAS.2025.1005000126

[7] Whitman, J., El-Karim, A., Nandakumar, P., Ortega, F., & Zheng, L. "AI for zero-day exploit detection in web applications", Unpublished manuscript, 2024.

[8] Micheal, D. "Benchmarking AI-powered anomaly detection techniques for zero-day attacks", Unpublished manuscript, 2024.

[9] Julakanti, S. R. "Computational intelligence approaches for analysis of the detection of zero-day attacks", *UW Journal of Science and Technology*, vol. 6, pp. 27–36, 2025.

[10] Hashim, K. A., Yussoff, Y. B. M., & Shahbudin, S. B. "Mitigating zero-day vulnerabilities in IIoT systems: Challenges and advances in AI-powered intrusion detection systems", *Mesopotamian Journal of Cybersecurity*, vol. 5, no. 3, pp. 1184–1198, 2025. https://doi.org/10.58496//MJCS/2025/063

[11] Hairab, B. I., Aslan, H. K., Elsayed, M. S., Jurcut, A. D., & Azer, M. A. "Anomaly detection of zero-day attacks based on CNN and regularization techniques", *Electronics*, vol. 12, no. 3, 2023. https://doi.org/10.3390/electronics12030573

[12] Kansal, S. "Utilizing deep learning techniques for effective zero-day attack detection", *Economic Sciences*, vol. 25, no. 1, pp. 246–257, 2025.

[13] Sarhan, M., Layeghy, S., Gallagher, M., & Portmann, M. "From zero-shot machine learning to zero-day attack detection", *arXiv preprint*, 2021. https://doi.org/10.1007/s10207-023-00676-0

[14] Hussain, A., Tordera, E. M., Masip-Bruin, X., and Leligou, H. C. "Rule-Based with Machine Learning IDS for DDoS Attack Detection in Cyber-Physical Production Systems (CPPS)", *IEEE Access*, vol. 12, pp. 114894–114905, 2024. https://doi.org/10.1109/ACCESS.2024.3445261

[15] Alsuwaiket, M. A. "ZeroDay-LLM: A large language model framework for zero-day threat detection in cybersecurity", *Information*, vol. 16, no. 11, 2025. pp. 1 - 46. https://doi.org/10.3390/info16110939

[16] James, F. "The role of AI in zero-day attack detection and mitigation", Unpublished manuscript, 2025.

[17] SakthiMurugan, S., Sanjay, K. A., Vishnu, V., & Santhi, P. "Assessment of zero-day vulnerability using machine learning approach", *EAI Endorsed Transactions on Internet of Things*, vol. 10, pp. 1–6, 2024. https://doi.org/10.4108/eetiot.4978

[18] Larry, J., and Smith, J. "AI-Augmented Threat Intelligence for Zero-Day Vulnerability Detection". International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence. Vol. 16, no. 1. pp. 84–97, 2025.

[19] Dai, Z., Por, L. Y., Chen, Y. L., Yang, J., Ku, C. S., Alizadehsani, R., & Pławiak, P. "An intrusion detection model to detect zero-day attacks in unseen data using machine learning", *PLOS ONE*, vol. 19, no. 9, 2024. https://doi.org/10.1371/journal.pone.0308469

[20] Benjamin, G., & Brandon, S. "Hybrid AI models for zero-day threat identification in distributed cloud architectures", *Journal of Advanced Research in Artificial Intelligence and Cybersecurity*, vol. 32, pp. 1–6, 2024.

[21] Zhang, W., Dong, N., Choi, T., Bai, G., & Ko, R. K. L. "Intelligent data refinement and analysis of real-world cyber-attacks on SCADA systems", in *Proceedings of ACM E-Energy Conference*, pp. 846–852, 2025.

[22] Brunner, C., Ko, A., & Fodor, S. (2021). An Autoencoder-Enhanced Stacking Neural Network Model for Increasing the Performance of Intrusion Detection. Journal of Artificial Intelligence and Soft Computing Research. 12. 149-163. https://doi.org/10.2478/jaiscr-2022-0010

[23] Krishnan, D., Singh, S., & Sugumaran, V. "Explainable AI for zero-day attack detection in IoT networks using attention fusion model", *Research Square*, 2025. https://doi.org/10.21203/rs.3.rs-5436116/v1

[24] Chhillar, K., Upadhyay, S., Tomar, D., & Singh, S. "An artificial intelligence-based predictive model for zero-day vulnerability detection", *International Journal of Computer Techniques*, vol. 12, no. 5, pp. 741–748, 2025.

[25] Berrios Vasquez, S. I., Hermosilla Monckton, P. A., Leiva Muñoz, D. I., & Allende, H. Zero-Day Threat Mitigation via Deep Learning in Cloud Environments. *Appl. Sci.* 2025, *15*, 7885. https://doi.org/10.3390/app15147885

[26] Rafy, A. L., Rahman, M. M., Nahar, S., Gony, M. N., & Bhuiyan, M. I. H. "The role of machine learning in predicting zero-day vulnerabilities", *International Journal of Science and Research Archive*, vol. 10, no. 1, pp. 1197–1208, 2023. https://doi.org/10.30574/ijsra.2023.10.1.0838

[27] Sayduzzaman, M., Tamanna, J. T., Kundu, D., & Rahman, T. "Interoperability and explicable AI-based zero-day attacks detection process in smart community", *arXiv preprint*, 2024.

[28] Wei, S. "Adaptive multi-scale feature extraction for zero-day vulnerability detection in system binaries", *Computer Science Bulletin*, vol. 8, no. 1, pp. 203–221, 2025. https://doi.org/10.71465/csb159

[29] Ali, M., Haque, M., Durad, M. H., Usman, A., Mohsin, S. M., Mujlid, H., & Maple, C. "Effective network intrusion detection using stacking-based ensemble approach", *International Journal of Information Security*, vol. 22, pp. 1781–1798, 2023. https://doi.org/10.1007/s10207-023-00718-7

[30] Adeniji, O. D., & Olatunji, O. O. "Zero-day attack prediction with parametric settings using bi-directional recurrent neural networks in cybersecurity", *International Journal of Computer Science and Information Security*, vol. 18, no. 3, pp. 111–118, 2020.

Research Field

Uchechukwu Samuel Nwankwo: Cybersecurity, Machine Learning, Expert Systems

Obi Chukwuemeka Nwokonkwo: IT Project Management, Digital System, IT Automation

Charles Ikerionwu: Software Process Models, Machine Learning, Deep Learning, Artificial Intelligence

Adetokunbo MacGregor John-Otumu: Machine Learning, Deep Learning, Computer Vision, NLP, Intelligent Systems

Udoka Feliata Eze: Database Systems, Data Mining, Data Science, Artificial Intelligence