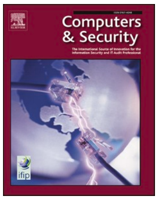
Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose



# Future of generative adversarial networks (GAN) for anomaly detection in network security: A review

Willone Lim [*], Kelvin Sheng Chek Yong, Bee Theng Lau, Colin Choon Lin Tan

*Faculty of Engineering, Computing, and Science, Swinburne University of Technology, Sarawak Campus, 93350 Kuching, Sarawak, Malaysia*

A R T I C L E I N F O

*Keywords:*
Network anomaly detection
Generative adversarial networks
Deep learning
Outlier detection

A B S T R A C T

Anomaly detection is crucial in various applications, particularly cybersecurity and network intrusion. However, a common challenge across anomaly detection techniques is the scarcity of data that accurately represents abnormal behavior, as such behavior is often detrimental to systems and, consequently, rare. This data limitation hampers the development and evaluation of effective anomaly detection methods. In recent years, Generative Adversarial Networks (GANs) have garnered significant attention in anomaly detection research due to their unique capacity to generate new data. This study conducts a systematic review of the literature to delve into the utilization of GANs for network anomaly detection, with a specific emphasis on representation learning rather than merely data augmentation. Our study also seeks to assess the efficacy of GANs in network anomaly detection by examining their key characteristics. By offering valuable insights, our research can aid researchers and practitioners in understanding the evolving landscape of network anomaly detection and the practical implementation of GANs while addressing the challenges in developing robust GAN-based anomaly detection systems.

*Abbreviations*

| | |
|---|---|
| GAN | Generative Adversarial Network |
| DAD | Deep Anomaly Detection |
| CPS | Cyber-Physical Systems |
| IoT | Internet of Things |
| PRISMA | Preferred Reporting Items for Systematic Reviews and Meta-Analyses |
| SLR | Systematic Literature Review |
| DNN | Deep Neural Network |
| DBSCAN | Density-Based Spatial Clustering of Applications with Noise |
| SVM | Support Vector Machines |
| FRNN | Fuzzy and Rough Set-based Nearest Neighborhood |
| LOF | Local Outlier Factor |
| IF | Isolated Forests |
| DNN | Deep Neural Network |
| RNN | Recurrent Neural Network |
| HTTP | Hypertext Transfer Protocol |
| CNN | Convolutional Neural Network |
| VAE | Variational Auto Encoder |
| FCN | Fully Connected Network |
| DBN | Deep Belief Network |
| PSO | Particle Swarm Optimization |
| PNN | Probabilistic Neural Network |
| GPU | Graphics Processing Unit |
| AE | Autoencoders |
| VAE | Variational Autoencoders |
| DAGMM | Deep Auto-encoding Gaussian Mixture Model |
| NIDS | Network Intrusion Detection System |
| IDS | Intrusion Detection System |
| ALAD | Adversarially Learned Anomaly Detection |
| FID-GAN | Fog-based Intrusion Detection |
| IGAN | Imbalanced Generative Adversarial Network |
| FNN | Feed-Forward Neural Network |
| ACGAN | Auxiliary Classifier Generative Adversarial Network |
| GANAD | Generative Adversarial Network Anomaly Detection |
| RCALAD | Regularized Complete Cycle Consistent GAN for Anomaly Detection |
| IF | Isolation Forest |
| GAN-SR | Generative Adversarial Network for Super-Resolution |
| FlowGAN | Flow Generative Adversarial Network |
| MLP | Multilayer Perceptron |
| Net-GAN | Network Generative Adversarial Network |
| IGAN | Inference Generative Adversarial Network |

[*] Corresponding author.
*E-mail address:* wilim@swinburne.edu.my (W. Lim).

https://doi.org/10.1016/j.cose.2024.103733
Received 8 November 2023; Received in revised form 2 January 2024; Accepted 17 January 2024
Available online 20 January 2024

## 1. Introduction

In modern systems, massive amounts of data are constantly transmitted over the network, moving from one network access point to one or more access points through hardware, software, and protocols. However, anomalies may arise within this process, potentially severely impacting the system and its environment (Lin et al., 2019). Therefore, to mitigate the impact, it is crucial to detect these anomalies promptly. The process of finding a system's anomalous behavior is called anomaly detection. The primary objective of anomaly detection is to differentiate between a system's expected and unexpected behavior. Anomaly detection techniques are widely used in many fields and applications, including network security, such as intrusion detection of potential attempts in network traffic (Koren et al., 2023). Therefore, anomaly detection is widely recognized as a crucial component within various decision systems.

Existing detection systems predominantly rely on pre-existing knowledge of attack patterns and normal network behavior (Dromard and Owezarski, 2020). Such systems, whether signature-based or behavior-based, necessitate ongoing updates to safeguard the network against the evolving attack vectors that consistently adapt to evade modern security measures (Lin et al., 2019). Nonetheless, constructing new signatures or updated profiles for these detectors demands significant time and resources, typically requiring the expertise of network specialists. As a result, these systems face limitations in effectively addressing novel network behaviors. Thus, various innovative detection techniques have been introduced to mitigate these challenges.

Anomaly detection techniques can be categorized as supervised, unsupervised, and mixed, with the most common approach being unsupervised anomaly detection (Smiti, 2020). The assumption is that data is normally distributed, and any data point that differs significantly from the expected behavior will be marked as an anomaly (Huyan et al., 2019). Unsupervised anomaly detection identifies abnormal patterns or observations in a dataset without requiring explicit labels for the anomalies. In supervised anomaly detection, a classifier will be trained using a dataset labeled as an 'anomaly' or 'normal' (Boukerche et al., 2020). One of the significant disadvantages of this process is the requirement to extensively train a dataset that includes both 'anomaly' and 'normal' observations. Due to the many types of anomalies, it is challenging for any algorithm to identify such patterns. Thus, the unsupervised approach is more popular because learning the expected behavior is considerably easier than learning the types of anomalies (Samariya and Thakkar, 2023).

Advanced techniques in deep learning, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Auto-Encoders (AEs) (Xia et al., 2022), have been effectively employed across various anomaly detection tasks, yielding positive results. With the surge in data complexity and size, these methods continually adapt to address the complexities through innovative deep learning algorithms tailored for anomaly detection. However, it is common to encounter datasets where most samples are normal, and a relatively small number of samples are abnormal, leading to imbalanced class distributions (Oksuz et al., 2021). Currently, the efficacy of Deep Anomaly Detection (DAD) methods often depends on extensive training samples, making data imbalance a notable burden in their application (Oksuz et al., 2021).

Furthermore, the absence of prior information, specifically regarding attack categories such as zero-day attacks, presents a significant challenge in the detection process (Blaise et al., 2020). Prompt detection of these attacks is essential to prevent substantial damage. This challenge is exacerbated by the existence of unknown vulnerabilities in network infrastructures and individual devices within Cyber-Physical Systems (CPS) and the Internet of Things (IoT), adding complexity to security solutions (Yaacoub et al., 2020). Additionally, the rapid development of 5G/6G networks and Cloud Services results in a considerable increase in the volume and bandwidth of cyber-attack traffic (Alnawayseh et al., 2022). Consequently, detecting these intrusions in real-time and maintaining consistent detection becomes increasingly challenging.

A recent trend in anomaly detection is the use of generative adversarial networks (GANs) (Bashar and Nayak, 2020). Generative adversarial networks are a type of unsupervised generative model that gained much attention from the research community. A well-trained GAN can create data that resembles real-world information, drawing from a learned data distribution. Comprising a generator and a discriminator model, GAN engages in a two-player zero-sum game situation, continually enhancing its capability in data generation and discrimination (Sabuhi et al., 2021). The appeal of GAN in anomaly detection research is twofold. First, they can aid in the generation of challenging-to-obtain anomalous data points. Second, they are well-suited to learn the data distribution representing the normal operation of a system, effectively serving as an anomaly or outlier detector.

Several issues and challenges related to the current network anomaly detection can be identified (Table 1). In the following sections, this paper will outline how the state-of-the-art GAN-based method tackles the specific issues (I) that have been identified.

## 2. Methodology

The current systematic literature review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, following the established standards outlined by Kitchenham (2014). The Systematic Literature Review (SLR) planning phase involves three key steps: recognizing the need for the systematic review, formulating the review protocol, and assessing the protocol's effectiveness. During the conducting phase, aligned with the established review protocol, we executed tasks such as identifying and selecting primary studies, extracting data from these studies, and synthesizing the accumulated data. All authors are involved in the primary study

**Table 1**
The issue faced by network anomaly detection.

| Issue (I) | Description |
|---|---|
| I1: Heavy reliance on pre-existing knowledge | Most current network anomaly detection systems heavily rely on pre-existing knowledge of attack patterns and normal network behavior. This dependency on historical data and predefined models can limit their ability to detect novel network behaviors effectively. |
| I2: Resources intensive and time-consuming | Constructing new signatures or updated profiles for existing detectors demands significant time and resources, typically requiring the expertise of network specialists. This process can be resource-intensive and time-consuming. |
| I3: Data imbalance | Imbalanced class distributions are common in network anomaly detection, where there are often many normal samples but only a few abnormal ones. Deep anomaly detection methods may depend on extensive training samples, making data imbalance a notable challenge. |
| I4: Absence of prior information on attack categories | The lack of prior information, specifically regarding attack categories like zero-day attacks, presents a significant challenge. Moreover, unknown vulnerabilities in network infrastructures and devices further complicate security solutions. |
| I5: Increased data complexity and lack of real-time detection | With the surge in data complexity and size, network traffic is continually growing, and detecting cyber-attack traffic in real-time and maintaining consistent detection becomes increasingly challenging. |
| I6: Data generation for anomalous patterns | Traditional anomaly detection approaches may struggle to generate challenging-to-obtain anomalous data points, which is essential for training and testing detection models. |

collection. The final phase concludes the systematic review by comprehensively presenting the gathered data and resultant findings.

### *2.1. Purpose of systematic review*

At present, Generative Adversarial Networks (GANs) represent a highly prominent research area, spawning a multitude of GAN-based anomaly detection approaches that have been widely employed in intrusion detection systems. The compelling capability of GAN to adeptly synthesize realistic data, coupled with their proficiency in representation learning, positions them as an attractive focal point for advancing anomaly detection methodologies (Cai et al., 2021). GAN provides benefits in anomaly detection by generating synthetic data for limited abnormal samples, learning intricate data features, adapting to changing distributions, detecting novel anomalies, and localizing specific anomalous patterns, thus improving accuracy and early detection (Bhattarai et al., 2020).

This work is centered around advancing research and the practical deployment of state-of-the-art Generative Adversarial Network (GAN) for anomaly detection within network security. The existing body of research has dedicated considerable effort to outlining various anomaly detection techniques and summarizing the advances in anomaly detection through deep learning. They only briefly introduce the application of GAN for anomaly detection. Di Mattia et al. (2019) have provided an overview of the three primary GAN-based anomaly detection methods and a performance analysis on publicly available datasets. However, this study falls short of providing a holistic synthesis of the most current developments and the potential trajectories that GAN technology might take in this field. Despite the potential, a notable significant gap exists concerning the GAN-based anomaly detection model within the network security domain. This gap points to the limited application and exploration of GANs' capabilities in effectively identifying anomalies and enhancing the security of network infrastructures. Hence, there is a pressing need for a comprehensive investigation that offers valuable insights into integrating GANs into anomaly detection within network security. The insights this study provides will hold immense significance, aiding researchers in understanding the latest advancements, the challenges, and the potential for future research within the domain of GAN-based network anomaly detection.

The paper's primary contribution evaluates different state-of-the-art GAN-based models applied to network anomaly detection. It does not limit itself to exploring and classifying GAN-based network anomaly detection. However, it offers further knowledge by considering each model's characteristics, their efficacy and applicability in network anomaly detection, and the limitations that hold them from performing effectively. The paper then deeply studies each GAN model to identify its weaknesses and provide recommendations for future studies.

To summarize, the main contributions of this paper are as follows:

1. This offers an up-to-date overview of the advancements in network anomaly detection, comprehensively examining various models based on conventional and deep neural network (DNN) approaches.
2. This paper provides a detailed classification of GAN roles in network anomaly detection.
3. This paper gives insights into the characteristics, effectiveness, and limitations of contemporary GAN-based models for network anomaly detection.
4. This paper uncovers important statistics on the existing GAN models and offers recommendations for future research.

### *2.2. Formulating the review protocol*

To systematically minimize the potential impact of researcher bias, it is essential to establish a comprehensive review protocol that outlines the procedural framework for conducting the systematic review. This protocol serves to define the subsequent elements precisely:

1. Research questions.
2. Search strategy.
3. Criteria for selecting studies.
4. Approach for extracting relevant data.
5. Methodology for synthesizing the extracted data.

### *2.3. Formulating the research questions*

The following research questions were formulated:

RQ1: What is the trend of network anomaly detection?
Motivation: To gain a deeper understanding of the advancements and progress made in network anomaly detection.
RQ2: What is the role of GAN in network anomaly detection?
Motivation: Investigate practical implementation of GANs for network anomaly detection.
RQ3: How effective GAN are for network anomaly detection?
Motivation: Identify hurdles to enhance reliable GAN-based anomaly detection in networks.

### *2.4. Search strategy*

Six electronic databases were searched for relevant studies, including IEEE Xplore, ScienceDirect, Scopus, ACM Digital Library, ProQuest, and Google Scholar. Backward searching was conducted, whereby the reference lists of relevant reviews and the included articles were also reviewed for additional included articles. The key search terms used were variations of ("anomaly" OR "anomalies" OR "anomalous" OR "abnormal" OR "outlier" OR "inconsistent" OR "variances") AND ("GAN" OR "GANs" OR "generative adversarial network" OR "generative adversarial networks" OR "GAN-based network" OR "GAN-based"). Fig. 1 illustrates the flow diagram of the SLR.

### *2.5. Study selection criteria*

The following research criteria, in Table 2, are adapted to ensure consistent evaluation and minimize subjectivity.

### *2.6. Data extraction strategy*

In order to optimize the data extraction process, we formulated a specialized data retrieval technique designed to methodically compile the essential information for each of our research inquiries (Table 3). We initiated the screening process by assessing the titles of the chosen studies to ensure the inclusion of solely pertinent research. Then, the information derived from these selected studies was evaluated to verify its relevance in addressing the predefined research questions. Furthermore, a comprehensive examination of the study information was conducted to ensure the thoroughness and accuracy of the data extraction.

### *2.7. Data synthesis*

In the data synthesis phase, we combine the gathered data from the data extraction forms to address the research questions. This comprehensive assembly of data provides insights into the best practices and architectures for anomaly detection using GANs.

### *2.8. Evaluation of the review protocol*

The protocol is a pivotal SLR component. It underwent evaluation by all authors, and the conclusive version of the protocol gained approval through multiple iterations. This final protocol was employed consistently during the execution phase of the SLR.

