IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 24 July 2024, accepted 2 September 2024, date of publication 6 September 2024, date of current version 11 October 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3455410*

SURVEY

# A Systematic Literature Review on AI-Based Methods and Challenges in Detecting Zero-Day Attacks

**LIP YEE POR[1], (Senior Member, IEEE), ZHEN DAI[1], SIEW JUAN LEEM[1], YI CHEN[1], JING YANG[1], (Graduate Student Member, IEEE), FARID BINBESHR[2], KOO YUEN PHAN[3], AND CHIN SOON KU[3]**

[1]Department of Computer System and Technology, Faculty of Computer Science and Information Technology, Universiti Malaya, Kuala Lumpur 50603, Malaysia
[2]Interdisciplinary Research Center for Intelligent Secure Systems, King Fahd University of Petroleum and Minerals, Dhahran, Eastern Province 31261, Saudi Arabia
[3]Department of Computer Science, Universiti Tunku Abdul Rahman, Kampar 31900, Malaysia

Corresponding authors: Lip Yee Por (porlip@um.edu.my) and Chin Soon Ku (kucs@utar.edu.my)

This work was supported by the KW IPPP (Research Maintenance Fee) Individual/Centre/Group at Universiti Malaya, Malaysia, under Grant RMF1506-2021.

**ABSTRACT** The detection of zero-day attacks remains one of the most critical challenges in cybersecurity. This systematic literature review focuses on the various AI-based methods employed for detecting zero-day attacks, identifying both the strengths and weaknesses of these approaches. By critically evaluating existing literature, this review provides new insights and highlights the gaps that future research must address. The findings suggest that while artificial intelligence, particularly machine learning, offers promising solutions, there are significant challenges related to data availability, algorithmic complexity, and real-time application. This review contributes to the field by providing a comprehensive analysis of current AI-driven methods and proposing future research directions to enhance zero-day attack detection.

**INDEX TERMS** Zero-day attack, CrowdStrike, intrusion detection, anomaly detection, machine learning, artificial intelligence, cybersecurity.

## I. INTRODUCTION

Cybersecurity encompasses technologies and strategies aimed at protecting computer networks, systems, and related infrastructure from threats that could compromise their confidentiality, integrity, and availability. Recent reports indicate a surge in cybercrime within organizations, with zero-day attacks emerging as a particularly significant concern. A zero-day attack exploits unknown or undisclosed vulnerabilities in software, enabling attackers to launch exploits before any fix is available. This creates a critical window where developers have zero days to address the flaw, hence the term ''zero-day.''

The incident involving CrowdStrike on July 19, 2024, underscored the urgency of addressing zero-day attacks. During this event, CrowdStrike's advanced detection systems identified and neutralized a sophisticated zero-day attack targeting a large financial institution. This specific attack involved a previously unknown vulnerability in widely used enterprise software, potentially compromising sensitive data. CrowdStrike's swift response highlighted the growing complexity of zero-day attacks and the critical need for continuous advancements in detection technologies to stay ahead of evolving cyber threats.

Given the significant risk posed by zero-day attacks, it is essential to explore the techniques and challenges involved in their detection, with a particular focus on artificial intelligence (AI). AI has rapidly become a cornerstone in cybersecurity, offering innovative solutions that traditional methods struggle to match. Understanding the current state-of-the-art AI methodologies and the primary difficulties encountered in detecting zero-day attacks is crucial. This

The associate editor coordinating the review of this manuscript and approving it for publication was Amjad Ali.

systematic literature review (SLR) aims to comprehensively analyze AI-based methods and challenges associated with zero-day attack detection. By examining a range of scholarly databases and publications, this review delves into the methods for detecting zero-day attacks using AI, as well as the challenges faced in this endeavor.

The primary research questions guiding this review focus on two key aspects of zero-day attack detection. First, the review discusses and evaluates the AI methods used for detecting zero-day attacks, aiming to identify and categorize the various techniques employed. This assessment includes evaluating the effectiveness and shortcomings of these methods. Second, the review explores the challenges encountered in detecting zero-day attacks, providing insights that could enhance the development of more robust detection mechanisms.

Aligned with these research questions, the objectives of this SLR are to explore the AI-based methods used to detect zero-day attacks and to investigate the challenges faced in this domain. To conduct a thorough systematic literature review, we employ a systematic approach, leveraging a wide range of renowned academic databases and repositories known for their scholarly articles and studies. By utilizing specific search keywords, this review aims to provide a comprehensive overview of the current state of knowledge, identifying gaps, trends, and emerging directions within the realm of zero-day attack detection.

The paper is structured as follows: The research method is detailed in the next section. Section II presents the method used to conduct systematic literature reviews. Section III presents the results of the systematic literature review, adhering to PRISMA guidelines. Section IV presents the taxonomy of the study. Section V discusses the taxonomy derived from the review. Lastly, Section VI provides the conclusions of the review.

## II. METHODS

This section outlines the methodology employed to conduct a SLR focused on exploring methods to detect zero-day attacks and the corresponding challenges. The methodology closely adhered to the 'Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework [1], as detailed in the PRISMA 2020 checklist (Appendix A). The PRISMA approach ensured a transparent and structured process in the selection of articles, encompassing their identification, screening, and inclusion. Adhering to PRISMA guidelines aimed to enhance the validity and reliability of our findings by providing a clear and visually comprehensive representation of the review's methodology.

The methodology was designed to thoroughly explore the methods for detecting zero-day attacks and the challenges of detecting zero-day attacks. We searched respected academic databases using specific keywords and criteria to find relevant studies. These studies were carefully assessed for their quality and relevance to our goals. We then analyzed and combined their findings to understand the various detection methods for zero-day attacks and how well the suggested defenses work. This methodological approach aimed to provide a detailed picture of the effectiveness of strategies designed to mitigate these threats and challenges.

### A. RESEARCH QUESTIONS AND OBJECTIVES

The primary objective of this systematic literature review is to examine the prevailing methods and challenges in detecting zero-day attacks. The research questions and research objectives are as follows:

Research Questions (RQ):

- What are the existing methods introduced to detect zero-day attacks using artificial intelligence?
- What are the challenges in detecting zero-day attacks?

Research Objectives:

- To identify and investigate the existing methods used to detect zero-day attacks using artificial intelligence.
- To identify and investigate the challenges in detecting zero-day attacks.

### B. INFORMATION SOURCES/DATABASES

We identified and selected a diverse range of databases, including the ACM Digital Library, IEEE Xplore, MDPI, SAGE Journals, Semantic Scholar, Science Direct, Scopus, and Web of Science. We conducted a comprehensive search of these databases to retrieve relevant journal articles and conference papers for the study.

### C. SEARCH STRATEGY/SEARCH TERMS

Based on the research questions and objectives, we devised a search strategy by combining keywords and controlled vocabulary terms relevant to intrusion detection, anomaly detection, and threat detection. To ensure a comprehensive search result, we employed different combinations of the search terms and used boolean operators, including AND and OR. For each database involved in the searching process, we customized the search strategy based on its syntax and functionalities. The chosen search terms for this study are as follows:

(''intrusion detection'' OR ''anomaly detection'' OR ''threat detection'') AND (''zero-day'' OR ''0-day'' OR ''unknown vulnerabilities'') AND (''artificial intelligence'' OR ''AI'')

### D. INCLUSION AND EXCLUSION CRITERIA

The formulation of inclusion and exclusion criteria was conducted to define the studies suitable for inclusion in the review and those that warranted exclusion. Table 1 outlines the criteria used to include or exclude papers from this review.

The decision to focus on the most recent 9 years, from April 2015 to April 2024, is based on several considerations. This period reflects a crucial phase of technological evolution, marked by substantial advancements in zero-day attack detection systems. By centering the literature review on this timeframe, the study aims to encompass the latest methods and challenges in detecting zero-day attacks, thereby contributing to a comprehensive understanding of zero-day attack detection. Meanwhile, the selected 9-year range strikes a balance between capturing a substantial body of relevant literature and preventing information overload, ensuring that the review remains manageable and that the selected studies are both recent and impactful.

**TABLE 1. Inclusion and exclusion criteria.**

| Inclusion Criteria | Exclusion Criteria |
|---|---|
| Studies that were published between 2015 and April 2024. | Review papers and survey papers. |
| Papers proposing or discussing the methods and challenges of detecting zero-day attacks. | Studies that are not accessible in full. |
| Papers published in the English language | |
| Peer-reviewed papers, articles, and conference | |

### E. SCREENING AND SELECTION PROCESS

The selection process consisted of four phases: identification, screening, eligibility, and inclusion. Phases 1 and 2 were conducted by all team members for the initial identification and screening of papers. Phases 3 and 4 involved all team members in the full-text screening and selection of the most pertinent studies.

During the identification process, we used search terms in the title, abstract, and keyword filters to locate relevant studies in the selected databases. All team members manually recorded the titles, DOIs, and URLs of the papers in a Google Sheets document. In the screening phase, we evaluated the papers for applicability and suitability by analyzing their titles and abstracts in relation to our research questions and objectives. Additionally, we manually excluded duplicated studies and those deemed inconsistent or irrelevant to our research topic.