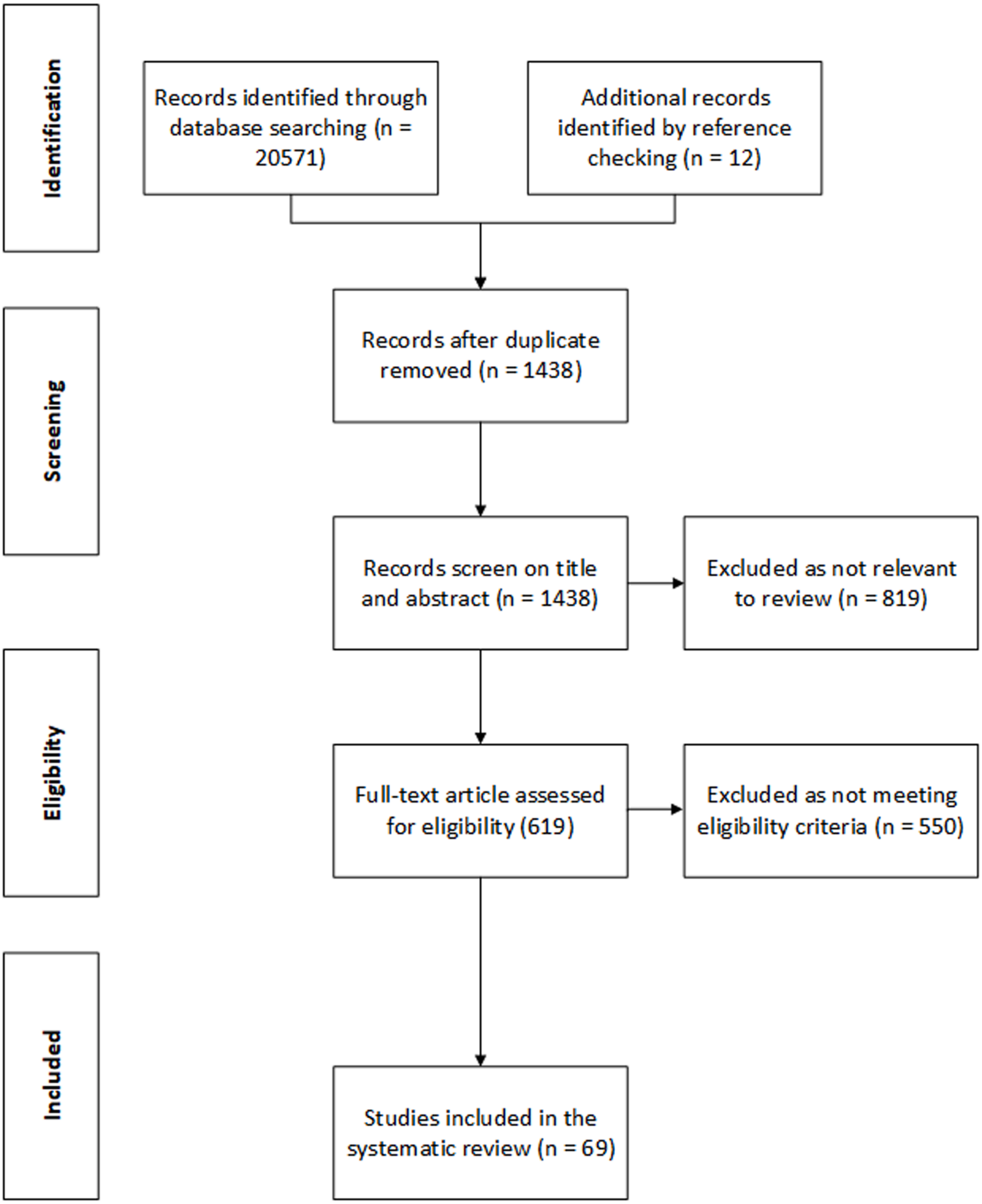

Identification
Screening
Eligibility
Included
Records identified through database searching (n = 20571)
Additional records identified by reference checking (n = 12)
Records after duplicate removed (n = 1438)
Records screen on title and abstract (n = 1438)
Excluded as not relevant to review (n = 819)
Full-text article assessed for eligibility (619)
Excluded as not meeting eligibility criteria (n = 550)
Studies included in the systematic review (n = 69)

**Fig. 1.** Flow diagram of the systematic review process.

**Table 2**
Inclusion and exclusion criteria.

| Criteria | Inclusion/ exclusion criteria | Description |
|---|---|---|
| Inclusion | IC1 | Research that is dedicated to GAN-based network anomaly detection system |
| | IC2 | Research on network anomaly detection methods based on machine learning |
| | IC3 | Research written in the English language |
| | IC4 | Research published from 2016 to 2023 |
| Exclusion | EC1 | Not a network anomaly-based detection paper |
| | EC2 | Not written in the English language |
| | EC3 | Not data augmentation |

**Table 3**
Data extraction method.

| | |
|---|---|
| Primary study information | Title of the Paper |
| | Authors |
| | Publication Year |
| | Journal/Conference/Book chapter |
| Research question alignment | Relevance to RQ1: Yes / No |
| | Relevance to RQ2: Yes / No |
| | Relevance to RQ3: Yes / No |
| Study Details | Model used |
| | Datasets Utilized |
| | Main Findings/Results |

### 2.9. Conducting the SLR

The execution phase of the SLR encompasses these four key stages: primary study search, primary study selection, primary study data extraction, and data synthesis. Throughout this process, we eliminated irrelevant or duplicated papers, adhering to the study selection criteria.

## 3. Results

### 3.1. Trend analysis of network anomaly detection?

Over the years, researchers and practitioners have dedicated substantial efforts to enhance the capabilities of network anomaly detection systems. These advancements have led to improved methods and tools for identifying and mitigating anomalies within computer networks.

#### 3.1.1. Traditional approaches

In the initial phase of network anomaly detection, conventional methods predominantly relied on supervised machine learning algorithms. For instance, one prevalent approach involved utilizing distance-based techniques, such as the one outlined in Xiong et al. (2011), which assessed data for anomalies by measuring the distances to nearest neighbors or clusters within the dataset. Distance-based techniques rely heavily on the choice of distance metrics, which can impact the detection accuracy. They may not effectively handle high-dimensional data and often assume that anomalies are distant from normal data points, which may not apply to all anomalies. Clustering-based strategies were also introduced, as shown by Blowers and Williams (2014), which introduced the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method for anomaly identification in networks. Clustering-based methods assume that anomalies are outliers in the clusters. They may struggle to identify anomalies that are part of small or dense clusters and might not work well with irregularly shaped clusters. Additionally, they are sensitive to the choice of parameters, such as the neighborhood size in DBSCAN.

Moving forward, Sadiq Ali Khan (2011) introduced an approach utilizing genetic algorithms to detect anomalies. This approach can be computationally expensive and may require substantial computational resources. They might not scale well enough to handle large and dynamic network datasets in real-time. Shone et al. (2018) employed random forest as a classifier to discriminate anomalies, but this approach has limitations in detecting anomalies that do not adhere to typical patterns in the data. It may also struggle with high-dimensional data and require careful parameter tuning. Mulay et al. (2010) proposed a method that combined Support Vector Machines (SVM) and decision trees to create a multi-classification anomaly detection system, constructing multi-classification SVMs through binary classification trees. However, combining SVMs and decision trees can be computationally intensive, particularly in multi-class scenarios. They may not perform optimally with imbalanced datasets, common in network anomaly detection.

In a separate development, Selvakumar et al. (2019) introduced an algorithm called Fuzzy and Rough Set-based Nearest Neighborhood (FRNN) for classifying network trace datasets. Although FRNN is proficient at managing data uncertainty, its suitability for highly dynamic network environments characterized by rapidly evolving anomaly patterns remains a concern. Several prominent machine learning methods have emerged based on density evaluation, including the Local Outlier Factor (LOF), Robust Covariance, and Isolated Forests (IF) (Fu et al., 2023). These methods have shown promise in mitigating the challenge of insufficient labeled data, providing a practical solution to an otherwise data-hungry problem. However, the landscape of network traffic is rapidly changing, marked by ultra-high dimensional data, and this progression has brought to light certain limitations of these density-based methods.

Though effective in specific scenarios, density-based approaches exhibit sensitivity to parameter settings (Liu et al., 2021a). The choice of these parameters can significantly impact their performance, making them less robust and potentially less reliable in practice. Moreover, they may encounter difficulties adapting to highly dynamic network environments where anomalies evolve rapidly (Lin et al., 2019). These methods often assume that anomalies are sparse, an assumption that may not hold for certain sophisticated attacks. As a result, the dynamic and complex nature of contemporary network traffic, coupled with the limitations of these conventional methods, necessitates exploring alternative approaches, such as deep neural networks, to tackle the challenges of modern network anomaly detection effectively.

#### 3.1.2. DNN-based approaches

Recent advancements in network anomaly detection have centered around utilizing deep neural networks (DNNs), which have gained widespread adoption in this domain. Initially, researchers explored the potential of recurrent neural networks (RNNs) for capturing temporal features in network data (Torres et al., 2016). However, the problem with the RNN detection model is that it has difficulty dealing with traffic behaviors that are not easily differentiable and certain cases of imbalanced network traffic. The detection model also struggles to detect some of the Hypertext Transfer Protocol (HTTP) traffic labeled as normal correctly, and it may have false alarm rates that are higher than expected during cross-validation.

In parallel, a novel approach was introduced that combined structural learning with graph neural networks, incorporating attention mechanisms to enhance the interpretability of detected anomalies (Deng and Hooi, 2021). However, the method for anomaly detection in high-dimensional time series data does not explicitly learn the structure of existing relationships between variables or use them to predict the expected behavior of time series. This limits their ability to capture complex inter-sensor relationships and detect and explain anomalies deviating from them.

Building upon the success of convolutional neural networks (CNNs), researchers introduced three distinct CNN architectures optimized for structural scalability to elevate the performance of network anomaly detection (Kwon et al., 2018). The evaluated CNN models occasionally outperform the Variational Auto Encoder (VAE) models but do not work better than the other deep learning models based on Fully Connected Network (FCN) and Seq2Seq-LSTM. The performance of the CNN model is generally worse than other deep learning models.

Another study proposed a network intrusion detection framework that harnessed the power of Deep Belief Networks (DBN) in conjunction with probabilistic neural networks, highlighting the superior efficacy of their combined approach (Zhao et al., 2017). The proposed model combines deep learning with the Particle Swarm Optimization (PSO) algorithm and the Probabilistic Neural Network (PNN) to address redundant information, large amounts of data, long training times, and the risk of falling into local optima in intrusion detection. While it appears effective based on the experimental results, the extent of its effectiveness is not quantified. Overall, deep neural networks involve training numerous parameters. The time required for training increases with larger network architectures. Computational resources, such as graphics processing unit (GPU) acceleration or big data processing technology, are needed.

Advancing further, DPLAN is a reinforcement learning technique designed to optimize the learning process for identifying unknown anomalies, effectively distinguishing marked abnormal data from unmarked cases (Pang et al., 2021). RDP is an unsupervised representation learning method that calculates data distances within a random projection space, achieved by training a neural network with random mappings (Wang et al., 2021). However, the training time for both models is relatively slow, which may limit its suitability for applications where real-time or near-real-time training is required. Although offline training is possible, the speed might not fully compensate for potential delays in certain applications.

DevNet is a methodology incorporating neural deviation learning to recognize abnormal score learning, enhancing anomaly representation by integrating neural networks, Gaussian priors, and Z-Score-based deviation loss functions (Pang et al., 2019). The method relies on having a few labeled anomalies and a prior probability, which may not always be available in real-world anomaly detection applications.

In addition to these approaches, techniques such as Autoencoders (AE) (Zhou and Paffenroth, 2017), Variational Autoencoders (VAE) (An and Cho, 2015), and Deep Auto-encoding Gaussian Mixture Models (DAGMM) (Zong et al., 2018) have found success in detecting abnormal data. However, these methods rely on modeling data distributions and deriving anomaly scores based on Gaussian mixture models. They only demonstrated on benchmark problems, and their performance on other real-world datasets was not evaluated. Kitsune is a plug-and-play intrusion detection system (NIDS) capable of autonomously learning to detect attacks within a local network without supervision (Mirsky et al., 2018). However, Kitsune relies on resource-intensive autoencoders, which require significant memory and processing power to train and execute, making them less suitable for devices with limited resources. The scalability of Kitsune or its performance in large-scale network environments is not analyzed, which can limit its effectiveness in such scenarios.

## 3.2. *Role of GAN in network anomaly detection*

### 3.2.1. *GAN architecture overview*

The GAN architecture is characterized by two essential models: generator and discriminator. The generator's primary responsibility is to produce synthetic data samples that accurately emulate the patterns observed in normal network traffic (Soleymanzadeh and Kashef, 2023). Achieving this involves the design of a generator capable of capturing the complexity inherent in the data distribution. In tandem, the discriminator, trained to distinguish real (normal) from fake (generated) data, assumes a central role in ensuring the overall efficacy of the anomaly detection system (Hong et al., 2019). The two models are trained in parallel, introducing a form of competition where each network tries to outperform the other (Aggarwal et al., 2021). The training process is complete when the two networks can no longer improve on their current state. This point is known as the Nash equilibrium. The objective of a GAN is to attain this equilibrium because, at that stage, the generated data from the generator model becomes indistinguishable from real data by the discriminator model (Yang et al., 2019). Consequently, the discriminator's output becomes a random guess regarding the authenticity of the input data.

The typical layout of GAN architecture is depicted in Fig. 2. In this setup, the generator takes a random sample from the latent space, denoted as $z$ as input. The generator then produces an output, represented as $G(z)$, which is subsequently evaluated by the discriminator alongside a sample drawn from the real data distribution.

The discriminator assigns a value to each sample, indicating whether it believes the sample to be real (assigned the value 1) or synthetic (assigned the value 0). These two outputs are crucial for assessing the performance of both models. Specifically, the generator is trained to minimize the function $\log(1 - D(G(z)))$, aiming to generate synthetic data that the discriminator cannot distinguish from real data (i.e., $D(G(z)) \approx 1$).

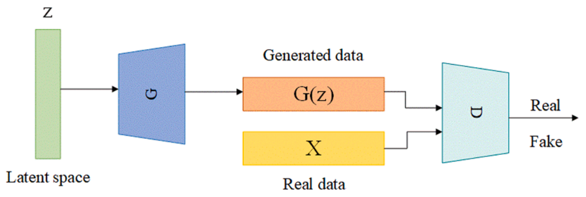


**Fig. 2.** GAN architecture (adapted from Sabuhi et al., 2021).

Simultaneously, the discriminator is trained to maximize the function $\log(D(\chi)) + \log(1 - D(G(z)))$. This objective guides the discriminator to maximize its ability to correctly identify real samples ($D(\chi)$) while also effectively distinguishing synthetic samples ($D(G(z))$).

### 3.2.2. *Classification of GAN roles*

GAN offers versatile capabilities, and their roles can be classified into several distinct categories, each addressing specific challenges in anomaly detection. Generally, the role of GAN can be classified into (i) generating abnormal data instances, (ii) generating normal and abnormal instances, (iii) learning normal behavior of a system, (iv) learning normal and abnormal behavior of a system and (v) learning complex data distributions. In this classification, applications i and ii fall under the category of data augmentation with GANs, while applications iii, iv, and v represent GANs' contribution to representation learning in the context of anomaly detection.

GAN is primarily created for data augmentation, which generates fresh data to supplement the existing dataset. They are also employed for representation learning, which involves learning data representations to facilitate information extraction when constructing classifiers or other predictive models (Zhao and Zhang, 2022). In networking, GAN is applied to network anomaly detection by leveraging their ability to model complex data distributions. In this scenario, a GAN generator and discriminator components can be employed to learn about the distribution of a specific data category, which could be normal or abnormal. Consequently, this acquired distribution knowledge can be applied to detect data instances that do not conform or deviate from the expected patterns. Most of the primary research studies have chosen to employ GANs to learn data representations rather than augment the datasets (Sabuhi et al., 2021). Moreover, the procedure of representation learning primarily centers on normal data. This choice is rooted in the inherent imbalance between normal and abnormal instances within the data. Creating a model that characterizes normal patterns rather than focusing on abnormal ones is often more practical. Additionally, by solely emphasizing the learning of the normal data distribution, the need for obtaining data related to abnormal conditions within the target system is eliminated (Sabuhi et al., 2021).

*Data augmentation.* Machine learning methods, particularly deep learning approaches, demand substantial data to excel in their intended tasks (Avci et al., 2021). Data augmentation, also referred to as oversampling, is employed to rectify the deficiency of data within a dataset, serving the dual purpose of preventing model overfitting and alleviating data imbalance (Xu et al., 2022). Data imbalance arises when there is a significant difference in the sizes of different classes within a dataset. For example, in a binary classification scenario, the class with fewer instances is termed the minority class, while the other is designated as the majority class (Arthur and Date, 2022). The training process associated with such an imbalanced dataset tends to be skewed in favor of the majority class, resulting in a classifier that exhibits higher accuracy for this class. Two approaches can be employed to tackle the challenge of imbalanced datasets: either the random removal of samples from the majority class to equalize class sizes (undersampling) or the introduction of artificially generated instances to increase the minority class (oversampling), achieved through suitable techniques (Jiang et al., 2019). The issue of imbalanced datasets is particularly vital in anomaly detection, primarily because obtaining data related to anomalous system behavior is challenging and resource-intensive (Y. Fu et al., 2022). Anomalous data instances are often scarce or absent. In such a scenario, GANs can provide a valuable solution by generating additional samples for the anomalous class.

*Representation learning.* The primary objective of GANs is to construct a generative model that generates realistic-looking data, achieved by sampling from the acquired data distribution. The potential of GANs in generating data with a high degree of realism was notably emphasized in the works of (Shin et al., 2023). Representation learning in anomaly detection exploits the ability of GANs to learn the distribution of specific data classes. Representation learning in the context of GAN for network anomaly detection refers to training GAN models to capture and understand the underlying patterns and structures within network traffic data (Li and Pi, 2020). It involves creating meaningful, compact, and abstract representations of the data that can be effectively used for identifying anomalies. GAN is designed to learn a rich and informative representation of network traffic data. The generator model within the GAN architecture is responsible for generating synthetic data that closely resembles normal data patterns (Syed Azahad and Shaik Hameeda, 2023). In the process, it learns to capture the data's complex details and statistical properties, which are essential for effective anomaly detection. GAN is trained to understand and encapsulate the normal behavior of network traffic. Exposing the GAN to a vast amount of legitimate network data allows it to model the distribution corresponding to normal operation (Xia et al., 2022). This learned representation serves as a baseline for identifying deviations from normalcy. Network traffic data often exhibits complex and dynamic distributions. GANs are well-suited for modeling such complex data distributions. They adapt to the evolving patterns and behaviors within the network data, ensuring that the learned representation remains up-to-date and can accommodate changing anomalies (X. Li et al., 2023). GANs continually update their learned representations as they encounter new data. This adaptability enables them to effectively handle zero-day attacks and evolving network behaviors (Umer et al., 2021). When previously unseen anomalies occur, GANs can detect them based on deviations from the established representation.

### 3.3. *Effectiveness of GAN in network anomaly detection*

Generally, GAN is a deep learning model that can generate realistic data from random noise. They consist of two competing networks: a generator that tries to create fake data and a discriminator that distinguishes between real and fake data. GAN has been applied to anomaly detection, identifying data points deviating from the normal distribution (Gui et al., 2023). However, anomaly detection is challenging because anomalies are rare, diverse, and unpredictable. Traditional methods, such as statistical tests, clustering, and classification, often rely on strong assumptions or prior knowledge about the data distribution, which may not hold in complex and dynamic scenarios (Wang et al., 2021b).

In comparison, Variational Autoencoders (VAEs) also adopt a dual-network structure, comprising an encoder and a decoder. VAEs focus on learning latent representations of data by mapping input data into a probabilistic latent space, from which they generate new samples. In contrast to GANs, VAEs emphasize data reconstruction and the acquisition of continuous, probabilistic latent representations (Li et al., 2021). VAEs offer a structured framework for capturing uncertainties within generated data, exhibiting proficiency in generating diverse samples while enabling controlled data generation through manipulations in the latent space.

However, VAE struggles to capture highly complex data distributions prevalent in network data, especially when anomalies exhibit diverse and complex patterns (Zavrak and Iskefiyeli, 2020). This limitation can impede their accuracy in distinguishing between anomalies and normal network behavior, potentially resulting in reduced accuracy. VAEs' sensitivity to noise and uncertainties within the data can also pose challenges, leading to difficulties in effectively separating relevant anomalies from noise, potentially resulting in false positives or negatives (Wang et al., 2020a, 2020b).

While GANs emphasize generating high-quality, realistic samples and are proficient in scenarios with data scarcity, VAEs prioritize learning latent representations and probabilistic modeling for generating diverse samples with controllable features. GANs require careful training to ensure stability and diversity in generated data, whereas VAEs inherently learn probabilistic distributions of data, making them more interpretable and capable of capturing uncertainties in generated samples. GANs might excel in producing more realistic but less interpretable data, while VAEs could provide a more structured approach with inherent uncertainty estimation, aiding in understanding the generated data's reliability (Gonzalez et al., 2023).

Network anomaly detection is a critical component of cybersecurity, where GAN has been applied to identify abnormal patterns in network traffic. However, the effectiveness of a typical GAN in this context is limited. Generally, a typical GAN faces three prominent challenges: imbalanced data, the generation of realistic anomalies, and resource-intensive processes that impede their overall performance. Network traffic data often exhibits a significant class imbalance, with most traffic being normal, while network anomalies represent a rare minority (Liu et al., 2021a). This imbalance poses a significant challenge for typical GAN, which may struggle to generate meaningful anomalies (Kim et al., 2020). The natural bias of GAN toward the majority class can result in ineffective modeling of rare but critical network anomalies. As a result, they tend to produce anomalies that resemble the majority class, making them less effective in capturing the subtleties of rare and critical network threats.

Moreover, generating realistic network anomalies is challenging for typical GAN (Xia et al., 2022). Anomalies can exhibit complex patterns that are difficult to model accurately (Schmidl et al., 2022). Standard GAN produces anomalies that are easily distinguishable from actual anomalies (Xia et al., 2022), diminishing their effectiveness for real-world applications. Moreover, the issue of mode collapse in GAN exacerbates the problem. Mode collapse can lead to the generation of repetitive or stereotypical data samples, potentially causing the model to miss many potential threats due to its inability to diversify anomaly patterns effectively (Pan et al., 2022).

The computational demands of training and deploying GAN for network anomaly detection can be challenging, particularly in large-scale networks (Das et al., 2022). These computational requirements can hinder the real-time processing and analysis of network traffic, leading to concerns about scalability and timely threat detection. The latency for GAN to process and analyze network traffic may not meet the real-time demands of some applications, potentially leaving networks vulnerable to emerging threats. Therefore, researchers have incorporated different methods to address the limitations of typical GAN.

The comprehensive examination of how state-of-the-art GAN models effectively tackle the issues (I) identified in network anomaly detection is provided below.

#### 3.3.1. *Issue 1: heavy reliance on pre-existing knowledge*

*3.3.1.1. Auxiliary classifier generative adversarial network (ACGAN).* Traditional network anomaly detection systems depend significantly on existing information about attack patterns and behavior (Mills et al., 2022). This reliance on historical data limits their efficacy in recognizing unfamiliar network behaviors. To address this limitation, Yuan et al. (2020) proposed ACGAN, that leverages edge computing and data augmentation to improve the precision and effectiveness of intrusion detection. The proposed system is designed to be deployed on edge nodes within a smart home environment. Edge computing allows data processing to occur closer to the source of data, reducing latency and bandwidth usage. This is particularly beneficial where real-time detection and response to potential intrusions are crucial. This also allows the model to effectively detect intrusions even without extensive pre-existing knowledge of that particular network environment. Network traffic data is transformed into images for network traffic analysis. Then CNN is applied to these images to classify network traffic into various categories, including distinguishing normal traffic from anomalies. CNNs are capable of automatically learning and extracting features from input data, reducing the need for pre-existing knowledge. To address the issue of limited and imbalanced datasets, which is a common challenge in intrusion detection, AC-GAN was employed for data augmentation. The AC-GAN is used to generate synthesized

**Table 4**
Evaluation Metrics Comparison of GAN Models Addressing Issue 1.

| GAN model | Datasets characteristics | Performance |
|---|---|---|
| ACGAN | UNSW-NB15<br>• Size: 175,341 records<br>• Type: Cyber Attacks<br>• Description: Real-world network traffic dataset | Accuracy: 96 %<br>Precision: 96 %<br>Recall: 98 %<br>F1-Score: 0.970 |
| N-GAN | CIC-IDS2017 network<br>• Size: 2830,743 records<br>• Type: Network Traffic<br>• Description: Real network traffic captured in controlled environment, containing multiple attack types, labeled data | Detection rate (DR..): 96.4 %<br>Area under the curve (AUC): 95 % |

samples of network traffic, effectively expanding the dataset and providing more diverse examples for CNN to learn from. The results demonstrated the effectiveness of the proposed system, particularly for minor categories of network traffic, where the precision was improved by 12 %. In tasks involving binary classification, the precision achieved in discerning between normal and anomalous traffic was 96 %.

*Weaknesses.* However, the extent to which ACGAN resolves the heavy reliance on pre-existing knowledge is limited. While it generates synthetic data, it primarily augments the existing dataset without guaranteeing the introduction of entirely new patterns or the capability to detect previously unseen anomalies effectively. There is a risk that the synthetic data might not fully capture the complexity and diversity of real-world network anomalies, potentially constraining the model's ability to adapt to entirely novel threats. The study does not explicitly discuss the potential challenges of handling class-imbalanced datasets, such as the risk of model bias toward the majority class. Assessing the effectiveness of class balance achieved by GAN and its impact on overall model performance is crucial. Moreover, the study outlines the architecture of the CNN but does not delve into the model's complexity for deployment on resource-constrained edge devices commonly found in smart homes. Ensuring model efficiency is vital for real-time intrusion detection, especially in edge computing environments. Given the intended deployment of this scheme on the edge nodes of smart homes, considerations of real-time performance, including latency and resource usage, become critical. Extensive testing in real-world smart home environments is essential to accurately evaluate these factors and assess the feasibility of the proposed model in practical settings. Furthermore, deploying intrusion detection systems in real-world scenarios entails challenges related to system maintenance and updates. The study does not address these practical deployment challenges or provide insights into considerations for ongoing system management. Further exploration of model complexity is necessary to enhance model efficiency for resource-constrained edge devices in smart homes. This can involve optimizing the architecture, implementing model compression techniques, or developing lightweight models tailored for edge computing environments (Han, et al., 2020). Future work should also delve into the practical challenges of deploying intrusion detection systems in real-world scenarios. This includes strategies for system maintenance, updates, and adaptability to evolving threats, ensuring the long-term effectiveness of the deployed system.

*3.3.1.2. Novel network intrusion detection technique based on generative adversarial networks (N-GAN).* Iliyasu and Deng (2022) introduced N-GAN, a new approach to network intrusion detection that addresses the reliance of existing network data. Unlike ACGAN which mitigates reliance on pre-existing knowledge by balancing class distribution through GAN-generated samples. N-GAN tackles the reliance on pre-existing knowledge by incorporating a few malicious samples, enabling the model to learn representations of new attacks weakly supervised. It focuses on learning effective representations for new or evolving attacks rather than balancing data distribution. The system focuses on the reconstruction error, Wasserstein distance-based GANs and autoencoder-driven deep learning models. N-GAN is a weakly supervised technique that incorporates a small number of known malicious samples during the training process. Incorporating malicious samples provides the model with prior knowledge about the distribution of anomalies, which can lead to improved representations. This also diminishes the necessity for substantial prior knowledge of network intrusions and enhances the model's ability to effectively adapt to emerging threats. Learning from regular data and malicious samples, N-GAN aims to enhance the quality of learned representations. This, in turn, can improve the model's robustness in distinguishing anomalies from regular network traffic.

*Weaknesses.* However, the N-GAN model faces limitations related to the diverse nature of cyberattacks. In the study, where novel attacks were assessed, the model's performance resembled baseline models. This suggests that the model may have difficulty learning a singular representation for various attacks with limited sample diversity, potentially limiting its ability to generalize to a wide range of unknown attacks. While incorporating a few malicious samples during training aids in learning representations of new attacks, the model's ability to generalize to entirely novel or evolving attack types might be limited. The model might struggle with entirely unprecedented threats or attacks that significantly differ from the few samples seen during training. Furthermore, the effectiveness of the N-GAN model is heavily dependent on the availability of even a small number of malicious samples for training. In cases where only a few samples are accessible for specific attacks, the model can significantly enhance detection for that attack. In real-world scenarios, securing enough samples for known attacks may be challenging, potentially restricting the model's practical applicability. While the model exhibits proficient performance in detecting variants of existing attacks, it is crucial to acknowledge that this success centers on the similarity of attributes between attack variants. Real-world attacks can possess highly diverse attributes, presenting a significant challenge for the model in effectively identifying all variants. Additionally, it is important to highlight that the experiments conducted are based on controlled datasets. The complex nature of real-world cyber threats demands assessing the model's performance in an operational environment where evolving and sophisticated attacks are prevalent.

Future research should investigate techniques for learning more robust representations of attack patterns to improve the model's ability to generalize to diverse and novel cyberattacks. This can involve developing adaptive approaches that capture the common features of different attack types, even with limited samples. Given the model's reliance on available malicious samples, researchers should explore data augmentation strategies to artificially increase the diversity of the training dataset (Al Olaimat et al., 2020). Techniques such as oversampling, undersampling, and synthetic data generation can be employed to ensure that the model has sufficient data for training, even for less common attacks (Park and Park, 2021). Research can delve into attribute-agnostic anomaly detection techniques that emphasize identifying unusual behavior patterns rather than specific attributes (Kumar and Sundaram, 2021). Researchers should explore techniques to mitigate the risk of overfitting by employing regularization methods or transfer learning approaches to ensure the model can generalize to new and unseen threats (Siahpour et al., 2022) (Table 4).

### *3.3.2. Issue 2: resources intensive and time-consuming*

*3.3.2.1. Bidirectional generative adversarial network (BiGAN).* Network anomaly detection often involves resource-intensive and time-consuming processes due to the complexity of analyzing vast amounts of network traffic data (Kaplan and Alptekin, 2020). Resource-intensive tasks in this domain typically involve computationally demanding algorithms or models that necessitate substantial computational power, memory, or storage capacities. Processing and analyzing large volumes of network traffic data to distinguish normal behavior from potentially malicious activities often demand extensive computational time. These resource-intensive and time-consuming aspects are critical considerations in designing efficient and responsive network anomaly detection systems, particularly in real-time network monitoring environments where quick and accurate detection of anomalies is crucial. The BiGAN exhibits the unique ability to learn anomalies by adding additional encoder that translates input data into a latent representation while simultaneously training a generator and discriminator (H. Zenati, Foo, et al., 2018). This innovative approach eliminates the need for computationally intensive latent representation recovery during testing. The model's design significantly accelerates testing time by several hundred-fold. This improvement directly targets the time-consuming nature commonly associated with numerous machine learning models. Unlike conventional GANs, where the discriminator solely assesses input data (real or generated), BiGAN's discriminator evaluates the input data and its corresponding latent representation, whether it originates from the generator or the encoder. Regarding resource demands, employing GANs enables the model to produce artificial network traffic samples, effectively broadening the dataset and offering a wider array of instances for the model's learning process. This diminishes the need for large quantities of actual data from real-world sources during training, resulting in a reduced demand for resources by the model. The model's primary objectives are to model the distribution of normal data samples, reconstruct them from their latent representations, and distinguish them as originating from the genuine data distribution. The model improved significantly, demonstrating inference times that were 700–900 times faster.

*Weaknesses.* The incorporation of an additional encoder within the GAN framework adds complexity to the model architecture. This increased complexity can lead to more demanding computational requirements during training and inference, potentially offsetting the intended reduction in resource intensiveness. Moreover, the model restricts its utilization of anomalous records, focusing exclusively on the normal records within the dataset during the training phase. This selectivity potentially constrains the model's capacity to detect anomalies deviating from the training data. Additionally, the threshold for distinguishing normal traffic from anomalies is determined based on the validation set, which does not include abnormal samples. This can potentially affect the model's ability to classify anomalies accurately. The simulated abnormal flows are generated with some deviation from the target class, irrespective of real anomalies. While this approach can provide some generalization, it does not fully represent the complexities of real-world network anomalies. Regarding performance evaluation, the model predominantly centers around the accuracy metric. While it holds significance, a more holistic evaluation can benefit from including other crucial metrics, such as precision, recall, and the F-score, particularly in anomaly detection tasks (Tharwat, 2018). Furthermore, while the model detection process is fast (45 µs on average), validating the model's real-time capability in a real-world network environment is important, which can introduce additional complexities and challenges (Ariyaluran Habeeb et al., 2019).