Moving on to the eligibility phase, all team members were assigned to conduct full-text screening to assess the appropriateness of the remaining papers based on the inclusion and exclusion criteria. They also summarized the studies according to the methods and challenges of detecting zero-day attacks. Finally, in the inclusion stage, all team members evaluated the shortlisted papers based on the summaries made and eliminated those that did not meet the predetermined inclusion and exclusion criteria or were found to be irrelevant to the research question.

### F. QUALITY ASSESSMENT

We undertook a comprehensive evaluation to assess the quality of the selected papers, aiming to rigorously analyze both the strengths and weaknesses of the quantitative and qualitative studies. This evaluation entailed a detailed examination of their design and analytical methodologies.

For the assessment of the quality of our quantitative studies, we employed the Mixed Methods Appraisal Tool (MMAT) [2], a well-regarded instrument recognized for its effectiveness in evaluating the quality and potential bias of quantitative research. In contrast, for the qualitative studies, we utilized the Critical Appraisal Skills Programme (CASP) checklists [3] as our evaluative tool. These checklists provide a systematic and meticulous approach to the examination of research evidence, allowing for an assessment of its trustworthiness, relevance, and value within a specific context.

Both the MMAT and CASP checklists consist of five criteria, ensuring a thorough and comprehensive assessment of the quality of the quantitative and qualitative studies, thereby enhancing the depth and integrity of our evaluation.

### G. DATA EXTRACTION

Data extraction is a critical component of this study, addressing research questions related to existing methods and the challenges of detecting zero-day attacks. Extracted elements, such as research titles, methods, and challenges, form a comprehensive foundation for analyzing and understanding the detection of zero-day attacks. This information is necessary in order to address the research questions pertaining to the current methods and challenges of detecting zero-day attacks.

### H. DATA SYNTHESIS AND ANALYSIS

The purpose of the data synthesis methodology was to evaluate and summarize the insights obtained from the selected papers and to present the data through tables. This synthesized data forms the primary body of evidence used to address the research questions, which specifically focus on the types of methods and challenges in detecting zero-day attacks.

Content analysis emerges as the most suitable analytical approach for our study, as our primary objective is to explore and discuss the types of methods and challenges in detecting zero-day attacks.

It is important to note that specific statistical techniques, such as measures of effect and meta-regression, were not within the scope of our study. Therefore, the review is specifically designed to provide an in-depth analysis and discussion of the types of methods and challenges in detecting zero-day attacks, with an emphasis on the security perspective.

### I. RESEARCH GAPS AND CONTRIBUTIONS

This review consolidates the current state of knowledge on AI-based mechanisms for zero-day attack detection, offering a comprehensive perspective for future researchers to identify areas that necessitate further inquiry. It advances our understanding of zero-day intrusion detection by critically examining AI methodologies and the challenges documented in the literature, thereby uncovering gaps and limitations in current AI-driven strategies for mitigating zero-day attacks. By identifying the detection challenges that have not been adequately addressed within the context of AI, this review paves the way for future research initiatives.

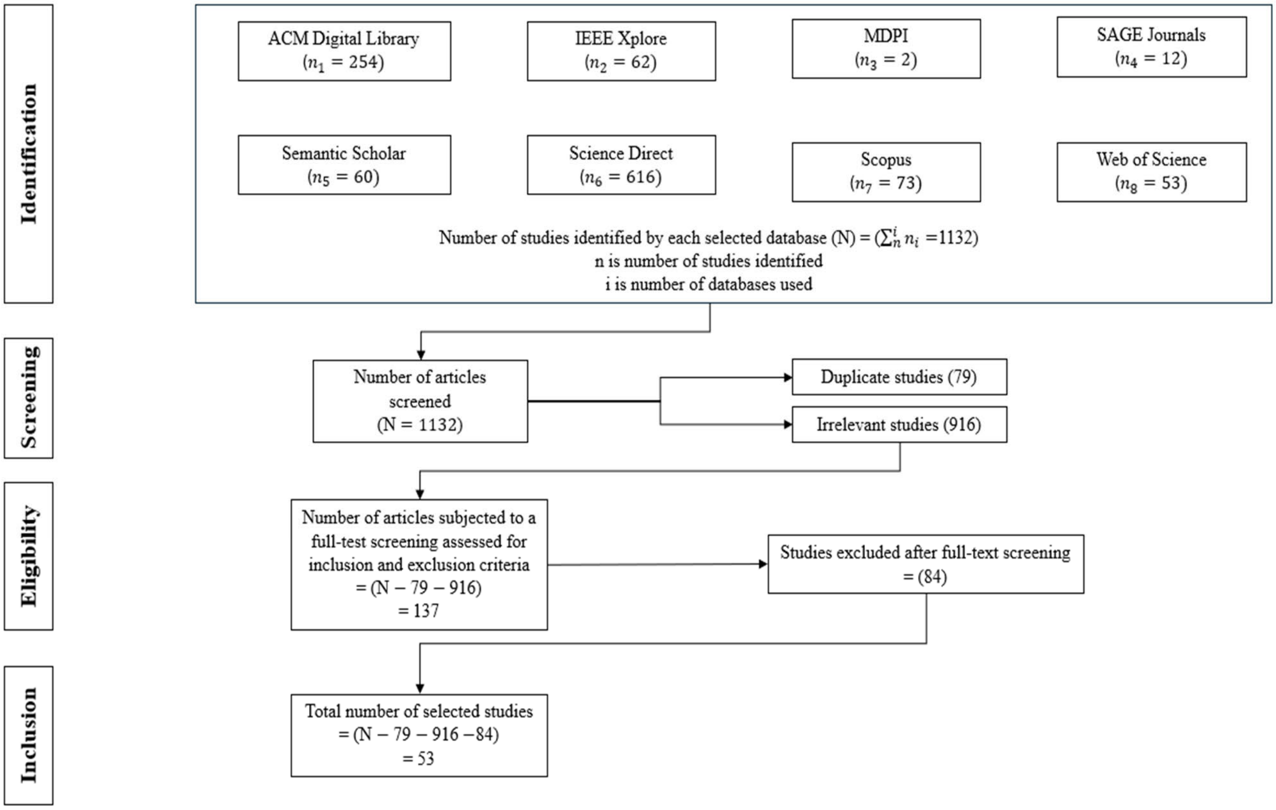

Identification
ACM Digital Library ($n_1 = 254$)
IEEE Xplore ($n_2 = 62$)
MDPI ($n_3 = 2$)
SAGE Journals ($n_4 = 12$)
Semantic Scholar ($n_5 = 60$)
Science Direct ($n_6 = 616$)
Scopus ($n_7 = 73$)
Web of Science ($n_8 = 53$)
Number of studies identified by each selected database (N) = ($\sum_n^i n_i$ =1132)
n is number of studies identified
i is number of databases used
Screening
Number of articles screened (N = 1132)
Duplicate studies (79)
Irrelevant studies (916)
Eligibility
Number of articles subjected to a full-test screening assessed for inclusion and exclusion criteria = (N − 79 − 916) = 137
Studies excluded after full-text screening = (84)
Inclusion
Total number of selected studies = (N − 79 − 916 −84) = 53

**FIGURE 1.** PRISMA flowchart for the study selection process.

Furthermore, the insights derived from this review will facilitate the design, development, and implementation of more robust and effective AI-based approaches to zero-day attack detection, thereby strengthening overall cybersecurity. By highlighting research gaps specific to AI and providing valuable insights, this systematic literature review aims to foster the development of highly resilient intrusion detection systems capable of identifying zero-day attacks through AI technologies, ultimately leading to enhanced intrusion detection systems with superior capabilities.

## III. RESULTS

### A. STUDY SELECTION

Figure 1 illustrates the four stages of the PRISMA flowchart: identification, screening, eligibility, and inclusion, conducted in this systematic literature review. During the identification stage, we found a total of 1132 studies across all selected databases by using the keywords Section II-C. The number of studies identified in each of the selected databases shown in Fig. 1.

We excluded a total of 79 duplicate studies and 916 irrelevant studies during the screening stage. After the screening stage, 137 of the remaining studies underwent full-text screening to assess their eligibility based on the inclusion and exclusion criteria. Following the full-text screening, 84 studies were excluded. In the final stage, 53 of the remaining studies that met all the inclusion and exclusion criteria were included in our systematic literature review.

### B. QUALITY OF INCLUDED STUDIES

Among the 53 quantitative studies analyzed, 52 (98.11%) received MMAT scores of 100%, signifying high quality as they fully met the evaluation criteria. Only one study (1.89%) was deemed of medium quality, with MMAT scores ranging from 60% to 80%, indicating partial adherence to the criteria. Notably, no quantitative studies were categorized as low quality. The one study classified as medium quality failed to meet criterion 1.2 (sample representativeness) due to insufficient information on the participants, such as the specific target group selected to test the developed tool. This raises concerns regarding whether the participants represent a valid sample of the overall target population. Detailed findings of the MMAT quality assessment for the quantitative studies can be found in the Assessment of Study Quality (Appendix B).

Since all analyzed studies utilized various datasets that can be interpreted in numerical form in developing their models and applied quantitative metrics in evaluating those models, all studies are quantitative. Hence, no qualitative studies are analyzed in this SLR.

### C. RQ1: WHAT ARE THE EXISTING METHODS INTRODUCED TO DETECT ZERO-DAY ATTACKS USING ARTIFICIAL INTELLIGENCE?

A total of 4 methods are reported in the articles. Table 2 displays the existing artificial intelligence method to detect zero-day attacks and the related articles discussing each method.

**TABLE 2. Methods to detect zero-day attacks and their related articles.**

| Method to detect a zero-day attack | Related Articles |
|---|---|
| Machine learning-based | [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15] |
| Deep learning-based | [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31] |
| Hybrid Model | [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47] |
| Anomaly-based | [48], [49], [50], [51], [52], [53], [54] |

There are various types of AI methods that are proposed to detect zero-day attacks. It can be categorized into four types, which are machine learning-based, deep learning-based, hybrid models, and anomaly-based. Supervised learning is a category of machine learning where the input data is labeled to train models to look for specific patterns. In supervised learning, each sample is a vector containing an input object (for instance, traffic features) and the expected outcome, often referred to as the supervisory signal or the class value. For example, supervised learning models like Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Multi-Layer Perceptron (MLP) were employed in detecting zero-day attacks for the Hikari 2021 dataset [6]. Unsupervised learning does not involve the use of labeled data and is mostly used to look for outliers. It is used for the purpose of isolating anomalous behaviors and also for the discovery of unknown attack patterns. [11]. When dealing with the detection of unknown threats or zero-day attacks, unsupervised algorithms can also be utilized as non-meta learners as well as base-level learners in meta-learning approaches [13]. Reinforcement learning refers to models that learn to do tasks based on inputs that an environment provides. Reinforcement learning (RL) is a subset of machine learning that mainly deals with the decision-making process of an agent as well as the selection of an appropriate action out of all those conceivable for a specific task. Basically, it involves an agent and the environment in which the agent acts depending on the learning gained and the status of the environment, while the environment gives rewards and punishments based on the expected action of the agent and the action it takes [14].

Deep learning uses neural networks with many layers to automatically learn hierarchical features from raw data. Common approaches include convolutional neural networks (CNN), recurrent neural networks (RNN), autoencoders, attention mechanisms, and generative adversarial networks (GAN). CNNs, for instance, have been applied to detect unfamiliar or novel forms of malware using their image-based similarity techniques [22]. Besides, federated deep learning employs deep neural network (DNN) models to detect zero-day botnet attacks on IoT-edge devices [20]. An RNN model has been developed to forecast malicious behavior, such as zero-day malware, using machine activity data [27]. RNN aids in reducing dynamic detection time while maintaining the benefits of a dynamic model. Long Short-Term Memory (LSTM), a type of RNN, was introduced in 1997 [55]. It has been utilized to identify cyberattacks and offer both global and local explanations for intrusion detection systems (IDS) [26]. A convolutional autoencoder at the character level was designed to capture features from URLs, while a deep autoencoder (AE) was developed for detecting zero-day phishing attempts [28]. This encoding/decoding framework facilitated the separation of classes, using convolution operations to capture extensive character-level URL attributes to establish anomaly scores based on reconstruction loss. Furthermore, a deep learning model for identifying malicious files employed attention-based methods for feature extraction in malware detection [29]. GANs were utilized to identify both known and zero-day attacks without relying on labeled attack data, utilizing temporal convolutional networks (TCNs) and self-attention mechanisms [30]. Additionally, an artificial neural network (ANN) was created to differentiate between known and unknown Distributed Denial of Service (DDoS) attacks based on distinct characteristic features (patterns) that distinguish DDoS attack traffic from legitimate traffic [31].

A hybrid model integrates diverse machine learning and deep learning techniques to maximize the benefits of multiple approaches. These models can enhance detection effectiveness when combined with complementary methods. For instance, the deep learning model RegNetY320 was employed to extract deep features, whereas a machine learning model like One Class Support Vector Machines (OCSVM) was utilized for detecting zero-day malware [32]. A two-stage intrusion detection framework combining the Light Gradient Boosting Machine (LightGBM) and an autoencoder was proposed for zero-day attack detection [34]. LightGBM's focal loss enhances learning from challenging samples and improves detection of attack instances in the first stage. The autoencoder used in the second stage enhances detection of misclassified samples. A hybrid deep learning model using LSTM and CNN was developed to detect zero-day attacks [37]. An active learning intrusion detection framework named Active Deep Q-Network (ADQN) was proposed to detect zero-day attacks [36]. The ADQN framework integrates active learning mechanisms with deep learning networks for zero-day attack detection. A hybrid intrusion detection model was proposed to detect various types of intrusions, encompassing both normal and recognized attacks [47]. In the realm of anomaly detection, fuzzy c-means (FCM) was employed to identify normal behaviors, utilizing relabeling techniques to uncover concealed attack data within normal patterns. For misuse detection, the model utilized Classification and Regression Trees (CART) to pinpoint previously identified attacks and subsequently employed Isolation Forest (iForest) to unveil new attack instances camouflaged within known attack patterns.

Anomaly-based detection techniques focus on pinpointing deviations from established norms in behavior patterns. These methods excel at detecting unfamiliar and zero-day attacks because they do not depend on predefined signatures. A Distributed Anomaly Detection (DAD) system was devised to identify zero-day attacks within edge networks. DAD leverages statistical learning models, including the Gaussian Mixture Model (GMM) and the Correntropy technique [48]. Additionally, a methodology involving data discretization and analysis of decision boundaries was developed to uncover zero-day attacks that exhibit behavior patterns akin to normal data. This approach discretizes data at decision points, facilitating point-by-point anomaly detection [49]. It relies on detecting statistical outliers to recognize unknown patterns.

### *D. RQ2: WHAT ARE THE CHALLENGES IN DETECTING ZERO-DAY ATTACKS?*

After gathering and collecting the difficulties faced by the researchers in RQ1, the challenges can be categorized into six categories: computation and resource challenges, data quality and processing challenges, model training challenges, accuracy and detection issues, adaptability and flexibility issues, and real-time and network adaptability issues. The specific issues faced by the researchers are detailed below based on the model used.

#### 1) MACHINE LEARNING-BASED (SUPERVISED)

i) Computation and Resource Challenges

- While the model in [5] is able to perform well in experiments, it does not guarantee good performance in practical environments due to the difficulty of adopting high-performance processors.
- The model in [9] demands resource-intensive training. The researchers also highlight the challenges in optimizing detection accuracy and computational efficiency in resource-constrained Internet of Medical Things (IoMT) and Industrial Internet of Things (IIoT) environments.

ii) Data Quality and Processing Challenges

- For [2], the detection capability is directly proportional to the data quality. Incomplete and unrealistic data that does not comprehensively represent all aspects that characterize real network traffic can lead to increases in false positives and false negatives. Reference [9] also mentions its dependency on dataset quality.
- The model developed in [10] is sensitive to noisy data and outliers, while [6] emphasizes that class imbalance adversely impacts learning performance.

iii) Accuracy and Detection Issues

- The study [6] highlights the difficulty in detecting zero-day attacks, as trained models may not be capable of identifying novel forms of attacks.
- The researchers in [7] highlighted various detection challenges affecting model performance. These include false positives caused by advertisement libraries in benign applications, false negatives associated with grayware, and inaccuracies stemming from outdated API calls. The system is also susceptible to multiple obfuscation techniques, such as encryption.
- The researchers in [10] were concerned that the false alarm rate has not completely diminished, although the model can reduce the false alarm rate.
- Certain malware has a high success rate in evading the classifiers labeled in [56].

iv) Adaptability and Flexibility Issues

- The proposed approach in [10] might not perform well for a dataset that differs from SCADA networks for gas pipelines, causing a risk to building exterior validity.

v) Real-Time and Network Adaptability Issues

- The proposed model in [9] is sensitive to network environment variations and susceptibility to evolving cyber threats. Ensuring adherence to data privacy standards is also stressed in real-world scenarios.

#### 2) MACHINE LEARNING-BASED (UNSUPERVISED)

i) Computation and Resource Challenges

- The mechanism in [11] demands graphics processing unit (GPU) hardware for inference; hence, it is challenging to implement directly on devices such as routers.

ii) Data Quality and Processing Challenges

- The study [12] overlooks certain attacks that might leave traces in memory or central processing unit (CPU) usage due to insufficient experimental data.

iii) Model Training Challenges

- The author of [13] faced challenges in tuning algorithm parameters when dealing with multiple datasets and algorithms. While automated tools test various parameter combinations, the optimal parameters for a specific algorithm in a particular system may not be included in these grid searches, leading to suboptimal algorithm performance.

iv) Adaptability and Flexibility Issues

- The localization in [12] requires enhancement by focusing exclusively on anomalous values and eventually developing a point-based rather than sequence-based detection system.