Future work includes exploring methods to incorporate anomalous records during training, optimizing threshold determination techniques, improving the realism of simulated abnormal data generation, and expanding the range of performance metrics for a more comprehensive evaluation. Additionally, transitioning from controlled environments to real-world network settings to validate the model's real-time capability will be crucial for assessing its practical utility in dynamic network security scenarios.

*3.3.2.2. Fog-based intrusion detection (FID-GAN).* Araujo-Filho et al. (2021) introduced an unsupervised Intrusion Detection System (IDS) tailored for cyber-physical systems and optimized for a fog computing architecture, resulting in enhanced detection rates. The model computes a reconstruction loss based on the reconstruction of data samples mapped to the latent space to achieve higher detection rates. The fog-based architecture is characterized by bringing computational resources closer to the endpoint nodes. This proximity is essential for meeting low-latency requirements in intrusion detection. Proximity ensures that data processing occurs nearby, reducing both time and resource consumption. The distinctive aspect of this model involves training an encoder to accelerate reconstruction loss computation, effectively addressing computational time issues. This enhancement enables faster anomaly detection and identification of potential cyber threats. Experimental results demonstrate that the proposed solution operates at least 5.5 times faster at baseline and achieves higher detection rates across the three examined datasets.

*Weaknesses.* Although FID-GAN employs a fast-mapping encoder to

**Table 5**
Evaluation metrics comparison of GAN models addressing issue 2.

| GAN model | Datasets characteristics | Performance |
|---|---|---|
| BiGAN | KDDCUP99<br>• Size: 1.4 million records<br>• Type: Network Traffic<br>• Description: Simulated network traffic dataset | Accuracy: 90 %<br>Precision: 85 %<br>Recall: 99 %<br>F1-Score: 0.908 |
| FID-GAN | NSL-KDD<br>• Size: 125,973 records<br>• Type: Network Traffic<br>• Description: Network intrusion detection dataset | Area under the curve (AUC): 80 % |
| GANAD | KDDCUP99<br>• Size: 805,050 records<br>• Type: Network Traffic<br>• Description: Simulated network traffic dataset<br>NSL-KDD<br>• Size: 148,517 records<br>• Type: Network Traffic<br>• Description: Simulated network traffic dataset<br>UNSW-NB15<br>• Size: 257,673 records<br>• Type: Network Traffic<br>• Description: Simulated network traffic dataset | KDDCUP99<br>• Precision: 97 %<br>• Recall: 97 %<br>• F1-Score: 0.9755<br>NSL-KDD<br>• Precision: 96 %<br>• Recall: 96 %<br>• F1-Score: 0.9581<br>UNSW-NB15<br>• Precision: 95 %<br>• Recall: 95 %<br>• F1-Score: 0.9482 |

map data patterns to latent spaces efficiently, the process might still pose computational challenges, especially if the encoder is complex or computationally intensive. Mapping high-dimensional data to latent spaces might demand substantial computational resources, potentially leading to time-consuming operations. The study also introduces additional components, such as spectral normalization and conditional entropy regularization, to enhance the core ALICE model, but their dependencies and potential limitations need careful consideration. Notably, the model's effectiveness appears to vary according to the data type employed without an extensive discussion of its adaptability to diverse data modalities. Finally, while acknowledging the challenges of training GANs and noting the field's rapid progress, the study briefly touches upon these issues, but a more in-depth exploration of the specific challenges and limitations encountered during training can offer valuable insights.

Further research can concentrate on optimizing the computational efficiency of the networks used within FID-GAN. Techniques such as model compression, pruning, or leveraging more lightweight architectures specifically tailored for intrusion detection could be explored to reduce computational overhead. Additionally, researchers should delve deeper into studying the components introduced to enhance the model, such as spectral normalization and conditional entropy regularization (Liu et al., 2021b). Investigating their dependencies, assessing their applicability across various use cases, and identifying potential limitations will aid in refining the model and maximizing its practical utility. The influence of data types on the model's performance should be extensively examined, focusing on understanding the factors driving performance variations across different data modalities. This analysis will contribute to developing a more adaptable and consistent IDS.

*3.3.2.3. Generative adversarial network anomaly detection (GANAD).* Existing GAN-based anomaly score methods built upon the generator network are designed for data synthesis, which would get undesired performance on the anomaly detection task. Their unstable performance makes detection tasks challenging. Therefore, Fu et al. (2023) proposed a GAN-based approach, GANAD, specifically designed for anomaly identification rather than data synthesis. It uses a similar auto-encoder architecture to address the time-consuming problem of traditional generator loss computation. This architecture allows for efficient computation and reduces the time taken for training the model. The model replaces the Jensen–Shannon (JS) divergence with the Wasserstein distance and incorporates a gradient penalty to enhance training stability and make it less resource-intensive. The training strategy of GANAD helps to learn minority abnormal distribution from normal data better, resulting in improved detection precision. This also ensures that the model can effectively learn from a smaller dataset reducing the need for large amounts of training data. In terms of performance, GANAD demonstrates superior anomaly detection performance compared to other models like ALAD, and FID-GAN by around 2–3 %. It excels in accuracy and demonstrates superior performance in a multi-classification setting. Specifically, GANAD achieves higher precision, recall, and F1-scores, indicating its effectiveness in identifying anomalies in network traffic. Additionally, GANAD proves more time-efficient than other GAN-based methods, especially when handling larger datasets, demonstrating its computational advantage in detecting anomalies within a shorter duration. Ablation studies confirm the effectiveness of GANAD's overall framework in maintaining balanced performance metrics across datasets, showcasing its consistency and robustness in anomaly detection.

*Weaknesses.* However, several critical considerations that should guide future work can be highlighted. The performance evaluations provided in the study are based on specific datasets, namely KDDCUP'99, NSL-KDD, and UNSW_NB15. Researchers must be mindful of the limitations of dataset dependency. The effectiveness of a model on one dataset may not necessarily extend to other datasets or real-world scenarios, urging the need for more extensive evaluation across diverse data sources. While the study emphasizes performance improvements, it does not address the challenges associated with imbalanced datasets. The impact of imbalanced class distributions on evaluation metrics and the model's performance across different data distributions should be explored in future research. The research underscores the significance of latency constraints in anomaly detection. Although the proposed approach aims to reduce time consumption, its actual time efficiency may vary depending on the specific hardware and infrastructure where it is deployed. Hence, future work should delve into practical considerations regarding the method's real-world deployment and its adaptability to various technological environments. Moreover, while the study asserts that the approach excels at learning minority abnormal distribution from normal data patterns, its generalizability to different types of network anomalies requires further exploration.

Future research should further address several key avenues to enhance its utility and adaptability. Firstly, to mitigate the limitations associated with dataset dependency, it is imperative to expand the evaluation of GANAD to encompass a broader range of datasets representative of diverse real-world scenarios. This inclusive approach will validate the model's robustness and ability to generalize effectively beyond specific data sources. Future work should delve into techniques that address class imbalances, aiming to provide more comprehensive insights into the model's performance across different data distributions to tackle the challenge of imbalanced datasets. Investigating approaches like class weighting or sampling methods can aid in addressing these challenges (Zhao et al., 2020). Future studies should explore the model's real-time performance in various hardware and infrastructure settings to ensure its practical viability in many technological environments. This could involve optimizing the model for specific hardware configurations and evaluating its resource efficiency. Investigating its performance across various network intrusion categories will provide a clear understanding of its utility in real-world security applications. Regarding scalability, researchers should focus on expanding the model's capabilities to accommodate larger datasets and the complexities of real-world network traffic scenarios (Table 5).

### 3.3.3. Issue 3: data imbalance

*3.3.3.1. Flow generative adversarial network (FlowGAN).* In network anomaly detection, it is common to encounter imbalanced class distributions, with an abundance of normal samples and a scarcity of abnormal instances. Deep anomaly detection techniques often rely on a large volume of training data, posing a significant challenge due to this imbalance in data distribution. Wang et al. (2019) proposed FlowGAN for identifying unbalanced network encrypted traffic. It addresses the problem of class imbalance in traffic classification by leveraging GAN's data augmentation capabilities to generate synthetic traffic data for classes with few samples. The model consists of several key stages aimed at mitigating data imbalance issues. Initially, the raw PCAP files undergo pre-processing steps such as filtering to remove unnecessary information, truncation to standardize packet length, and normalization to achieve uniform input sizes for the subsequent GAN model training. The GAN model is then trained in three distinct steps, training the discriminator with real data, fixing its parameters, and training the generator to produce authentic traffic samples. Notably, the generator and discriminator are concurrently trained in a competitive manner to enhance the generation of realistic samples while improving the discriminator's ability to distinguish between real and generated data. The third stage involves data balancing, where FlowGAN is utilized to mix generated and real samples, specifically focusing on balancing each category, particularly the minor classes within the dataset. The model aims to rectify data imbalance in encrypted traffic classification by using FlowGAN training to augment the smaller class, thereby creating a more equitable representation of the various classes. The performance of

**Table 6**
Evaluation metrics comparison of GAN models addressing issue 3.

| GAN model | Datasets characteristics | Performance |
|---|---|---|
| FlowGAN | ISCX VPN- nonVPN<br>• Size: 206,688 records<br>• Type: Network Traffic<br>• Description: Dataset containing VPN and non-VPN network traffic | Accuracy: 99 %<br>Precision: 99 %<br>Recall: 99 %<br>F1-Score: 0.991 |
| GAN RF | CICIDS 2017<br>• Size: Large volume<br>• Type: Cyber Attacks<br>• Description: Real-world network traffic dataset for cybersecurity | Accuracy: 99 %<br>Precision: 98 %<br>Recall: 93 %<br>F1-Score: 0.950 |
| GAN-SR | UNSW-NB15<br>• Size: 175,341 records<br>• Type: Network Traffic<br>• Description: Intrusion detection dataset with diverse attacks | UNSW-NB15<br>• Precision: 85 %<br>• Recall: 52 %<br>• F1-Score: 0.625 |
| | SWaT<br>• Size: 19,237 records<br>• Type: Network Traffic<br>• Description: Intrusion detection dataset with diverse attacks | SWaT<br>• Precision: 93 %<br>• Recall: 90 %<br>• F1-Score: 0.928 |
| | Gas Pipeline<br>• Size: 97,019 records<br>• Type: Network Traffic<br>• Description: Intrusion detection dataset with diverse attacks | Gas Pipeline<br>• Precision: 95 %<br>• Recall: 98 %<br>• F1-Score: 0.910 |
| IGAN | NLS-KDD<br>• Size: Large volume<br>• Type: Network Traffic<br>• Description: Intrusion detection dataset with diverse attacks | Accuracy: 89 %<br>Precision: 98 %<br>Recall: 89 %<br>F1-Score: 0.934 |
| IGAN-IDS | NSL-KDD<br>• Size: 148,517 records<br>• Type: Network Traffic<br>• Description: Intrusion detection dataset with diverse attacks | NSL-KDD<br>• Accuracy: 84 %<br>• F1-Score: 0.841<br>• AUC: 95 % |
| | UNSW-NB15<br>• Size: 2540,044 records<br>• Type: Network Traffic<br>• Description: Intrusion detection dataset with diverse attacks | UNSW-NB15<br>• Accuracy: 82 %<br>• F1-Score: 0.828<br>• AUC: 97 % |
| | CICIDS2017<br>• Size: 2827,829 records<br>• Type: Network Traffic<br>• Description: Intrusion detection dataset with diverse attacks | CICIDS2017<br>• Accuracy: 99 %<br>• F1-Score: 0.99<br>• AUC: 99 % |
| NIDS | NSL-KDD<br>• Size: 125 973 records<br>• Type: Network Traffic<br>• Description: Intrusion detection dataset with diverse attacks | NSL-KDD<br>• Accuracy: 90 %<br>• Precision: 85 %<br>• Recall: 97 %<br>• F1-Score: 0.909 |
| | UNSW-NB15<br>• Size: 175 341 records<br>• Type: Network Traffic<br>• Description: Intrusion detection dataset with diverse attacks | UNSW-NB15<br>• Accuracy: 91 % |
| | IoT-23<br>• Size: 23 145 records<br>• Type: Network Traffic<br>• Description: Network traffic from IoT | IoT-23<br>• Accuracy: 93 % |

**Table 7**
Evaluation metrics comparison of GAN models addressing issue 4.

| GAN model | Datasets characteristics | Performance |
|---|---|---|
| N-GAN | CIC-IDS2017 network<br>• Size: 2830,743 records<br>• Type: Network Traffic<br>• Description: Real network traffic captured in controlled environment, containing multiple attack types, labeled data | Detection rate (DR..): 96.4 %<br>Area under the curve (AUC): 95 % |

FlowGAN was evaluated using a Multilayer perceptron (MLP) based network traffic classifier, and experimental results showed that FlowGAN outperformed both unbalanced and balanced datasets regarding data augmentation. Particularly, FlowGAN demonstrated enhanced recognition performance on smaller samples. When compared with the unbalanced dataset, the Precision, Recall, and F1-score exhibited an average increase of 13.2 %, 17.0 %, and 15.6 %, respectively. In comparison, when compared with the balanced dataset, there was an average increase of 2.15 % in Precision, 2.06 % in Recall, and 2.12 % in F1-score.

*Weaknesses.* While FlowGAN exhibits significant improvements in recognizing data with limited samples, its effectiveness may be tied to the specific dataset and problem it applied to. The generalizability of these improvements to other datasets and diverse applications remains unexplored. FlowGAN's performance enhancement primarily stands out compared to imbalanced and oversampled datasets. This implies that its efficacy can be closely linked to the dataset characteristics it trained on. Consequently, the model's performance in varying real-world scenarios warrants further investigation to establish its practical versatility. While the FlowGAN method demonstrates its ability to improve true positive rates, assessing its impact on false positives is equally crucial. High false positive rates can have disruptive and costly consequences, especially in operational security contexts. Future research should thoroughly examine the model's false positive rates to assess its practical utility comprehensively.

Efforts should be made to enhance the model's robustness by recognizing a wide range of traffic characteristics. This could involve developing adaptive mechanisms that can adjust to variations in traffic patterns and ensure consistent performance across diverse applications. Future research should focus on developing strategies to reduce false positives while maintaining high true positive rates, striking a balance between precision and recall. Conducting experiments and tests in real-world operational settings is essential to validate the model's practical utility. This includes evaluating its performance in dynamic and complex network environments, where the consequences of misclassification can be significant.

*3.3.3.2. Imbalanced generative adversarial network intrusion detection system (IGAN-IDS).* In contrast, Huang and Lei (2020) devised a solution, IGAN-IDS, to tackles the class imbalance issue and simulates unknown anomalies. It also involves the generation of new, representative instances for underrepresented minority classes. This was achieved by integrating an imbalanced data filter and incorporating convolutional layers into the conventional GAN framework. It comprises three modules such as Feature Extraction (FE), IGAN, and Deep Neural Network (DNN). The FE module transforms raw network attributes into latent feature vectors using a Feed-forward Neural Network. The IGAN component generates new data samples expressed in the latent space. These samples are designed to represent the minority class, effectively addressing the class imbalance issue. The final intrusion detection is done by a DNN, including convolutional and fully connected layers. This network analyzes the generated and transformed data to detect network intrusions. Experimental evaluation demonstrates IGAN-IDS achieves improvements in Precision, Recall, F1 Score, and AUC by at least 1 %, 6 %, 10 %, and 1 %, respectively, compared to state-of-the-art methods on NSL-KDD. The robustness of IGAN-IDS is further examined by varying generated ratios and imbalance ratios, showing stable and effective performance even under different conditions. Overall, IGAN-IDS effectively handles class imbalance by generating samples and demonstrates efficient intrusion detection capabilities.