#### 3) MACHINE LEARNING-BASED (REINFORCEMENT)

i) Computation and Resource Challenges

- In [14], computational and network costs are not taken into account. Moreover, the authors point out the challenge of compromising routers, noting their fewer potentially exploitable services compared to hosts.

ii) Data Quality and Processing Challenges

- The proposed model in [15] involves manual labeling, which has the risk of mislabeling due to human error.

### 4) DEEP LEARNING-BASED

i) Computation and Resource Challenges

- The proposed attention-based model in [29] demands significantly longer training time than the SC-LSTM-based and CNN-based models. Similarly, the proposed federated deep learning model in [19] also requires a longer training time when compared to centralized deep learning, localized deep learning, and distributed deep learning. However, both proposed methods are rewarded when they outperform their competitors in detection capability.
- The proposed system in [19] involves the time-consuming process of extracting dynamic features that significantly impact the number of applications available for training the model. Additionally, complex deep learning-based models may demand substantial computational resources, making them unsuitable for on-device detection applications and necessitating the use of cloud computing services.
- The stochastic models in [25] are computationally expensive and require mitigation of computational complexity to enhance their effectiveness in practical applications.

ii) Data Quality and Processing Challenges

- The shortage of packet-level datasets constrained the capacity to test the models in [25] on novel intrusions with packet-level labels. Similarly, in [23], the limited number of available samples poses a huge challenge when conducting their experiments.

iii) Model Training Challenges

- While the AttackNet proposed in [18] demonstrates superior performance in IoT botnet attack detection, its dependence on labeled data and specific attack types could restrict its generalizability. The model's effectiveness may fluctuate as attack strategies evolve, necessitating ongoing adaptation.
- The researchers in [21] emphasize the challenge of designing a robust and inclusive collection of features that accurately describe network behaviors since both image-like representations and extracted characteristics sacrifice specific network traffic details.
- The GAN proposed in [30] requires frequent training to ensure the model is up-to-date. However, the extensive volume of network data poses significant operational challenges during retraining. Additionally, gathering training data from various nodes can present significant privacy issues.

iv) Accuracy and Detection Issues

- In [16], due to benign traffic characteristics, certain attacks such as infiltration and DdoS-LOIC UDP remain stealthy for CNN and DAE, respectively. Besides, the DAE model has a high false alarm rate that needs to be resolved.
- The proposed framework in [21] falls short of detecting unknown attacks since it still relies on specific malicious samples. The detection accuracy and recall still have room for improvement.
- In [23], the domain shift between the two domains poses challenges to achieving generalization during fine-tuning, causing reduced approach accuracy.
- The proposed method in [24] lacks a specific mechanism for dealing with minority classes.
- The proposed framework in [26] is unable to identify the specific vulnerability exploited by an attack class.
- The robustness of the approach in [27] is constrained if adversaries are aware that the initial five seconds are used to assess whether a file will execute on the network. By incorporating extended delays or benign behavior at the beginning of a malicious file, adversaries could evade detection within the virtual machine environment.
- For the model proposed in [57], the increase in attack categories will lead to a decline in the model's performance.

v) Adaptability and Flexibility Issues

- The proposed method in [17] necessitates a pre-existing corpus of URLs for data transformation and model training.
- The proposed approach in [22] is susceptible to adversarial attacks and can produce erroneous results, as even minor changes in the image can lead to misclassification.
- The proposed methodology in [28] is optimized with character-level features among the various components of URLs due to confusion in character-level features significantly contributing to the performance degradation. Hence, enhancements are required by incorporating word-level features, which encompass domains, subdomains, typos, and keywords listed in blacklists, thereby considering the structure of web addresses comprehensively.
- The proposed solution in [31] faces challenges in detecting DDoS attacks when protocol headers are encrypted because the solution is not designed to analyze encrypted packets.

vi) Real-Time and Network Adaptability Issues

- The proposed method in [37] conducts the learning stage in a batch manner without incorporating any concept drift detection mechanism to effectively adapt the learned model to a dynamic streaming environment. The authors emphasize that this is a challenge inherent to all security systems confronting adversaries.

### 5) HYBRID MODEL

i) Computation and Resource Challenges

- In [32], the use of deep learning models often necessitates substantial computational resources and power, which can limit their applicability on resource-constrained mobile ad hoc network (MANET) devices.

- In [40], which also consists of deep learning models, they suffer from a time-consuming training process and substantial computational resource consumption.

ii) Data Quality and Processing Challenges

- The proposed framework in [32] is sensitive to noisy data. Specialized data augmentation and resampling techniques for malware detection are among the challenges that require consideration.
- In [35], the performance of the model during training and testing phases may degrade with fewer datasets. Moreover, inaccuracies in samples, whether they pertain to malware or benign instances, can result in inadequate malware identification.
- The limited number of real-time attacks conducted in [37] generated reduced zero-day attack data through a generative adversarial approach.
- The sampling ratio of historical data significantly influences the accuracy and update time of the proposed intrusion detection system in [39], which are crucial in determining the detection performance and response speed for zero-day attacks.
- In [41], the absence of labeled multi-attack datasets specifically tailored to MANETs can impede model training and evaluation.
- The performance of the model in [43] is significantly dependent on the quality of the data utilized for training and testing. In situations where the data is noisy, incomplete, or otherwise compromised, the accuracy and effectiveness of the model may be adversely impacted.
- The imbalanced data issue in the NSL-KDD may resist the true potential of the proposed model in the study [48].

iii) Accuracy and Detection Issues

- In [42], certain attack classes characterized by fewer distinctive patterns or features present challenges for accurate generation.
- The proposed framework in [40] is unable to identify minor modifications on the device that are not affecting the device's network performance, for example, changes in screen brightness.
- The proposed model in [48] requires improvements in lowering false-positive rates of anomaly detection and misuse detection.

iv) Adaptability and Flexibility Issues

- The proposed method in [34] may achieve a high recall in one dataset but fall short in another. Besides, the precision of the method is not yet satisfied.
- The study [38] faces challenges when dealing with multiple languages and an executable binary. The proposed approach only supports the C/C++ language; detecting vulnerabilities in binary files without access to the source code is more challenging.
- The framework proposed in [42] did not consider secure aggregation and authentication of model sharing that ensure its integrity and trustworthiness.
- The proposed system in [44] could misclassify the attack class during prediction, resulting in misclassification errors and incorrect attack mitigation strategies. Additionally, the model may fail to detect zero-day attacks where the network traffic profiles generated closely resemble legitimate network profiles.
- There are several adaptability issues mentioned in [45]. Some cyclostationary features were lost when addressing the incompatibility issue that occurred in the framework. Besides, the study's findings may not be applicable to other datasets or contexts. Additionally, there is a decrease in the classification of unknown traffic, and an increase in unclassified traffic was observed following the implementation of the UGRansome properties-based rule.

v) Real-Time and Network Adaptability Issues

- There are certain adaptability limitations in the proposed model [33], such as that Ipv6 is not accounted for, OptiFilter occasionally requires manual configuration, and the alarm system should be more interactive in ideal situations.
- The authors of [43] underscore several limitations and obstacles associated with their model. Variations in network architecture and traffic patterns, distinct from those encountered during the model's initial testing and configuration, pose scalability challenges in large network environments. The model's adaptability across a broader spectrum of network environments is called into question, given its emphasis on specific network customizations. In scenarios with high traffic volumes and environments featuring limited computational resources, such as edge devices in IoT networks, the model faces challenges in real-time data processing. Moreover, the unpredictable and fluctuating nature of live data streams complicates maintaining consistent model performance. Fluctuations in network conditions, including changes in traffic levels and congestion, significantly impact the model's ability to reliably detect intrusions.
- In [45], during real-time testing of UGRansome, the study observed a packet drop anomaly, indicated by a reduction in network traffic when UGRansome data was integrated into the IDS to block infected traffic. The underlying cause of this issue remains unidentified.

### 6) ANOMALY-BASED

i) Computation and Resource Challenges

- The approach in [46] is time-consuming, as the system can become overly intricate when handling large volumes of data.
- The proposed system in [49] would require significant computational resources, especially if it monitors diverse data sources like network traffic, operating system audit traces, and telemetry data from IoT devices.

ii) Data Quality and Processing Challenges

- The data processing in [52] can cause feature loss to a certain extent, which impacts the intrusion detection's accuracy.

iii) Model Training Challenges

- The proposed system in [49] requires an excessive number of normal vectors during the training phase to maintain high reliability.

iv) Accuracy and Detection Issues

- On occasion, the proposed anomaly detection model in [50] is unable to correctly identify data that is close to the decision boundary.
- The researchers in [51] highlight that the detection of Honeypots poses a common challenge in the field of Honeypot studies, requiring thorough investigation and research within the Honeypot research domain.

v) Adaptability and Flexibility Issues

- The approach in [46] lacks scalability as the system becomes unduly complex when significant volumes of data are employed.
- In [52], the encoding time increases with data volume, making the proposed method potentially unsuitable for detecting large-scale data.
- The out-of-distribution (OOD) detection in [53] is limited in the absence of OOD data and domain expertise. Besides, the data generation method has low performance on high-dimensional datasets, which could be attributed to the restricted variety of the generated OOD samples.
- The proximity-based features in the approach [54] are not efficient, as most new blacklisted IPs originated from /24 subnetworks characterized by a low density of blacklisted IPs.