*Weaknesses.* However, the evaluation primarily focuses on three datasets (NSL-KDD, UNSW-NB15, and CICIDS2017). The limitations of the model's generalization to a broader range of network intrusion detection datasets are not thoroughly explored. The model touches upon the effect of the generated ratio (r) but overlooks the impact of other hyperparameters on the performance. A more comprehensive analysis of the sensitivity of IGAN-IDS to various hyperparameters is essential to fine-tune and optimize the model effectively. Furthermore, the model focuses on benchmark datasets and does not address real-world complexities in network intrusion detection, such as zero-day attacks, adversarial examples, and data drift. The study suggests controlling the generated ratio within a specific range ($r = 0.5:1$ to $r = 1.5:1$). The choice of this range is not thoroughly justified, and a more in-depth analysis of the optimal range could be informative.

Future work should expand the evaluation to encompass more diverse network intrusion detection datasets to enhance the model's robustness. This can involve testing the model on a broader array of real-world scenarios, such as zero-day attacks, adversarial examples, and data drift, to assess its performance under these complex conditions. To fine-tune the model effectively, in-depth research into the sensitivity of IGAN-IDS to various hyperparameters is necessary. Understanding how different settings impact the model's performance can lead to more optimized configurations. A more detailed exploration of the computational efficiency and runtime performance of IGAN-IDS is crucial, especially for its applicability in real-time intrusion detection systems. This analysis will help determine its practical feasibility in operational environments.

*3.3.3.3. Generative adversarial network with random forest (GAN RF).* Lee and Park (2021) proposed a model that combines GAN with Random Forest classification to address the issue of imbalanced data in intrusion detection systems, which can lead to data loss or overfitting. It operates on the CICIDS 2017 dataset, containing various attack types and predominantly consisting of normal network traffic data, with minority classes such as Infiltration, Web attacks, and Heartbleed attacks. The GAN RF model aims to generate additional instances for the rare classes by learning the underlying data distribution and creating synthetic data that resembles the real data. This process addresses the issue of data imbalance by enhancing the representation of underrepresented classes, ensuring a more balanced dataset for classification purposes. The resampled data obtained through this method is utilized in conjunction with a Random Forest classification algorithm for intrusion detection. The goal is to provide a more robust and balanced dataset to the classification model, enhancing its ability to accurately detect and classify various types of network intrusions, especially those belonging to the rare classes present in the dataset. The experiments aimed to assess the model's performance before and after data resampling using GAN and to compare it with other algorithms. A Random Forest (RF) model was used for classification, with and without prior data resampling through GAN. The results demonstrated notable improvements in the GAN-resampled RF model, particularly regarding Recall and F1-Score metrics, essential in unbalanced data scenarios. Specifically, minority classes like Bot, Infiltration, and Heartbleed, which were oversampled, exhibited enhanced performance, indicating that better understanding of the minority class characteristics improved the classification performance of both the minority and normal classes. The experiment highlighted how Precision and Recall metrics can conflict depending on data volume, where enhancing Precision often reduces Recall and vice versa.

*Weaknesses.* However, several pertinent considerations arise from the research findings, and these aspects should guide future work. Firstly, the study employs GANs for data resampling, a promising approach for mitigating the challenges associated with imbalanced datasets. Nevertheless, it is crucial to acknowledge the potential limitations and challenges linked to using GAN. GANs can be susceptible to training instability, mode collapse, and sensitivity to hyperparameter settings. The research does not delve into these challenges, and future work should explore these aspects to provide a more comprehensive view of the approach. Secondly, the research selects Random Forest (RF) as the classification algorithm. RF is well-known for its robustness and interpretability, but various factors can influence its performance, including the quality of features and the choice of hyperparameters. A robust justification for the selection of RF should be provided, as well as the limitations associated with this algorithm. Furthermore, the study briefly touches on the enhanced performance of minority classes such as Bot, Infiltration, and Heartbleed due to GAN-based resampling. However, the research does not delve into the specifics of these class-specific improvements. Lastly, the study hints at the inadequacy of Synthetic Minor Oversampling Technique (SMOTE) in fully addressing data imbalance due to data overlap among classes, with GANs being considered a more effective alternative. However, a more comprehensive discussion and in-depth analysis of how GAN precisely addresses the issue of data overlap is necessary to gain a deeper understanding of the method's effectiveness.

*3.3.3.4. Inference generative adversarial network (IGAN).* Shah and Das (2022) presented different IGAN that detects malicious strings with decent accuracy. It consists of an encoder, a decoder, and a discriminator. The encoder and decoder form the generative unit, reconstructing the input and mapping it to a latent space variable. IGAN exploits latent space and adversarial training with the discriminator to enhance learning of the normal distribution. The model enhances its understanding of the normal data distribution by using the latent space and adversarial training with the discriminator. The discriminator plays a dual role in IGAN. It acts as both a classifier and a feature extractor. This means it helps distinguish between normal and anomalous instances while extracting meaningful features from the data. IGAN calculates an anomaly score after passing the data through the trained model to identify whether an instance is an anomaly or normal. The anomaly score measures the different data from the learned normal distribution. It employs one-hot encoding for continuous features, transforming them into a binary form for classification. The model utilized the discriminator as a feature extractor improved the performance of all the baseline and proposed architectures, resulting in an increase in the AUC score of 1.15942 % for the IGAN model.

*Weaknesses.* However, it is worth noting that both the NSL-KDD dataset and its precursor, the KDD Cup 1999 dataset, are modified datasets, and their ability to represent the complexity and diversity of real-world network traffic fully may be limited. This raises questions about the generalizability of the research results to real-world scenarios with more complex and varied data. Secondly, the data preprocessing steps described in the research, such as converting continuous features to one-hot encoding and transforming attack types into binary format, may oversimplify the data. Moreover, the research highlights a data imbalance issue, with significantly more anomalous entries than normal entries in the dataset. While this is a common characteristic of intrusion detection datasets, it can introduce challenges in model training and evaluation, potentially leading to biased model performance. Furthermore, the training of IGAN is conducted for a fixed number of epochs (100) with a specified learning rate. Deep learning models often require careful hyperparameter tuning, and the choice of hyperparameters can substantially impact model performance.

*3.3.3.5. Generative adversarial network for super-resolution (GAN-SR).* Instead of using random forest method in their model like GAN RF, Wang et al. (2023) address the low minority class identification rate caused by data imbalance in anomaly detection tasks using a GAN-Super Resolution intrusion detection model. The model corrects the imbalance of minority classes in the dataset using a GAN and completes high-dimensional feature extraction using a stacked asymmetric depth self-encoder. The high-dimensional feature extraction is performed to address the issues of low reconstruction error and lengthy training times. Then a random forest decision tree is built, and intrusion detection is carried out using the features retrieved by the stacked asymmetric depth self-encoder. The GAN module employs adversarial training between the generator G and discriminator D to generate new samples aligned with the original data distribution. Through the process of passing random noise into the generator, it generates samples that the discriminator is tasked with differentiating from genuine data. The goal of the generator is to create samples resembling the original distribution, while the discriminator aims to accurately distinguish between real and generated samples. The SR module utilizes SNDAE for feature extraction from complex and high-dimensional data, enhancing hierarchical learning and providing significant data representations. SNDAE's capacity for unsupervised hierarchical feature learning reduces training time, preserving efficiency without sacrificing accuracy. The GAN-SR model demonstrated enhanced performance in expanding minority sample sizes and detecting anomalies compared to other models. It significantly improved detection accuracy, precision, recall, and F1-values across various minority classes and datasets. The model effectively reduced data imbalance, achieving better anomaly detection results, especially at an 80% expansion ratio, showcasing its robustness in handling imbalanced data.

*Weaknesses.* However, using a GAN to correct the imbalance of minority classes in the dataset may introduce additional complexity and potential challenges in training the model effectively. The study sheds light on several pertinent considerations that warrant attention in future research. The model's performance demonstrates high dependence on the specific dataset and class distribution. It excels when a certain ratio expands the minority class. This observation underscores the need for parameter tuning to align with the dataset's characteristics, implying that the model's performance may fluctuate significantly across different datasets. Moreover, the research primarily focuses on conventional evaluation metrics like accuracy, precision, recall, and F1-score, which are well-suited for binary classification tasks but may fall short of capturing the multi-class problems. Future work should explore the complexities of setting appropriate detection thresholds, especially in scenarios involving multiple minority classes, to enhance the model's performance.

Future work should improve the model's capacity to generalize across a broader range of datasets. This can involve developing adaptive techniques that automatically adjust model parameters based on the dataset characteristics, making it more flexible and reducing the need for manual parameter tuning (Fu et al., 2022). Researchers should explore developing robust threshold selection methods for multi-class anomaly detection that align with the model's performance goals. Given the potential challenges introduced by GANs in class imbalance adjustment, future research should optimize the model's training process for efficiency. This includes exploring techniques for reducing training time and computational resource requirements ensuring the model remains practical for real-world deployment in resource-constrained environments.

*3.3.3.6. Network intrusion detection system (NIDS).* Park et al. (2023) proposed AI-based NIDS that addresses data imbalance in network intrusion detection systems. The model focuses on the reconstruction error and Wasserstein distance-based GAN to generate plausible synthetic data. It leverages a generative adversarial network (GAN), specifically the Boundary Equilibrium Generative Adversarial Networks (BEGAN) model, to generate synthetic data that mirrors the underrepresented classes in the dataset. These artificially created instances are merged with the original data, creating a more balanced training dataset, particularly the less frequent attack types. Additionally, an autoencoder aids in extracting pertinent features while reducing dimensionality. This model helps in capturing essential characteristics of the input data, assisting in the creation of representations that are more robust and discriminative, thus aiding in addressing the data imbalance issue. Various classifiers, including Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), Support Vector Machines (SVM), and Decision Trees (DT), are trained on this augmented dataset. These models are employed to classify network traffic data and identify intrusions or attacks. Experimental results including real-world datasets indicate that the model outperforms conventional machine learning and deep learning approaches in addressing data imbalance. However, the accuracy values varied from approximately 80% to around 93% in different experiments, depending on the dataset and the specific model configurations used.

*Weaknesses.* Nevertheless, the diversity of data sources in the experiments, spanning synthetic and real data, might not fully encapsulate the intricacies and diversity of network intrusion scenarios encountered in real-world settings. Network attacks and traffic patterns can exhibit much greater variation, potentially resulting in varying system performance. Secondly, data imbalance, highlighted in several experiments, raises concerns, particularly in cases where certain attack classes are underrepresented. This imbalance can introduce bias in model performance, especially when dealing with rare attack types. While the experiments attempt to mitigate this by generating synthetic data, potential limitations in effectively representing these infrequent cases should be further explored. Furthermore, using synthetic data in certain experiments may prompt inquiries regarding the quality and fidelity of the generated data. The performance improvements observed based on synthetic data may not necessarily translate to the complexities and nuances of real-world data. It is essential to scrutinize the generalizability of the proposed models to unseen data to avoid overfitting the specific datasets used in the experiments. Lastly, while the primary performance metric used in the evaluation is accuracy, it is vital to consider additional metrics like precision, recall, and the F1-score. These metrics provide a more comprehensive understanding of system performance, especially in imbalanced datasets (Table 6).

#### 3.3.4. Issue 4: absence of prior information on attack categories

*3.3.4.1. Novel network intrusion detection technique based on generative adversarial networks (N-GAN).* Reconstruction-based anomaly detection, N-GAN (Iliyasu and Deng, 2022) was used to address the issue of absence of prior information on attack categories. This model is trained in two distinct phases, GAN training and encoder training. During the GAN training phase, the model is trained on large-scale normal network traffic data using a regular GAN setup. The generator (G) creates fake network traffic features from random noise sampled from latent space, and the discriminator (D) attempts to distinguish between real and generated network traffic features. Through adversarial training, the GAN aims to generate network traffic features that closely match the distribution of real traffic, allowing the generator to learn the structural variability of normal network traffic. In the subsequent Encoder training phase, the trained GAN is used to train an encoder (E), where the GAN generator acts as the decoder. The encoder is trained using a contrastive loss function, considering both normal and a few malicious samples. The loss function constrains the model to well-reconstruct benign samples while forcing malicious samples to have poor reconstructions. This contrastive training allows the encoder to learn to map network traffic features into latent features while distinguishing between normal and potentially malicious patterns. The anomaly detection process involves passing network traffic features through the encoder to transform them into latent features. These latent features are then reconstructed back to the network traffic feature space by the generator acting as a decoder. The model calculates the degree of deviation between the original and reconstructed features, considering minimal deviations for normal traffic and larger deviations for anomalous traffic, especially those lying outside the learned manifold of normal network traffic.

*Weaknesses.* N-GAN, designed to address the absence of prior information on attack categories, exhibits strengths but also notable weaknesses. This model heavily relies on diverse and representative training data, posing a challenge if it lacks variations or specific attack patterns. While its contrastive training approach distinguishes normal from malicious patterns, N-GAN may struggle with novel or rare attack types outside learned patterns, leading to false positives or negatives. Additionally, its limited ability to adapt to new attack categories, interpret latent representations, handle imbalanced data, and scale for real-time network traffic poses challenges in accurately detecting emerging anomalies.

Expanding the model's adaptability to novel attack types by devising mechanisms to efficiently incorporate and learn from evolving threat landscapes is crucial. Improving the model's capability to handle imbalanced data and rare attack instances would strengthen its performance in detecting anomalies accurately. Further exploration into methods that enable real-time scalability for network traffic analysis, such as optimizing computational efficiency or streamlining latent feature representation, would be valuable. Additionally, research into interpretability methods for latent representations and developing techniques to mitigate false positives and negatives could refine the model's anomaly detection capabilities (Table 7).

#### 3.3.5. Issue 5: increased data complexity and lack of real-time detection

*3.3.5.1. Adversarially learned anomaly detection (ALAD).* Zenati et al. (2018b) presents a unique approach rooted in bi-directional GANs' principles for anomaly detection in complex and high-dimensional data. ALAD's core methodology involves deriving adversarially learned features and using reconstruction errors based on these features to discern anomalies. Additionally, ALAD effectively incorporates recent advancements toward ensuring both dataspace and latent-space cycle consistency, thereby promoting the stability of GAN training. ALAD utilizes GANs to model normal data distribution. GANs are adept at capturing complex data distributions, including high-dimensional data. ALAD attempts to understand the pattern of regular data to create authentic samples that correspond to the complexities present in the data it was trained upon. This capability helps in handling increased data complexity without requiring explicit manual feature engineering. However, once trained, the inference or anomaly detection phase, where it evaluates whether a sample is normal or anomalous, can be relatively fast compared to training. The ability to generate reconstructions and compute anomaly scores based on these reconstructions allows for near real-time detection once the model is deployed. However, the actual real-time performance might vary depending on the size of the model, hardware capabilities, and the complexity of the data. ALAD demonstrates competitive results against state-of-the-art methods in anomaly detection with a high accuracy of 94 %.

*Weaknesses.* Despite its efficiency in inference time, it does not delve into the specifics of this aspect, such as the hardware or software used for testing. Providing more details about the computational resources can enhance the credibility of the performance claims. ALAD's training phase can be computationally intensive. Yet, during inference, the model demonstrates relatively faster anomaly detection capabilities. The reconstruction-based approach for anomaly scoring allows for potential real-time detection upon deployment, but the actual real-time performance may fluctuate based on model size, hardware limitations, and data complexity. Additionally, the model primarily restricts its comparison to AnoGAN and a variational auto-encoder (VAE), which can narrow the scope of the evaluation. The anomaly detection field encompasses many techniques, and the study's breadth of comparison can be enriched by incorporating a more extensive array of baseline methods. The study briefly mentions the encoder's role in eliminating the need for a costly procedure to recover the latent representation. However, it does not explore the encoder's accuracy or impact on overall anomaly detection performance.

Future work should provide a more detailed exposition of experimental setups and network architectures, enhancing transparency and facilitating reproducibility. Furthermore, delving into the specifics of

**Table 8**
Evaluation metrics comparison of GAN models addressing issue 5.

| GAN model | Datasets characteristics | Performance |
|---|---|---|
| ALAD | KDDCup99<br>• Size: 494,021 records<br>• Type: Network Traffic<br>• Description: Simulated network traffic dataset for intrusion detection | Precision: 94 %<br>Recall: 96 %<br>F1-Score: 0.950 |
| RCALAD | KDDCup99<br>• Size: 5,000,000 records<br>• Type: Network Traffic<br>• Description: Simulated network traffic dataset for intrusion | Precision: 95 %<br>Recall: 96 %<br>F1-Score: 0.954 |

efficiency regarding inference time by elucidating the hardware and software used in testing can verify performance claims. Expanding the scope of comparative evaluations by including a broader array of baseline anomaly detection techniques can yield a more comprehensive understanding of ALAD's relative merits. Lastly, a deeper investigation into the accuracy and functional impact of the encoder within the ALAD framework is essential for comprehensive anomaly detection assessment and should be a focal point of future research.

*3.3.5.2. Regularized complete cycle consistent GAN for anomaly detection (RCALAD).* Rooted with similar adversarial learning principles, Dehghanian et al. (2023) introduced RCALAD, an adversarial method that leverages GANs through cycle consistency in reconstruction error. The proposed method introduces a novel discriminator to the structure and employs a supplementary distribution in the input space to steer reconstructions toward the normal data distribution. This model comprises an encoder and a generator trained in the structure of an adversarial neural network. The encoder maps input data to a latent space, while the generator performs the reverse mapping. The model employs various discriminators (Dxz, Dxx, Dzz, Dxxzz) to ensure cycle consistency between input and latent spaces. RCALAD utilizes a supplementary distribution in the input space to guide the reconstruction process toward the normal data distribution. This helps distinguish anomalous samples from their reconstructions and enables more accurate anomaly detection. The model achieves 95.3 % in terms of performance accuracy.

*Weaknesses.* While RCALAD may effectively handle increased data complexity during the anomaly detection process, its training phase could be computationally intensive. Training an adversarial neural network (ANN) structure, particularly when employing multiple discriminators, might demand significant computational resources and time. The use of multiple discriminators and the complex adversarial network structure could lead to a more complex model. Such complexity might hinder interpretability, making it challenging to understand and interpret the model's decision-making process or provide transparent explanations for detected anomalies. Furthermore, the study indicates that the model's performance on the thyroid dataset falls short of the Isolation Forest (IF) model. This suggests that for datasets with an abundance of features, alternative approaches such as feature selection or dimensionality reduction may outperform deep learning models (Zebari et al., 2020). Additionally, a slightly increased training time for the proposed model was recorded compared to the ALAD model. While this additional training time correlates with performance improvements, it is crucial to consider the computational resources and infrastructure required for training. In resource-constrained environments, longer training times may not be a viable option.

Future research can simplify the model architecture without sacrificing performance. This might involve pruning redundant components, reducing the number of discriminators, or exploring alternative network structures that maintain effectiveness while decreasing complexity (Li et al., 2022). In cases where the model's performance lags alternative approaches, such as the Isolation Forest (IF) model, further work can concentrate on optimizing the RCALAD model for datasets with numerous features. Feature selection methods and dimensionality reduction techniques should be explored as viable solutions to augment the model's performance. Moreover, the study's emphasis on the F1 score as a performance metric calls for a more detailed approach to evaluation. Future research should investigate the suitability of different evaluation metrics for diverse datasets (Kim et al., 2021). Tailoring the choice of metrics to align with specific anomaly detection task objectives is crucial, allowing for a more comprehensive assessment of the model's capabilities. While performance improvements justify the model's increased training time, it is essential to consider the practical implications of longer training times, particularly in resource-constrained environments. Future work should seek to optimize training processes to strike a balance between performance gains and computational efficiency. Considering the hardware dependency observed in the study, researchers should evaluate the generalizability of the RCALAD model to various hardware configurations (Table 8).

### *3.3.6. Issue 6: data generation for anomalous patterns*

*3.3.6.1. End-to-end deep architecture GAN.* An end-to-end deep architecture for Intrusion Detection Systems (IDS), using GANs to train deep models in a semi-unsupervised manner, was proposed by Mohammadi and Sabokrou (2019). The proposed model for Intrusion Detection Systems (IDS) tackles the challenge of generating data for anomalous patterns by employing a sophisticated two-module architecture: the Reconstructor Network (R) and the Anomaly Detector Network (A). This approach is designed to autonomously generate synthetic anomalous traffic and distinguish it from genuine normal traffic without relying on explicitly labeled anomaly data during the training phase. The Reconstructor Network (R) serves as the core component responsible for simulating anomalous network patterns. By reconstructing normal traffic patterns adversarially, R aims to generate synthetic anomalies. R utilizes an encoder–decoder architecture to map incoming normal traffic into a latent space and then reconstruct it. However, crucially, R is trained to stop perfectly reconstructing normal traffic. This deliberate imperfection ensures that the generated anomalies deviate from the learned distribution of normal traffic. The emphasis is on introducing deviations or irregularities while retaining certain characteristics of normal traffic patterns. Complementing R is the Anomaly Detector Network (A), functioning as a classifier to discern between the generated synthetic anomalies and real normal traffic. Leveraging knowledge obtained from training on genuine data, A is equipped to identify and reject simulated anomalous traffic. Its primary role is to detect deviations from the learned distribution of normal traffic, allowing it to effectively differentiate between simulated anomalies and authentic normal traffic. The joint adversarial training of R and A follows a model like GAN but with a different objective. While conventional GANs aim to generate samples from a specific distribution, R in this context strives to create anomalies that distinctly differ from the learned distribution of normal traffic. A, on the other hand, learns to discriminate between the synthetic anomalies generated by R and genuine normal traffic. The model achieves 91.39 % in terms of performance accuracy.

*Weaknesses.* The semi-supervised employs normal records from the dataset during training, neglecting the anomalous records. This omission can impede the model's effectiveness in detecting anomalies do not present in the training data. The model's threshold for distinguishing between normal and anomalous traffic is established using the validation set, which lacks abnormal samples. This can potentially impact the model's ability to classify anomalies accurately. The model also introduces the concept of generating simulated abnormal flows with deviations from the target class rather than relying on actual anomalies. While this approach offers some degree of generalization, it does not fully capture the complexities of real-world network anomalies.

Future work in this field should consider enhancing the model's performance by incorporating more diverse evaluation metrics to ensure a comprehensive assessment of its capabilities in detecting network anomalies. Additionally, researchers should explore more sophisticated methods for generating simulated abnormal flows that better emulate real-world anomalies. Further investigations into the model's scalability and adaptability in dynamic, real-time network environments are essential to validate its practical utility. Finally, efforts should be made to collect and utilize more extensive datasets of actual network anomalies to improve the model's ability to detect novel and evolving threats, thus advancing state-of-the-art intrusion detection systems.

*3.3.6.2. Network generative adversarial network (Net-GAN).* González et al. (2020) introduced Net-GAN, an approach to network anomaly detection in multivariate time-series data using recurrent neural networks (RNNs) and generative adversarial networks (GAN). It focuses on detecting anomalies in multivariate time-series data, leveraging temporal dependencies through RNNs. It does not make any assumptions about the nature of the data, making it suitable for complex network monitoring data. Net-GAN uses GANs and RNNs to capture temporal correlations in multivariate time-series data. Within the GAN framework, Net-GAN features a generator (G) that learns to create synthetic sample sequences by inputting Gaussian noise from a latent space, aiming to produce sequences resembling normal data patterns. Simultaneously, a discriminator (D) is trained to distinguish between real and synthetic sequences. Through adversarial training, where the generator attempts to deceive the discriminator by creating sequences deviating from the normal data distribution, Net-GAN encourages the generation of data patterns that differ from the baseline. In the subsequent anomaly detection phase, the trained discriminator acts as an anomaly detector, leveraging its learned ability to recognize deviations from the baseline. Notably, it offers a model-agnostic and data-driven solution for capturing the underlying data distribution, making it applicable to complex network monitoring data. Additionally, the trained generator assists in anomaly detection by performing an inverse search in the latent space, aiding in computing a residual loss. Net-GAN demonstrates its efficacy in anomaly detection by achieving notable detection rates with a low False Positive Rate (FPR) in various attack scenarios. For instance, it successfully detects approximately 93 % of botnet-related attacks, all infiltration incidents (100 %), around 89 % of port scan activities, and 78 % of DDoS attacks, all with an FPR below 1 %.

*Weaknesses.* While the model aims to create synthetic sequences deviating from normal patterns, it might face limitations in generating a wide range of diverse anomalies. Certain types of anomalies that significantly differ from the learned normal data distribution might be challenging to replicate accurately. Moreover, performance evaluation relies on publicly available datasets, including synthetically generated ones. The applicability of these datasets to real-world scenarios is unclear, necessitating further examination of the method's effectiveness in handling diverse and evolving network traffic. The imbalanced datasets, notably the SYN—NET dataset consisting of 83 % normal and 17 % malicious, raise the issue of model performance under such conditions. Lastly, while Net-GAN is noted for its low false positive rate in detecting attacks, it is important to consider a broader spectrum of performance metrics, including precision, recall, and F1-score. These additional metrics offer a more comprehensive perspective on the model's anomaly detection capabilities, ensuring a well-rounded assessment.

Future work should enhance dataset diversity by integrating a wider array of anomalies, encompassing rare or outlier patterns that deviate from the learned normal distribution. Enriching the dataset with varied and extensive anomalies can facilitate a more comprehensive learning experience for the model, enabling it to detect a broader range of anomalies. Additionally, future work should prioritize using real-world network traffic data to validate the practical utility of Net-GAN. Extensive testing in diverse and evolving network environments will help assess the model's ability to handle the complexity and dynamics of actual network traffic (Tables 9 and 10).

### 3.4. Practical implementation of GAN in network anomaly detection

One of the notable instances illustrating the practical application of GAN in network scenarios is demonstrated by Labonne and Olivereau (2020) in the domain of unsupervised protocol-based intrusion detection within real-world networks. The utilization of GAN in network intrusion detection was motivated by the evolving complexity of cyber threats, necessitating the development of more robust systems capable of identifying zero-day attacks within authentic network environments. However, to create an unsupervised anomaly-based intrusion detection solution, specifically focusing on analyzing protocol headers, posed a unique challenge. The intrusion detection classification problem demands advanced metrics to effectively detect and differentiate anomalies in network traffic. Addressing this challenge, experts leveraged the CICIDS2017 dataset, characterized by realistic network representations and credible attacks. Each protocol underwent conversion into normalized numeric features and was subjected to diverse neural network architectures, including deep autoencoders, deep MLPs, LSTMs, BiLSTMs, and GANs. The output from these algorithms produced anomaly scores, subsequently normalized and amalgamated with anomaly scores from other protocols. Encouragingly, the approach showcased promising results, successfully identifying 7 out of 11 previously unseen attacks during the training phase, notably without any false positives. Furthermore, post-processing refinements enhanced packet anomaly scores to consolidate anomalies into continuous attack sequences. These outcomes underscore the potential feasibility of deploying IDS supported by GAN in real-world network scenarios. The adoption of unsupervised learning techniques can effectively contribute to intrusion detection within real network infrastructures.

While explicit examples directly correlating to network anomaly detection were limited, the examination has been expanded to include insights from industry leaders leveraging Generative AI in diverse anomaly-related contexts. Companies such as GemmoAI, ServiceNow and Dataiku offer noteworthy instances of Generative AI utilization, although their applications might not specifically target network anomaly detection. Their integration of Generative AI technology into anomaly-related domains demonstrates a rising trend of interest, demand, and product offerings within the broader spectrum of anomaly detection solutions. Because of this, emphasizing the broader applications of Generative AI highlights the importance of conducting a comprehensive survey on GANs. Understanding their applications becomes critical in realizing their full potential within network anomaly detection.

It is worth highlighting that the widespread use of GAN for network anomaly detection has been extensively documented within academic research, often relying on publicly available datasets as foundational testing grounds. However, the transition from controlled experimental setups to the practical implementation of GANs in live network environments remains notably scarce in existing literature. While leveraging public datasets serves as a crucial benchmark for model development, the actual deployment and practical utilization of GANs within live network environments remains noticeably limited. The transition from controlled experimental settings to the implementation of GANs in real-world network infrastructures poses notable challenges and

**Table 9**
Evaluation metrics comparison of GAN models addressing issue 6.

| GAN model | Datasets characteristics | Performance |
|---|---|---|
| End-to-end deep architecture GAN | NSL-KDD<br>• Size: Large volume<br>• Type: Network Traffic<br>• Description: Network intrusion detection dataset refined from KDD Cup '99 | Accuracy: 91 % |
| Net-GAN | SYN—NET<br>• Size: Large volume<br>• Type: Network Traffic<br>• Description: Synthetically generated dataset for network intrusion detection | Accuracy: 93% |

**Table 10**
Characteristics of GAN-based model in network anomaly detection.

| Focus | Author | Model | Characteristics | Dataset | Strengths | Weaknesses |
|---|---|---|---|---|---|---|
| I1 | Yuan et al. (2020) | ACGAN | • IDS deployed on edge nodes for local data processing, reducing privacy risks during transmission.<br>• Converts network traffic into images for CNN-based classification.<br>• Utilizes AC-GAN to generate synthetic samples. | UNSW-NB15 | • Expanded dataset<br>• High precision for binary classification | • Imbalanced dataset<br>• Model complexity<br>• Real-time deployment challenges |
| I1 I4 | Iliyasu and Deng (2022) | N-GAN | • Adding a few malicious samples in training enhances robustness against false alarm rates (FARs) and improves learned representations.<br>• N-GAN excels in general anomaly detection, class-specific attack detection, and mutant attack detection.<br>• Leveraging attack samples during training, N-GAN discovers effective representations for detecting attacks. | CIC-IDS2017 | • Robustness<br>• Suitability for zero-day attacks | • Limited data diversity<br>• Small sample size<br>• Attack variants<br>• Lack of real-world evaluation<br>• Overfitting risk |
| I2 | Kaplan and Alptekin (2020) | BiGAN | • Simultaneous encoder learning eliminates the need for a costly latent representation recovery procedure.<br>• Effectiveness relies on A(x) score function, combining reconstruction and discriminator-based loss for anomaly measurement. | • MNIST<br>• KDD99 | • Encoder learning<br>• Anomaly measurement | • Limited use of anomalous records<br>• Threshold determination<br>• Simulated abnormal flow<br>• Performance evaluation<br>• Real-time capability |
| I2 | Araujo-Filho et al. (2021) | FID-GAN | • Utilizes a reconstruction loss based on latent space mapping for higher detection rates.<br>• Addresses latency with an encoder, achieving at least 5.5x faster results. | • SWaT<br>• WADI<br>• NSL-KDD | • Low latency<br>• Faster detection<br>• High detection rates | • Limited diversity of datasets<br>• Comparison with limited baselines<br>• Dependency on additional components<br>• Influence of data types<br>• Limited insight into training challenges |
| I2 | Fu et al. (2023) | GANAD | • GANAD focuses on anomaly identification, not data synthesis.<br>• Utilizes an auto-encoder architecture to expedite generator loss computation.<br>• Enhances training stability with Wasserstein distance and gradient penalty in discriminator training.<br>• Adopts a novel training strategy for learning the minority abnormal distribution from normal data. | • KDDCUP99<br>• NSL-KDD<br>• UNSW-NB15 | • Consistent improvements<br>• Efficiency in computation<br>• Multi-classification capability | • Dataset dependency<br>• Imbalanced datasets<br>• Latency constraints<br>• Limited comparison<br>• Generalization<br>• Scalability<br>• Network intrusion categories |
| I3 | Wang et al. (2019) | FlowGAN | • Utilizes GAN's data augmentation for classes with limited samples.<br>• Experimental results demonstrate FlowGAN's superiority over oversampling methods in both unbalanced and balanced datasets. | ISCX | • Improved performance<br>• Oversampling | • Limited applicability<br>• Dataset dependency<br>• Traffic characteristics<br>• Impact on false positives |
| I3 | Lee and Park (2021) | GAN RF | • Utilizes GAN for generating virtual data to address data imbalance.<br>• Employs Wasserstein GAN (WGAN) with gradient penalty for training stability.<br>• Utilizes Random Forest classifier to evaluate detection performance post data imbalance correction. | CICIDS 2017 | • Handling imbalanced datasets | • Data resampling<br>• Algorithm selection<br>• Class-specific performance<br>• Data overlap |
| I3 | Wang et al. (2023) | GAN-SR | • GAN-SR effectively corrects data imbalance in anomaly detection using GAN for generating new minority class training samples.<br>• Utilizes a stacked asymmetric depth self-encoder for high-dimensional feature extraction, improving reconstruction error and training efficiency. | • UNSW-NB15<br>• SWaT<br>• Gas Pipeline | • Improved detection performance<br>• Stable | • Dataset specificity<br>• Threshold selection |
| I3 | Shah and Das (2022) | IGAN | • IGAN uses an encoder and decoder for input reconstruction and latent space mapping.<br>• IGAN discriminator functions as both a classifier and feature extractor. | NSL-KDD | • Enhanced normal distribution learning | • Dataset selection<br>• Data preprocessing<br>• Data imbalance<br>• Model training<br>• Performance metrics<br>• Real-world applicability |