Table 3 shows the summary of existing challenges for each method category.

## IV. STUDY TAXONOMY

Figure 2 presents a taxonomy of methods and corresponding challenges in detecting zero-day attacks. This taxonomy is developed based on the findings from research questions 1 and 2, which explore the landscape of methods used to detect zero-day attacks and the challenges faced in implementing these methods.

In this taxonomy, the methods are categorized into four primary groups, which are Machine Learning-Based, Deep Learning-Based, Hybrid Models, and Anomaly-Based. Each category is further broken down into specific techniques and their associated challenges, offering a detailed view of the current state of zero-day attack detection using artificial intelligence.

The first category, the Machine Learning-Based method, encompasses models such as Supervised Learning, Unsupervised Learning, and Reinforcement Learning. This method faces challenges such as computation and resource challenges, data quality and processing challenges, model training challenges, accuracy and detection issues, adaptability and flexibility issues, and real-time and network adaptability issues.

Moving on, Deep Learning-Based method faces challenges such as computation and resource challenges, data quality and processing challenges, model training challenges, accuracy and detection issues, adaptability and flexibility issues, and real-time and network adaptability issues.

Hybrid models are in the third category. It faces challenges such as computation and resource challenges, data quality and processing challenges, accuracy and detection issues, adaptability and flexibility issues, and real-time and network adaptability issues. Lastly, the Anomaly-Based method faces challenges such as computation and resource challenges, data quality and processing challenges, model training challenges, accuracy and detection issues, and adaptability and flexibility issues.

**TABLE 3. Existing detection method and their challenges.**

| Method | Challenges |
|---|---|
| Machine Learning-Based (Supervised) | • Computation and Resource Challenges<br>• Data Quality and Processing Challenges<br>• Accuracy and Detection Issues<br>• Adaptability and Flexibility Issues<br>• Real-Time and Network Adaptability Issues |
| Machine Learning-Based (Unsupervised) | • Computation and Resource Challenges<br>• Data Quality and Processing Challenges<br>• Model Training Challenges<br>• Adaptability and Flexibility Issues |
| Machine Learning-Based (Reinforcement) | • Computation and Resource Challenges<br>• Data Quality and Processing Challenges |
| Deep Learning-Based | • Computation and Resource Challenges<br>• Data Quality and Processing Challenges<br>• Model Training Challenges<br>• Accuracy and Detection Issues<br>• Adaptability and Flexibility Issues<br>• Real-Time and Network Adaptability Issues |
| Hybrid Model | • Computation and Resource Challenges<br>• Data Quality and Processing Challenges<br>• Accuracy and Detection Issues<br>• Adaptability and Flexibility Issues<br>• Real-Time and Network Adaptability Issues |
| Anomaly-Based | • Computation and Resource Challenges<br>• Data Quality and Processing Challenges<br>• Model Training Challenges<br>• Accuracy and Detection Issues<br>• Adaptability and Flexibility Issues |

This taxonomy offers a structured framework to understand the diverse array of methods and challenges in detecting zero-day attacks. By categorizing these methods and their associated challenges, it provides valuable insights for researchers, practitioners, and developers working on enhancing zero-day attack detection systems.

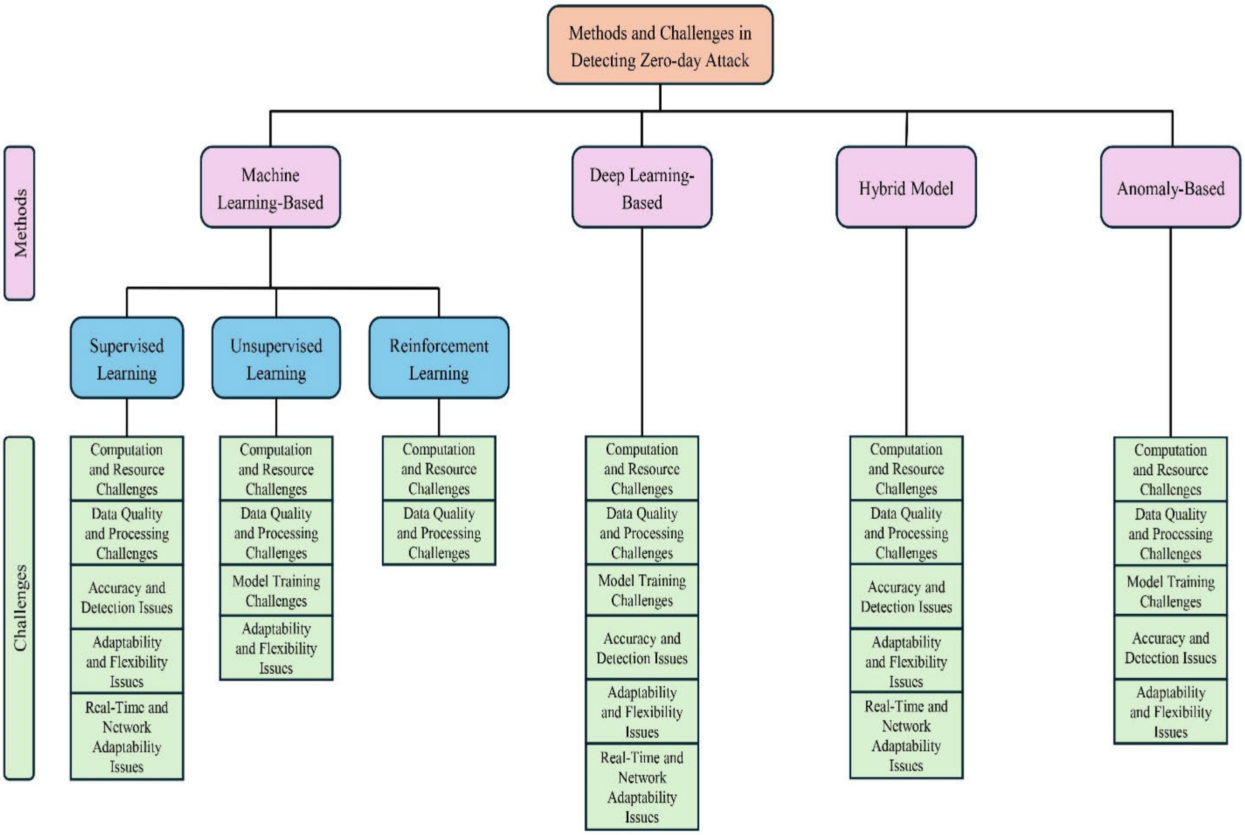

Methods and Challenges in Detecting Zero-day Attack
Methods
Machine Learning-Based
Deep Learning-Based
Hybrid Model
Anomaly-Based
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Challenges
Computation and Resource Challenges
Data Quality and Processing Challenges
Accuracy and Detection Issues
Adaptability and Flexibility Issues
Real-Time and Network Adaptability Issues
Model Training Challenges

**FIGURE 2.** Taxonomy of methods and challenges in detecting zero-day attacks.

## V. DISCUSSION

Machine learning approaches have significantly advanced intrusion detection systems, and consequently, machine learning-based methods, particularly supervised and unsupervised techniques, are frequently proposed for zero-day attack detection. In supervised techniques, numerous studies indicate that the effectiveness of these methods heavily relies on the quality of the training data. Additionally, some researchers have emphasized the challenges in detecting zero-day and other specific types of attacks, often reporting high false alarm rates. Moreover, certain machine learning techniques face challenges related to computational constraints.

Beyond conventional machine learning, AI techniques such as deep learning, hybrid models, and anomaly-based methods are also employed to detect zero-day attacks, offering distinct strategies for addressing these intrusions. Deep learning utilizes neural networks to extract hierarchical features from raw data, with various types of neural networks being applied, including ANN, CNN, DNN, and RNN. Hybrid models combine and integrate multiple models to leverage the strengths of each component within the ensemble. These hybrid models can be flexibly composed using a combination of machine learning, deep learning techniques, or other methodologies to achieve superior results. Anomaly-based approaches focus on identifying deviations from established normal behavior patterns, thereby eliminating the need for predefined signatures.

However, several challenges are associated with these AI algorithms. For deep learning, issues related to the time-consuming nature of data training frequently arise. Additionally, deep learning approaches can be computationally expensive due to the complexity of the algorithms involved. The main data quality issue is the lack of available data. Concerns about scalability persist, as researchers are often uncertain about the performance of the proposed models as attacks evolve and diversify. Due to experimental designs, proposed deep learning models often have restrictions, making them suitable for detecting only specific types of attacks and adaptable to limited environments. Therefore, the usability of proposed deep learning models should be broadened to better meet real-world requirements.

For hybrid models, challenges inherent to individual algorithms persist if other components in the model do not adequately address these issues. For instance, significant computational resources may be required when incorporating deep learning algorithms. Additionally, a lack of sufficient quality or quantity of data can diminish the performance of these models. Although hybrid models appear to enhance detection capabilities—only 3 out of 17 studies report related issues—adaptability challenges remain a prominent concern, with certain models struggling to perform effectively in real-time network environments.

Lastly, for anomaly-based approaches, scalability is the primary issue, as data processing performance may degrade dynamically when confronted with increasing data volumes and high-dimensional data.

The incident involving CrowdStrike on July 19, 2024, where the company's advanced AI-driven detection systems successfully identified and neutralized a sophisticated zero-day attack, exemplifies both the strengths and challenges of current detection methodologies. This incident underscores the critical need for robust, real-time detection capabilities and the importance of continuous advancements in AI detection technologies to keep pace with evolving cyber threats. It also highlights the effectiveness of hybrid and anomaly-based approaches in identifying sophisticated attacks while pointing out the ongoing need to address issues such as computational constraints and data quality.

In general, the challenges reported in the literature can range from general to very specific. Some challenges might also be present in other studies, although the authors may not report them due to a lack of consideration, particularly concerning real-world usability. Most literature does not discuss real-time detection, and thus the related challenges are often overlooked. Therefore, future research should not only address the existing challenges highlighted in previous literature but also explore other unmentioned challenges to enhance the robustness and applicability of the solutions.

Our research has encountered challenges, such as the absence of standardized zero-day attack detection systems, which complicates comparisons among studies and their findings. Some studies lack clarity on zero-day intrusion detection, making it challenging to determine their adherence to established criteria. Diverse implementations and variations of intrusion detection models may impede the establishment of universal conclusions.

Additionally, our SLR encompasses articles published only until April 30, 2024, potentially excluding more recent research. Moreover, our focus is solely on zero-day attack detection, neglecting existing discovered attacks. Given the rapid evolution of AI-driven detection mechanisms driven by the increasing prevalence of zero-day attacks in recent trends, it is critical to acknowledge the time frame covered by our SLR. The challenges identified may be addressed in subsequent research. The diverse AI methodologies employed in the studies included in this review for detecting zero-day intrusions may also pose challenges for comparability. Additionally, certain proposed frameworks may become outdated and unsuitable for real-world usability in the face of rapid security threat evolution.

Despite these limitations, our review provides a valuable overview of the current research landscape in this domain and highlights opportunities for future research to mitigate the identified challenges.

Moving forward, it is crucial to underscore that security attacks are evolving in increasingly sophisticated ways, necessitating heightened attention and investigation to effectively safeguard privacy and information security. The CrowdStrike incident serves as a timely reminder of the ongoing threat posed by zero-day attacks and the critical importance of developing adaptive and resilient AI-driven detection systems capable of responding to the ever-changing cyber threat landscape.

## VI. CONCLUSION

This SLR explores the current landscape of zero-day attack detection within the field of AI and examines the associated challenges. The review involved a comprehensive search across eight databases, yielding a total of 1,132 articles through the use of specific keywords. After a meticulous filtering and screening process, 53 studies were identified that met all the predefined inclusion and exclusion criteria for this SLR.

In this study, we identified various AI-driven mechanisms aimed at zero-day intrusion detection. Notably, machine learning, deep learning, anomaly detection approaches, and hybrid models combining multiple algorithms have been prominently applied in the existing literature. While the proposed AI algorithms have generally achieved satisfactory results in detecting zero-day attacks, they often exhibit limitations in various aspects. Furthermore, several prior studies are characterized by narrow scopes, concentrating exclusively on specific types of attacks or environments. Despite the availability of more advanced and evolving AI approaches, we have endeavored to compile a comprehensive collection of methods and challenges identified through our extensive research.

This review contributes to the field in two notable ways. Firstly, it amalgamates existing knowledge by presenting a well-structured overview of the current state of AI research in this domain. Secondly, it highlights gaps and constraints in the literature, thus outlining a roadmap for future research and innovation in AI-based zero-day attack detection.

Our comprehensive analysis of these articles has unveiled a wide array of challenges and limitations that existing AI-based intrusion detection models should address to facilitate their advancement. These challenges encompass conventional issues such as constraints in model performance as well as real-world adaptability concerns, including computational demands, real-time responsiveness, and scalability difficulties. This range of limitations underscores the urgent necessity for ongoing research and the development of enhanced or innovative AI approaches in the field of zero-day attack defense.

Focusing on the defensive aspect, each AI approach has its own unique role and strengths in detecting unknown intrusions, as justified by the authors, and can be effectively applied in appropriate environments. The use of hybrid models is particularly prevalent because they can counterbalance the shortcomings of individual components within the model and amplify the strengths of each. The integration of multiple AI models often yields more promising results.

However, despite advancements in AI detection mechanisms, significant limitations remain in mitigating zero-day attacks, especially considering the evolving and increasingly sophisticated nature of attack technologies. The recent CrowdStrike incident, where the company's advanced AI-driven detection systems successfully identified and neutralized a sophisticated zero-day attack, exemplifies the critical importance of robust, real-time detection capabilities. This event underscores the necessity for continuous advancements in AI detection technologies to keep pace with evolving cyber threats.

The synthesis of findings from numerous studies highlights the critical need for a comprehensive AI-based approach to intrusion detection systems to effectively safeguard against zero-day intrusions. Moving forward, it is essential to address not only technical challenges but also the issues of real-world and real-time adaptability. Developing effective strategies requires a multidisciplinary approach involving domain experts such as security specialists, computer scientists, and AI engineers, recognizing the interplay of various factors in protecting against zero-day attacks.

In conclusion, while significant progress has been made in the field of zero-day attack detection through AI, the dynamic and sophisticated nature of cyber threats necessitates ongoing research and innovation. The insights gained from the CrowdStrike incident on July 19, 2024, along with the challenges identified in this review, highlight the urgent need for adaptive and resilient AI detection systems. Future research should focus on enhancing the real-world applicability of these systems, ensuring they are equipped to handle the complexities and scale of modern cyber threats.

## REFERENCES

[1] Q. N. Hong et al. (2018). *Mixed Methods Appraisal Tool (MMAT)*. Accessed: Jul. 2, 2024. [Online]. Available: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf

[2] Critical Appraisal Skills Programme (CASP). (2018). *CASP Qualitative Studies Checklist*. Accessed: Jul. 2, 2024. [Online]. Available: https://casp-uk.net/checklists/casp-qualitative-studies-checklist-fillable.pdf

[3] G. D'angelo, F. Palmieri, M. Ficco, and S. Rampone, "An uncertainty-managing batch relevance-based approach to network anomaly detection," *Appl. Soft Comput.*, vol. 36, pp. 408–418, Nov. 2015, doi: 10.1016/j.asoc.2015.07.029.

[4] K. Bong and J. Kim, "Analysis of intrusion detection performance by smoothing factor of Gaussian NB model using modified NSL-KDD dataset," in *Proc. 13th Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2022, pp. 1471–1476, doi: 10.1109/ICTC55196.2022.9952381.

[5] D. Kwon, R.-M. Neagu, P. Rasakonda, J. T. Ryu, and J. Kim, "Evaluating unbalanced network data for attack detection," in *Proc. Syst. Netw. Telemetry Anal.*, Jul. 2023, pp. 23–26, doi: 10.1145/3589012.3594898.

[6] D. Zou, Y. Wu, S. Yang, A. Chauhan, W. Yang, J. Zhong, S. Dou, and H. Jin, "IntDroid," *ACM Trans. Softw. Eng. Methodol.*, vol. 30, no. 3, pp. 1–32, May 2021, doi: 10.1145/3442588.

[7] N. Usman, S. Usman, F. Khan, M. A. Jan, A. Sajid, M. Alazab, and P. Watters, "Intelligent dynamic malware detection using machine learning in IP reputation for forensics data analytics," *Future Gener. Comput. Syst.*, vol. 118, pp. 124–141, May 2021, doi: 10.1016/j.future.2021.01.004.

[8] U. Zukaib, X. Cui, C. Zheng, M. Hassan, and Z. Shen, "Meta-IDS: Meta-learning-based smart intrusion detection system for Internet of Medical Things (IoMT) network," *IEEE Internet Things J.*, vol. 11, no. 13, pp. 23080–23095, Jul. 2024, doi: 10.1109/JIOT.2024.3387294.