(continued on next page)

**Table 10** (*continued*)

| Focus | Author | Model | Characteristics | Dataset | Strengths | Weaknesses |
|---|---|---|---|---|---|---|
| I3 | Huang and Lei (2020) | IGAN-IDS | • Employs an anomaly-based GAN architecture with encoder, decoder, and discriminator to detect malicious strings accurately.<br>• Encoder and decoder work together to reconstruct input and improve normal distribution learning in the latent space.<br>• The discriminator acts as both a classifier and a feature extractor, aiding precise identification of anomalies and normal instances. | • NLS-KDD<br>• UNSW-NB15<br>• CICIDS2017 | • Generating new representative instances for minority classes<br>• Improved detection precision | • Limited dataset diversity<br>• Sensitivity to hyperparameters<br>• Lack of discussion on real world complexity<br>• Imbalance ratio range selection |
| I3 | Park et al. (2023) | NIDS | • Utilizes GAN and autoencoder-driven deep learning to generate synthetic data for minor attack traffic.<br>• Efficiently detects network threats in distributed environments. | • NSL-KDD<br>• UNSW-NB15<br>• IoT-23 | • Improved classification for minor classes | • Data diversity<br>• Data imbalance<br>• Synthetic data quality<br>• Performance metrics<br>• Generalization |
| I5 | H. Zenati et al. (2018) | ALAD | • ALAD employs bi-directional GANs for effective real-world anomaly detection.<br>• It uses reconstruction errors from adversarial features to enhance detection performance.<br>• ALAD maintains dataspace and latent-space cycle-consistencies, stabilizing GAN training for improved performance. | KDDCUP99 | • Effective anomaly detection<br>• Adversarial features | • Limited anomaly detection models<br>• Accuracy of encoder and performance |
| I5 | Dehghanian et al. (2023) | RCALAD | • Uses GANs and cycle consistency for anomaly detection, capturing complex patterns and generating realistic reconstructions.<br>• Introduces an efficient discriminator to the structure.<br>• Utilizes a supplementary distribution in the input space to guide reconstructions toward normal data distribution.<br>• Introduces two novel anomaly scores to enhance model performance. | • KDDCup99 | • Complex patterns<br>• Realistic reconstructions | • Dataset specificity<br>• Feature selection<br>• Performance metrics<br>• Training time<br>• Hardware dependency |
| I6 | Mohammadi and Sabokrou (2019) | End-to-end deep architecture GAN | • Compensates for lack of anomalous traffic by approximating them from normal flows.<br>• Mitigates bias toward available intrusions in the training set. | NSL-KDD | • Handling anomalous records | • Limited use of anomalous records<br>• Threshold determination<br>• Simulated abnormal flow<br>• Performance evaluation<br>• Real-time capability |
| I6 | González et al. (2020) | Net-GAN | • Makes no assumptions about data nature, enabling robust anomaly detection.<br>• Utilizes generative models for data-driven, model-agnostic learning of data distributions. | SYN—NET | • Detecting complex anomalies | • Limited real-world data<br>• Imbalanced datasets<br>• Performance metrics |

complexities yet to be comprehensively addressed.

The primary challenge in the practical application of GANs lies in obtaining authentic and diverse network data that encapsulates the complexities and variations inherent in operational networks without compromising confidentiality or integrity (González et al., 2020). Adapting GAN architectures to dynamically changing network topologies, addressing ethical considerations surrounding data use, and validating models in real-time operational scenarios are pivotal factors that significantly impact the integration of GANs into real network security frameworks.

Despite the predominance of studies using public datasets, there exists a critical need to bridge the gap between research advancements and real-world deployment of GANs in network anomaly detection. The limited literature focusing on practical implementation underscores the necessity for comprehensive case studies or field experiments that address the challenges of deploying GANs within operational networks. Future research should pivot towards these real-world applications, striving to develop robust methodologies and frameworks that can effectively integrate GANs into authentic network security systems while addressing concerns of scalability, adaptability, and ethical implications.

## 4. Future directions and recommendations

The utilization of GAN for anomaly detection in networks is a developing area with numerous prospects for further research. However, most of these prospects are within the domain of GANs. In this section, we outline potential pathways for future investigations involving the application of GANs in network anomaly detection.

### *4.1. Real-time training and detection*

Intrusion detection frameworks must operate online to provide network administrators with timely alerts for mitigation and remediation purposes. However, many implemented systems perform offline classification of network telemetry (Dromard and Owezarski, 2020; Zaman and Lung, 2018). In these cases, timely remediation is impossible since detection occurs much later than actual malicious behavior. These framework types often attempt to marginally increase detection rates on outdated data sets, which is a problem facing network anomaly detection systems (Mills et al., 2022). This is due to the requirement for vast training data and the sheer amount of data within modern networks, which needs to be assessed. The existing GAN models for network anomaly detection are trained offline and then used to detect anomalies in real-time network traffic. The detection process is conducted in real-time as network traffic is being processed, but the model is not updated in real time. Instead, it undergoes periodic retraining on new data to maintain its effectiveness.

One key optimization strategy in deploying GAN for real-time network anomaly detection is mini-batch training (Azahad and Hameeda, 2023). Training GANs involves iteratively updating the model's parameters using a dataset. Traditional approaches feed the entire dataset into the GAN for each training iteration. While this can yield accurate results, it is computationally expensive and time-consuming, often impractical for real-time analysis. Mini-batch training, on the other hand, breaks the dataset into smaller, manageable portions or batches, typically containing a few examples (Kong et al., 2023). Each training iteration then updates the model using one of these mini-batches. The primary advantage of mini-batch training is its ability to reduce computation time during training. When dealing with extensive datasets, such as network traffic logs, processing the entire dataset in one go can be prohibitively slow and resource-intensive (Sinha et al., 2020). Real-time analysis in network security involves rapidly detecting anomalies or suspicious activities as they occur or shortly after they are identified. Mini-batch training enables GANs to be more responsive and adaptive. The reduced computation time means the model can update its understanding of normal network behavior more frequently, making it well-suited for real-time or near-real-time analysis of incoming network traffic.

### 4.2. Resource efficiency enhancement

Generative adversarial networks (GANs) have shown great potential in network anomaly detection. Novel GAN-based methods such as GANAD, RCALAD, and GAN-SR have been proposed to detect network anomalies with high accuracy and efficiency. However, these methods can be computationally expensive and require significant resources to train and deploy. Therefore, forthcoming research should focus on exploring GAN architectures that are resource-efficient and lightweight (Labonne and Olivereau, 2020). Addressing the need for resource-efficient and lightweight Generative Adversarial Network (GAN) architectures is critical due to the considerable computational demands of traditional GANs, which pose challenges for practical deployment in resource-constrained settings (Shuvo et al., 2023). Existing GAN models like ALAD involve resource-intensive operations during both training and inference phases. These resource requirements primarily stem from large model sizes, complex network structures, and high computational demands, making them unsuitable for real-time applications, edge computing, or devices with limited computational capabilities. For instance, the substantial memory and power consumption, as well as the lengthy training times associated with these architectures, hinder their usability in scenarios requiring efficiency and speed.

However, several potential strategies and innovations could enhance the resource efficiency of GAN architectures. Proposing a similar auto-encoder architecture like GANAD can make up for the time-consuming problem of the traditional generator loss computation (Fu et al., 2023). By using an auto-encoder architecture, the generator's loss can be computed more efficiently by comparing the reconstructed data with the original data. Additionally, utilizing a new training strategy to better learn minority abnormal distribution from normal data can contribute to the detection precision. This strategy could involve techniques such as oversampling the minority class, undersampling the majority class, or generating synthetic examples of the minority class.

Models like GANAD, have made significant progress in improving resource efficiency, particularly in the field of network anomaly detection. However, the need for further research persists due to several reasons. Firstly, the landscape of problems that GANs are applied to is continuously evolving, presenting new challenges that may not be addressed by existing models. Secondly, while models like GANAD have enhanced efficiency, there is always scope for further optimization to make models faster, reduce their memory footprint, or improve scalability. Thirdly, GANs' performance can be sensitive to various factors like the choice of hyperparameters, the architecture of the generator and discriminator, and the type of loss function used. Further research could lead to more robust models that are less sensitive to these factors. Lastly, every model has its limitations. For instance, GANs can sometimes struggle with mode collapse, where they fail to generate diverse examples (Bhagyashree et al., 2020).

### 4.3. Hyperparameter optimization

One unaddressed area of study is the impact of GAN hyperparameter selection on anomaly detection performance (Sabuhi et al., 2021). Hyperparameter optimization refers to the process of fine-tuning various parameters, configurations, and functions within the GAN model. It involves seeking an optimal balance for parameters such as learning rates, network architectures, activation functions, and loss functions. Optimizing these elements is crucial as they influence the overall performance of the anomaly detection model. Future research in this area could focus on improving the performance of GAN-based methods by exploring different architectures and training strategies (Xia et al., 2022). Furthermore, enhancing the interpretability of GAN could be achieved by devising visualization techniques for the attention maps generated by the residual channel attention module (Dehghanian et al., 2023). In addition, to reinforce the robustness of GAN, it is worth considering exploring diverse techniques and developing strategies for handling missing data (Wang et al., 2023).

### 4.4. Integration of zero-shot learning

Zero-shot learning is a valuable approach for GAN-based network anomaly detection due to its ability to recognize anomalies without prior exposure to specific anomaly types during training (Wang et al., 2021). Traditional learning methods require labeled data for each anomaly type, which may not cover all potential anomalies in real-world scenarios. By contrast, zero-shot learning enables GANs to detect and differentiate anomalies that were not part of the training dataset, allowing for increased adaptability and flexibility in anomaly detection. Implementing zero-shot learning in future work involves designing GAN architectures that are capable of learning from latent representations and generalizing effectively. Techniques such as leveraging generative models to create diverse anomaly representations and developing novel loss functions that encourage the network to learn about diverse anomalies without explicit training on them can be explored. Additionally, refining the GAN's ability to extract meaningful features and representations from data, especially in scenarios with scarce labeled anomaly data, can further enhance zero-shot learning capabilities for network anomaly detection. Future research should focus on improving the robustness and generalization ability of GAN-based models through zero-shot learning to address the challenge of detecting previously unseen anomalies in network traffic effectively.

### 4.5. Transfer learning and generalization capabilities

GAN-based anomaly detection models currently face a critical challenge regarding their adaptability and generalization across diverse network environments (Azahad and Hameeda, 2023). These models struggle to extend their learned knowledge effectively beyond the specific domain they were trained in, impeding their performance when confronted with varied datasets. To address this issue, it is essential to enhance the transfer learning capabilities of these models. Developing transfer learning capabilities allows these models to acquire knowledge from one domain and apply it across various network settings without necessitating extensive retraining. Implementing transfer learning methods would facilitate the efficient transfer of knowledge from a pre-trained source domain to a target domain, enabling the model to adapt and perform effectively in new environments. Future research should aim to refine transfer learning approaches, focusing on preventing negative knowledge transfer between domains and exploring unsupervised transfer learning methods. These efforts are vital for reducing the model's dependency on labeled data, allowing it to learn and adapt efficiently to new network environments without extensive human annotation.

Overall, GAN-based methods represent a promising approach to network anomaly detection, and further research in this area could lead to even more effective and efficient methods. Despite the promise of GAN-assisted anomaly detection, challenges remain. These include more efficient GAN architectures, robustness to adversarial attacks, and better interpretability.

## 5. Conclusion

In conclusion, the landscape of network anomaly detection has undergone significant transformations over the years, marked by a transition from traditional supervised methods to unsupervised approaches. Therefore, GAN has been proposed to predict and detect complex anomalous network traffic data. GANs are unsupervised generative models that leverage their capacity to model intricate data distributions. Specifically, GANs employ a two-component architecture composed of a generator and discriminator. The generator creates synthetic data that closely mimics normal network traffic patterns, while the discriminator distinguishes between real and generated data, constantly improving their capabilities through competitive training. GANs find practical applications in generating data that closely resembles real data, contributing to data augmentation and enhancing representation learning for network anomaly detection. GANs have emerged as a powerful tool for addressing network anomaly detection, a challenging task marked by the rarity and unpredictability of anomalies. GANs generate realistic data by pitting a generator against a discriminator, and various architectural approaches exemplify their application in anomaly detection. These approaches range from BiGAN architectures and bi-directional GANs to deep models for Intrusion Detection Systems (IDS) and GAN-based solutions for unsupervised cyber-attack detection. GANs have also been instrumental in overcoming class imbalance in anomaly detection datasets, with methods introducing novel discriminator structures, supplementary input distributions, and auxiliary classifiers. Notably, GANAD and RCALAD offer promising innovations in anomaly identification, reducing computational time and enhancing precision. However, future research should focus on developing resource-efficient GAN architectures and exploring the impact of GAN hyperparameter selection on anomaly detection performance. While GAN-based methods show great potential, further work is required to address challenges, such as the development of more interpretable GANs, the incorporation of diverse techniques, and to explore the integration of GANs with mini-batch training for adaptive real-time responses.

## CRediT authorship contribution statement

**Willone Lim:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Kelvin Sheng Chek Yong:** Conceptualization, Investigation, Supervision, Writing – review & editing. **Bee Theng Lau:** Conceptualization, Writing – review & editing. **Colin Choon Lin Tan:** Conceptualization, Investigation, Supervision, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

No data was used for the research described in the article.

## References

Aggarwal, A., Mittal, M., Battineni, G., 2021. Generative adversarial network: an overview of theory and applications. Int. J. Inform. Manag. Data Insights 1 (1). https://doi.org/10.1016/j.jjimei.2020.100004.

Al Olaimat, M., Lee, D., Kim, Y., Kim, J., Kim, J, 2020. A learning-based data augmentation for network anomaly detection. In: *Proceedings - International Conference on Computer Communications and Networks, ICCCN, 2020-Augus*. https://doi.org/10.1109/ICCCN49398.2020.9209598.

Alnawayseh, S.E.A., Al-Sit, W.T., Ghazal, T.M., 2022. Smart congestion control in 5G/6G networks using hybrid deep learning techniques. Complexity 2022. https://doi.org/10.1155/2022/1781952.

Araujo-Filho, P.F.de, Kaddoum, G., Campelo, D.R., Gondim Santos, A., Macedo, D., Zanchettin, C, 2021. Intrusion detection for cyber-physical systems using generative adversarial networks in fog environment. IEEe Internet. Things. J. 8 (8), 6247–6256. https://doi.org/10.1109/JIOT.2020.3024800.

Ariyaluran Habeeb, R.A., Nasaruddin, F., Gani, A., Hashem, I.A.T., Ahmed, E., Imran, M., 2019. Real-time big data processing for anomaly detection: A Survey. Int J Inf Manage 45, 289–307. https://doi.org/10.1016/j.ijinfomgt.2018.08.006.