[9] A. Verma, R. Saha, G. Kumar, M. Conti, and T.-H. Kim, "PREVIR: Fortifying vehicular networks against denial of service attacks," *IEEE Access*, vol. 12, pp. 48301–48320, 2024, doi: 10.1109/ACCESS.2024.3382992.

[10] F. Carrera, V. Dentamaro, S. Galantucci, A. Iannacone, D. Impedovo, and G. Pirlo, "Combining unsupervised approaches for near real-time network traffic anomaly detection," *Appl. Sci.*, vol. 12, no. 3, p. 1759, Feb. 2022, doi: 10.3390/app12031759.

[11] W. Haider, N. Moustafa, M. Keshk, A. Fernandez, K.-K.-R. Choo, and A. Wahab, "FGMC-HADS: Fuzzy Gaussian mixture-based correntropy models for detecting zero-day attacks from Linux systems," *Comput. Secur.*, vol. 96, Sep. 2020, Art. no. 101906, doi: 10.1016/j.cose.2020.101906.

[12] T. Zoppi, M. Gharib, M. Atif, and A. Bondavalli, "Meta-learning to improve unsupervised intrusion detection in cyber-physical systems," *ACM Trans. Cyber-Phys. Syst.*, vol. 5, no. 4, pp. 1–27, Sep. 2021, doi: 10.1145/3467470.

[13] K. Sethi, Y. V. Madhav, R. Kumar, and P. Bera, "Attention based multi-agent intrusion detection systems using reinforcement learning," *J. Inf. Secur. Appl.*, vol. 61, Sep. 2021, Art. no. 102923, doi: 10.1016/j.jisa.2021.102923.

[14] S. Shen, C. Cai, Z. Li, Y. Shen, G. Wu, and S. Yu, "Deep Q-network-based heuristic intrusion detection against edge-based SIoT zero-day attacks," *Appl. Soft Comput.*, vol. 150, Jan. 2024, Art. no. 111080, doi: 10.1016/j.asoc.2023.111080.

[15] J. C. J. Badji and C. Diallo, "A CNN-based attack classification versus an AE-based unsupervised anomaly detection for intrusion detection systems," in *Proc. Int. Conf. Electr., Comput. Energy Technol. (ICECET)*, Jul. 2022, pp. 1–7, doi: 10.1109/ICECET55527.2022.9873072.

[16] Q. Ma, C. Sun, B. Cui, and X. Jin, "A novel model for anomaly detection in network traffic based on kernel support vector machine," *Comput. Secur.*, vol. 104, May 2021, Art. no. 102215, doi: 10.1016/j.cose.2021.102215.

[17] H. Nandanwar and R. Katarya, "Deep learning enabled intrusion detection system for industrial IoT environment," *Expert Syst. Appl.*, vol. 249, Sep. 2024, Art. no. 123808, doi: 10.1016/j.eswa.2024.123808.

[18] A. R. Nasser, A. M. Hasan, and A. J. Humaidi, "DL-AMDet: Deep learning-based malware detector for Android," *Intell. Syst. Appl.*, vol. 21, Mar. 2024, Art. no. 200318, doi: 10.1016/j.iswa.2023.200318.

[19] S. I. Popoola, R. Ande, B. Adebisi, G. Gui, M. Hammoudeh, and O. Jogunola, "Federated deep learning for zero-day botnet attack detection in IoT-edge devices," *IEEE Internet Things J.*, vol. 9, no. 5, pp. 3930–3944, Mar. 2022, doi: 10.1109/JIOT.2021.3100755.

[20] J. Yang, H. Li, S. Shao, F. Zou, and Y. Wu, "FS-IDS: A framework for intrusion detection based on few-shot learning," *Comput. Secur.*, vol. 122, Nov. 2022, Art. no. 102899, doi: 10.1016/j.cose.2022.102899.

[21] R. Kumar, Z. Xiaosong, R. U. Khan, I. Ahad, and J. Kumar, "Malicious code detection based on image processing using deep learning," in *Proc. Int. Conf. Comput. Artif. Intell.*, Mar. 2018, pp. 81–85, doi: 10.1145/3194452.3194459.

[22] S. Bhardwaj, A. S. Li, M. Dave, and E. Bertino, "Overcoming the lack of labeled data: Training malware detection models using adversarial domain adaptation," *Comput. Secur.*, vol. 140, May 2024, Art. no. 103769, doi: 10.1016/j.cose.2024.103769.

[23] G. Andresini, A. Appice, F. P. Caforio, D. Malerba, and G. Vessio, "ROULETTE: A neural attention multi-output model for explainable network intrusion detection," *Expert Syst. Appl.*, vol. 201, Sep. 2022, Art. no. 117144, doi: 10.1016/j.eswa.2022.117144.

[24] J. A. Wong, A. M. Berenbeim, D. A. Bierbrauer, and N. D. Bastian, "Uncertainty-quantified, robust deep learning for network intrusion detection," in *Proc. Winter Simulation Conf. (WSC)*, vol. 31, Dec. 2023, pp. 2470–2481, doi: 10.1109/wsc60868.2023.10407559.

[25] M. Keshk, N. Koroniotis, N. Pham, N. Moustafa, B. Turnbull, and A. Y. Zomaya, "An explainable deep learning-enabled intrusion detection framework in IoT networks," *Inf. Sci.*, vol. 639, Aug. 2023, Art. no. 119000, doi: 10.1016/j.ins.2023.119000.

[26] M. Rhode, P. Burnap, and K. Jones, "Early-stage malware prediction using recurrent neural networks," *Comput. Secur.*, vol. 77, pp. 578–594, Aug. 2018, doi: 10.1016/j.cose.2018.05.010.

[27] S.-J. Bu and S.-B. Cho, "Deep character-level anomaly detection based on a convolutional autoencoder for zero-day phishing URL detection," *Electronics*, vol. 10, no. 12, p. 1492, Jun. 2021, doi: 10.3390/electronics10121492.

[28] S. Choi, J. Bae, C. Lee, Y. Kim, and J. Kim, "Attention-based automated feature extraction for malware analysis," *Sensors*, vol. 20, no. 10, p. 2893, May 2020, doi: 10.3390/s20102893.

[29] P. F. de Araujo-Filho, M. Naili, G. Kaddoum, E. T. Fapi, and Z. Zhu, "Unsupervised GAN-based intrusion detection system using temporal convolutional networks and self-attention," *IEEE Trans. Netw. Service Manage.*, vol. 20, no. 4, pp. 4951–4963, Dec. 2023, doi: 10.1109/TNSM.2023.3260039.

[30] A. Saied, R. E. Overill, and T. Radzik, "Detection of known and unknown DDoS attacks using artificial neural networks," *Neurocomputing*, vol. 172, pp. 385–393, Jan. 2016, doi: 10.1016/j.neucom.2015.04.101.

[31] K. Shaukat, S. Luo, and V. Varadharajan, "A novel machine learning approach for detecting first-time-appeared malware," *Eng. Appl. Artif. Intell.*, vol. 131, May 2024, Art. no. 107801, doi: 10.1016/j.engappai.2023.107801.

[32] M. Salem and A.-K. Al-Tamimi, "A novel threat intelligence detection model using neural networks," *IEEE Access*, vol. 10, pp. 131229–131245, 2022, doi: 10.1109/ACCESS.2022.3229495.

[33] H. Zhang, L. Ge, G. Zhang, J. Fan, D. Li, and C. Xu, "A two-stage intrusion detection method based on light gradient boosting machine and autoencoder," *Math. Biosci. Eng.*, vol. 20, no. 4, pp. 6966–6992, Jan. 2023, doi: 10.3934/mbe.2023301.

[34] S. K. Smmarwar, G. P. Gupta, and S. Kumar, "AI-empowered malware detection system for industrial Internet of Things," *Comput. Electr. Eng.*, vol. 108, May 2023, Art. no. 108731, doi: 10.1016/j.compeleceng.2023.108731.

[35] Y. Wu, Y. Hu, J. Wang, M. Feng, A. Dong, and Y. Yang, "An active learning framework using deep Q-network for zero-day attack detection," *Comput. Secur.*, vol. 139, Apr. 2024, Art. no. 103713, doi: 10.1016/j.cose.2024.103713.

[36] M. Asaduzzaman and M. M. Rahman, "An adversarial approach for intrusion detection using hybrid deep learning model," in *Proc. Int. Conf. Inf. Technol. Res. Innov. (ICITRI)*, Nov. 2022, pp. 18–23, doi: 10.1109/ICITRI56423.2022.9970221.

[37] S. Jeon and H. K. Kim, "AutoVAS: An automated vulnerability analysis system with a deep learning approach," *Comput. Secur.*, vol. 106, Jul. 2021, Art. no. 102308, doi: 10.1016/j.cose.2021.102308.

[38] D. Jin, S. Chen, H. He, X. Jiang, S. Cheng, and J. Yang, "Federated incremental learning based evolvable intrusion detection system for zero-day attacks," *IEEE Netw.*, vol. 37, no. 1, pp. 125–132, Jan. 2023, doi: 10.1109/MNET.018.2200349.

[39] Y. K. Saheed, O. H. Abdulganiyu, and T. A. Tchakoucht, "Modified genetic algorithm and fine-tuned long short-term memory network for intrusion detection in the Internet of Things networks with edge capabilities," *Appl. Soft Comput.*, vol. 155, Apr. 2024, Art. no. 111434, doi: 10.1016/j.asoc.2024.111434.

[40] R. Reka, R. Karthick, R. S. Ram, and G. Singh, "Multi head self-attention gated graph convolutional network based multi-attack intrusion detection in MANET," *Comput. Secur.*, vol. 136, Jan. 2024, Art. no. 103526, doi: 10.1016/j.cose.2023.103526.

[41] D. Hamouda, M. A. Ferrag, N. Benhamida, H. Seridi, and M. C. Ghanem, "Revolutionizing intrusion detection in industrial IoT with distributed learning and deep generative techniques," *Internet Things*, vol. 26, Jul. 2024, Art. no. 101149, doi: 10.1016/j.iot.2024.101149.

[42] S. M. S. Bukhari, M. H. Zafar, M. A. Houran, S. K. R. Moosavi, M. Mansoor, M. Muaaz, and F. Sanfilippo, "Secure and privacy-preserving intrusion detection in wireless sensor networks: Federated learning with SCNN-Bi-LSTM for enhanced reliability," *Ad Hoc Netw.*, vol. 155, Mar. 2024, Art. no. 103407, doi: 10.1016/j.adhoc.2024.103407.

[43] B. M. Serinelli, A. Collen, and N. A. Nijdam, "Training guidance with KDD cup 1999 and NSL-KDD data sets of ANIDINR: Anomaly-based network intrusion detection system," *Proc. Comput. Sci.*, vol. 175, pp. 560–565, Jan. 2020, doi: 10.1016/j.procs.2020.07.080.

[44] M. W. Nkongolo, "Zero-day vulnerability prevention with recursive feature elimination and ensemble learning," Cryptol. ePrint Arch., CA, USA, Tech. Rep. 2023/1843, 2023. Accessed: Jul. 2, 2024. [Online]. Available: https://eprint.iacr.org/2023/1843

[45] I. A. Khan, D. Pi, Z. U. Khan, Y. Hussain, and A. Nawaz, "HML-IDS: A hybrid-multilevel anomaly prediction approach for intrusion detection in SCADA systems," *IEEE Access*, vol. 7, pp. 89507–89521, 2019, doi: 10.1109/ACCESS.2019.2925838.

[46] G.-Y. Shin, D.-W. Kim, S.-S. Kim, and M.-M. Han, "Unknown attack detection: Combining relabeling and hybrid intrusion detection," *Comput., Mater. Continua*, vol. 68, no. 3, pp. 3289–3303, Jan. 2021, doi: 10.32604/cmc.2021.017502.

[47] N. Moustafa, M. Keshk, K.-K.-R. Choo, T. Lynar, S. Camtepe, and M. Whitty, "DAD: A distributed anomaly detection system using ensemble one-class statistical learning in edge networks," *Future Gener. Comput. Syst.*, vol. 118, pp. 240–251, May 2021, doi: 10.1016/j.future.2021.01.011.

[48] G.-Y. Shin, D.-W. Kim, and M.-M. Han, "Data discretization and decision boundary data point analysis for unknown attack detection," *IEEE Access*, vol. 10, pp. 114008–114015, 2022, doi: 10.1109/ACCESS.2022.3215269.

[49] M. Mohammadzad, J. Karimpour, and F. Mahan, "MAGD: Minimal attack graph generation dynamically in cyber security," *Comput. Netw.*, vol. 236, Nov. 2023, Art. no. 110004, doi: 10.1016/j.comnet.2023.110004.

[50] T. Hou, H. Xing, X. Liang, X. Su, and Z. Wang, "Network intrusion detection based on DNA spatial information," *Comput. Netw.*, vol. 217, Nov. 2022, Art. no. 109318, doi: 10.1016/j.comnet.2022.109318.

[51] S. Koda and I. Morikawa, "OOD-robust boosting tree for intrusion detection systems," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, vol. 30, Jun. 2023, pp. 1–10, doi: 10.1109/ijcnn54540.2023.10191603.

[52] D. Likhomanov and V. Poliukh, "Predicting malicious hosts by blacklisted IPv4 address density estimation," in *Proc. IEEE 11th Int. Conf. Dependable Syst., Services Technol. (DESSERT)*, May 2020, pp. 102–109. Accessed: Sep. 17, 2023. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9125012, doi: 10.1109/DESSERT50317.2020.9125012.

[53] A. H. Celdrán, P. M. S. Sánchez, J. von der Assen, D. Shushack, Á. L. P. Gómez, G. Bovet, G. M. Pérez, and B. Stiller, "Behavioral fingerprinting to detect ransomware in resource-constrained devices," *Comput. Secur.*, vol. 135, Dec. 2023, Art. no. 103510, doi: 10.1016/j.cose.2023.103510.

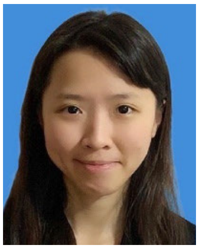
[54] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[55] N. Nissim, Y. Lapidot, A. Cohen, and Y. Elovici, "Trusted system-calls analysis methodology aimed at detection of compromised virtual machines using sequential mining," *Knowl.-Based Syst.*, vol. 153, pp. 147–175, Aug. 2018, doi: 10.1016/j.knosys.2018.04.033.

[56] C. Redino, D. Nandakumar, R. Schiller, K. Choi, A. Rahman, E. Bowen, A. Shaha, J. Nehila, and M. Weeks, "Zero day threat detection using graph and flow based security telemetry," in *Proc. Int. Conf. Comput., Commun., Intell. Syst. (ICCCIS)*, Nov. 2022, pp. 655–662, doi: 10.1109/ICCCIS56430.2022.10037596.

**LIP YEE POR** (Senior Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees from Universiti Malaya, Malaysia. He currently holds the position of a Professor with the Faculty of Computer Science and Information Technology, Universiti Malaya. His research interests include various aspects of information security and quality assurance (National Education Code (NEC) 2020: 0611), including authentication, graphic passwords, PIN-entry, cryptography, data hiding, steganography, and watermarking. Additionally, he specializes in machine learning (NEC 2020: 0613), with expertise in extreme learning machines, support vector machines, deep learning, long-short-term memory, computer vision, and the AIoT.

**ZHEN DAI** received the B.Sc. degree in geomatics engineering from Anhui Agricultural University, in 2015, and the M.Sc. degree in applied computing from Universiti Malaya, in 2019, where he is currently pursuing the Ph.D. degree. He is a Research and Development Engineer with Sun-Create Electronics Company Ltd., which is part of China Electronics Technology Group Corporation. His research interests include cybersecurity, image processing, deep learning, machine learning, and optimization algorithms.



**SIEW JUAN LEEM** received the Bachelor of Chemical Engineering Technology degree from Universiti Malaysia Perlis, in 2021. She is currently pursuing the master's degree with Universiti Malaya. Her current research interests include cybersecurity and machine learning.

**YI CHEN** received the Bachelor of Applied Sciences (Operation Research) degree from Universiti Sains Malaysia, in 2022. He is currently pursuing the master's degree with the Universiti Malaya. His current research interests include cybersecurity and machine learning.

**JING YANG** (Graduate Student Member, IEEE) received the Bachelor of Engineering degree in navigation technology from Shandong Jiaotong University, in 2022, and the master's degree (Hons.) in data science from Universiti Malaya, Malaysia, in 2024, where he is currently pursuing the Ph.D. degree. His primary research interests include medical image processing, deep learning, the IoT, and blockchain.

**FARID BINBESHR** received the bachelor's degree in computer science with scientific excellence from Hadhramout University, the master's degree in computer networks from the King Fahd University of Petroleum and Minerals (KFUPM), and the Ph.D. degree in information security from the Faculty of Computer Science and Information Technology, Universiti Malaya, Malaysia. Currently, he is a Postdoctoral Research Fellow with the Interdisciplinary Research Center for Intelligent Secure Systems, KFUPM. His research interests include computer networks, computer security, and artificial intelligence.

**KOO YUEN PHAN** received the Ph.D. degree in business information technology in Malaysia and the M.Sc. degree in information studies in Singapore. He is currently an Assistant Professor of computer science with the Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman. His research interests include the domains of information systems, information technology, business intelligence, and firm performance.

**CHIN SOON KU** received the Ph.D. degree from Universiti Malaya, Malaysia, in 2019. He is currently an Assistant Professor with the Department of Computer Science, Universiti Tunku Abdul Rahman, Malaysia. His research interests include AI techniques (such as genetic algorithm), computer vision, decision support tools, graphical authentication (authentication, picture-based password, graphical password), machine learning, deep learning, speech processing, natural language processing, and unmanned logistics fleets.