An, J., Cho, S., 2015. Variational autoencoder based anomaly detection using reconstruction probability. Spec Lect IE 2, 1–18.

Arthur, D., & Date, P. (2022). *A hybrid quantum-classical neural network architecture for binary classification*. https://arxiv.org/abs/2201.01820.

Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M., Inman, D.J., 2021. A review of vibration-based damage detection in civil structures: from traditional methods to machine learning and deep learning applications. Mech. Syst. Signal. Process. 147, 107077 https://doi.org/10.1016/j.ymssp.2020.107077.

Azahad, S., Hameeda, S., 2023. A deep exposition of GAN and its applications. Int. J. Eng. Technol. Manag. Sci. 7 (2), 32–37. https://doi.org/10.46647/ijetms.2023.v07i02.005.

Kitchenham, B., 2014. Procedures for Performing Systematic Reviews. In: Procedures for Performing Systematic Reviews, 33. Keele University Technical Report, pp. 1–26.

Bashar, M.A., Nayak, R., 2020. TAnoGAN: time series anomaly detection with generative adversarial networks. In: 2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020, 3, pp. 1778–1785. https://doi.org/10.1109/SSCI47803.2020.9308512.

Bhagyashree, Kushwaha, V., Nandi, G.C, 2020. Study of prevention of mode collapse in generative adversarial network (GAN). In: *4th IEEE Conference on Information and Communication Technology, CICT 2020*. https://doi.org/10.1109/CICT51604.2020.9312049.

Bhattarai, B., Baek, S., Bodur, R., Kim, T.K., 2020. Sampling strategies for gan synthetic data. In: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, pp. 2303–2307. https://doi.org/10.1109/ICASSP40776.2020.9054677.

Blaise, A., Bouet, M., Conan, V., Secci, S., 2020. Detection of zero-day attacks: an unsupervised port-based approach. Comput. Netw. 180 https://doi.org/10.1016/j.comnet.2020.107391.

Blowers, M., Williams, J., 2014. Machine learning applied to cyber operations. Adv. Inform. Security 55. https://doi.org/10.1007/978-1-4614-7597-2. V–VIII.

Boukerche, A., Zheng, L., Alfandi, O., 2020. Outlier detection: methods, models, and classification. ACM. Comput. Surv. 53 (3) https://doi.org/10.1145/3381028.

Cai, Z., Xiong, Z., Xu, H., Wang, P., Li, W., Pan, Y., 2021. Generative adversarial networks: a survey toward private and secure applications. ACM. Comput. Surv. 54 (6) https://doi.org/10.1145/3459992.

Das, T., Shukla, R.M., Sengupta, S., 2022. What could possibly go wrong?: identification of current challenges and prospective opportunities for anomaly detection in internet of things. IEEE Netw. 37, 194–200. https://doi.org/10.1109/MNET.119.2200076.

Dehghanian, Z., Saravani, S., Amirmazlaghani, M., & Rahmati, M. (2023). *Spot the odd one out: regularized complete cycle consistent anomaly detector GAN*. https://arxiv.org/abs/2304.07769.

Deng, A., Hooi, B., 2021. Graph neural network-based anomaly detection in multivariate time series. In: 35th AAAI Conference on Artificial Intelligence, AAAI 2021, 5A, pp. 4027–4035. https://doi.org/10.1609/aaai.v35i5.16523.

Di Mattia, F., Galeone, P., De Simoni, M., & Ghelfi, E. (2019). *A survey on GANs for anomaly detection*. http://arxiv.org/abs/1906.11632.

Dromard, J., Owezarski, P., 2020. Study and evaluation of unsupervised algorithms used in network anomaly detection. In: Advances in Intelligent Systems and Computing, 1070. Springer International Publishing. https://doi.org/10.1007/978-3-030-32523-7_28.

Fu, J., Wang, L., Ke, J., Yang, K., & Yu, R. (2023). GANAD: a GAN-based method for network anomaly detection. *World Wide Web*. https://doi.org/10.1007/s11280-023-01160-4.

Fu, Y., Du, Y., Cao, Z., Li, Q., Xiang, W., 2022. A deep learning model for network intrusion detection with imbalanced data. Electronics. (Basel) 11 (6), 1–13. https://doi.org/10.3390/electronics11060898.

González, G.G., Casas, P., Fenández, A., Gómez, G., 2020. Network anomaly detection with net-GAN, a generative adversarial network for analysis of multivariate time-series. In: Proceedings of the SIGCOMM 2020 Poster and Demo Sessions, SIGCOMM 2020, pp. 62–64. https://doi.org/10.1145/3405837.3411393.

Gonzalez, G.G., Tagliafico, S.M., Fernandez, A., Gomez, G., na, J.A., Casas, P., 2023. One model to find them all deep learning for multivariate time-series anomaly detection in mobile network data. IEEE Trans. Netw. Service Manag. 1–16. https://doi.org/10.1109/TNSM.2023.3340146.

Gui, J., Sun, Z., Wen, Y., Tao, D., Ye, J., 2023. A review on generative adversarial networks: algorithms, theory, and applications. IEEE Trans. Knowl. Data Eng. 35 (4), 3313–3332. https://doi.org/10.1109/TKDE.2021.3130191.

Hong, Y., Hwang, U., Yoo, J., Yoon, S., 2019. How generative adversarial networks and their variants work: an overview. ACM. Comput. Surv. 52 (1), 1–43. https://doi.org/10.1145/3301282.

Huang, S., Lei, K., 2020. IGAN-IDS: an imbalanced generative adversarial network towards intrusion detection system in ad-hoc networks. Ad Hoc Netw. 105 https://doi.org/10.1016/j.adhoc.2020.102177.

Huyan, N., Zhang, X., Zhou, H., Jiao, L., 2019. Hyperspectral anomaly detection via background and potential anomaly dictionaries construction. IEEE Trans. Geosci. Remote Sens. 57 (4), 2263–2276. https://doi.org/10.1109/TGRS.2018.2872590.

Iliyasu, A.S., Deng, H., 2022. N-GAN: a novel anomaly-based network intrusion detection with generative adversarial networks. Int. J. Inf. Technol. (Singapore) 14 (7), 3365–3375. https://doi.org/10.1007/s41870-022-00910-3.

Jiang, W., Hong, Y., Zhou, B., He, X., Cheng, C., 2019. A GAN-based anomaly detection approach for imbalanced industrial time series. IEEe Access. 7, 143608–143619. https://doi.org/10.1109/ACCESS.2019.2944689.

Kaplan, M.O., Alptekin, S.E., 2020. An improved BiGAN based approach for anomaly detection. Proc. Comput. Sci. 176, 185–194. https://doi.org/10.1016/j.procs.2020.08.020.

Kim, H.J., Lee, J., Park, C., Park, J.G., 2021. Network anomaly detection based on GAN with scaling properties. In: *International Conference on ICT Convergence, 2021-Octob*, pp. 1244–1248. https://doi.org/10.1109/ICTC52510.2021.9621052.

Kim, J., Jeong, K., Choi, H., Seo, K., 2020. GAN-based anomaly detection in imbalance problems. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 12540 LNCS. Springer International Publishing. https://doi.org/10.1007/978-3-030-65414-6_11.

Kong, K., Kim, K., & Kang, S. (2023). *Robust GAN training with selective conditional matching.*

Koren, O., Koren, M., Peretz, O., 2023. A procedure for anomaly detection and analysis. Eng. Appl. Artif. Intell. 117, 105503 https://doi.org/10.1016/j.engappai.2022.105503.

Kumar, S., Sundaram, H., 2021. Attribute-guided network sampling mechanisms. ACM. Trans. Knowl. Discov. Data 15 (4). https://doi.org/10.1145/3441445.

Kwon, D., Natarajan, K., Suh, S.C., Kim, H., Kim, J., 2018. An empirical study on network anomaly detection using convolutional neural networks. In: Proceedings - International Conference on Distributed Computing Systems, 2018, pp. 1595–1598. https://doi.org/10.1109/ICDCS.2018.00178.

Labonne, M., Olivereau, A., 2020. Unsupervised protocol-based intrusion detection for real-world networks. In: 2020 International Conference on Computing, Networking and Communications (ICNC), pp. 299–303.

Lee, J.H., Park, K.H., 2021. GAN-based imbalanced data intrusion detection system. Pers. Ubiquitous. Comput. 25 (1), 121–128. https://doi.org/10.1007/s00779-019-01332-y.

Li, B., Pi, D., 2020. Network representation learning: a systematic literature review. Neural Comput. Appl. 32 (21) https://doi.org/10.1007/s00521-020-04908-5.

Li, L., Yan, J., Wang, H., Jin, Y., 2021. Anomaly detection of time series with smoothness-inducing sequential variational auto-encoder. IEEe Trans. Neural Netw. Learn. Syst. 32 (3), 1177–1191. https://doi.org/10.1109/TNNLS.2020.2980749.

Li, X., Ouyang, W., Pan, M., Lv, S., Ma, Q., 2023. Continuous learning method of radar HRRP based on CVAE-GAN. IEEE Trans. Geosci. Remote Sens. 61 https://doi.org/10.1109/TGRS.2023.3268219.

Li, Z., Shi, S., Wang, L., Xu, M., Li, L., 2022. Unsupervised generative adversarial network with background enhancement and irredundant pooling for hyperspectral anomaly detection. Remote Sens. (Basel) 14 (5). https://doi.org/10.3390/rs14051265.

Lin, P., Ye, K., Xu, C.Z., 2019. Dynamic network anomaly detection system by using deep learning techniques. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 11513 LNCS. Springer International Publishing. https://doi.org/10.1007/978-3-030-23502-4_12.

Liu, L., Wang, P., Lin, J., Liu, L., 2021a. Intrusion detection of imbalanced network traffic based on machine learning and deep learning. IEEE Access. 9, 7550–7563. https://doi.org/10.1109/ACCESS.2020.3048198.

Liu, S., Li, X., Zhai, Y., You, C., Zhu, Z., Fernandez-Granda, C., Qu, Q., 2021b. Convolutional normalization: improving deep convolutional network robustness and training. Adv. Neural Inf. Process. Syst. 34, 28919–28928.

Liu, Y., Yang, C., Wei, P., Zhou, P., Du, J., 2021c. An ODE-driven level-set density method for topology optimization. Comput. Methods Appl. Mech. Eng. 387, 114159 https://doi.org/10.1016/j.cma.2021.114159.

Mills, R., Marnerides, A.K., Broadbent, M., Race, N., 2022. Practical intrusion detection of emerging threats. IEEE Trans. Netw. Service Manag. 19 (1), 582–600. https://doi.org/10.1109/TNSM.2021.3091517.

Mirsky, Y., Doitshman, T., Elovici, Y., & Shabtai, A. (2018). *Kitsune : an ensemble of autoencoders for online network intrusion detection.* 18–21.

Mohammadi, B., Sabokrou, M., 2019. End-to-end adversarial learning for intrusion detection in computer networks. In: *Proceedings - Conference on Local Computer Networks*. LCN, pp. 270–273. https://doi.org/10.1109/LCN44214.2019.8990759.

Mulay, S.A., Devale, P.R., Garje, G.V., 2010. Intrusion detection system using support vector machine and decision tree. Int. J. Comput. Appl. 3 (3), 40–43. https://doi.org/10.5120/758-993.

Oksuz, K., Cam, B.C., Kalkan, S., Akbas, E., 2021. Imbalance problems in object detection: a review. IEEE Trans. Pattern. Anal. Mach. Intell. 43 (10), 3388–3415. https://doi.org/10.1109/TPAMI.2020.2981890.

Pan, Z., Niu, L., Zhang, L., 2022. UniGAN: reducing mode collapse in GANs using a uniform generator. Adv. Neural Inf. Process. Syst. 35.

Pang, G., Shen, C., Van Den Hengel, A., 2019. Deep anomaly detection with deviation networks. Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 353–362. https://doi.org/10.1145/3292500.3330871.

Pang, G., Van Den Hengel, A., Shen, C., Cao, L., 2021. Toward deep supervised anomaly detection: reinforcement learning from partially labeled anomaly data. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1298–1308. https://doi.org/10.1145/3447548.3467417.

Park, C., Lee, J., Kim, Y., Park, J.G., Kim, H., Hong, D., 2023. An enhanced AI-based network intrusion detection system using generative adversarial networks. IEEE Internet Things. J. 10 (3), 2330–2345. https://doi.org/10.1109/JIOT.2022.3211346.

Park, S., Park, H., 2021. Combined oversampling and undersampling method based on slow-start algorithm for imbalanced network traffic. Computing 103 (3), 401–424. https://doi.org/10.1007/s00607-020-00854-1.

Sabuhi, M., Zhou, M., Bezemer, C.P., Musilek, P., 2021. Applications of generative adversarial networks in anomaly detection: a systematic literature review. IEEE Access. 9, 161003–161029. https://doi.org/10.1109/ACCESS.2021.3131949.

Sadiq Ali Khan, M., 2011. Rule based network intrusion detection using genetic algorithm. Int. J. Comput. Appl. 18 (8), 26–29. https://doi.org/10.5120/2303-2914.

Samariya, D., Thakkar, A., 2023. A comprehensive survey of anomaly detection algorithms. Ann. Data Sci. 10 (3), 829–850. https://doi.org/10.1007/s40745-021-00362-9.

Schmidl, S., Wenig, P., Papenbrock, T., 2022. Anomaly detection in time series: a comprehensive evaluation. Proc. VLDB Endowm. 15 (9), 1779–1797. https://doi.org/10.14778/3538598.3538602.

Selvakumar, K., Karuppiah, M., SaiRamesh, L., Islam, S.H., Hassan, M.M., Fortino, G., Choo, K.K.R., 2019. Intelligent temporal classification and fuzzy rough set-based feature selection algorithm for intrusion detection system in WSNs. Inf. Sci. (N.Y.) 497, 77–90. https://doi.org/10.1016/j.ins.2019.05.040.

Shah J., Das M. IGAN: Intrusion Detection Using Anomaly-Based Generative Adversarial Network. In: Iyer B, Ghosh D, Balas VE, editors. Appl. Inf. Process. Syst., Singapore: Springer Singapore; 2022, p. 371–9.

Shin, A.H., Kim, S.T., Park, G.M., 2023. Time series anomaly detection using transformer-based GAN with two-step masking. IEEE Access 11, 74035–74047. https://doi.org/10.1109/ACCESS.2023.3289921.

Shone, N., Ngoc, T.N., Phai, V.D., Shi, Q., 2018. A deep learning approach to network intrusion detection. IEEE Trans. Emerg. Top. Comput. Intell. 2 (1), 41–50. https://doi.org/10.1109/TETCI.2017.2772792.

Shuvo, M.M.H., Islam, S.K., Cheng, J., Morshed, B.I., 2023. Efficient acceleration of deep learning inference on resource-constrained edge devices: a review. Proc. IEEE 111 (1), 42–91. https://doi.org/10.1109/JPROC.2022.3226481.

Siahpour, S., Li, X., Lee, J., 2022. A novel transfer learning approach in remaining useful life prediction for incomplete dataset. IEEE Trans. Instrum. Meas. 71, 1–11. https://doi.org/10.1109/TIM.2022.3162283.

Sinha, S., Zhang, H., Goyal, A., Bengio, Y., Larochelle, H., Odena, A., 2020. Small-GAN: speeding up GAN Training using Core-Sets. In: 37th International Conference on Machine Learning, ICML 2020, pp. 8952–8962. *PartF168147*2018.

Smiti, A., 2020. A critical overview of outlier detection methods. Comput. Sci. Rev. 38, 100306 https://doi.org/10.1016/j.cosrev.2020.100306.

Soleymanzadeh, R., Kashef, R., 2023. Efficient intrusion detection using multi-player generative adversarial networks (GANs): an ensemble-based deep learning architecture. Neural Comput. Appl. 35 (17), 12545–12563. https://doi.org/10.1007/s00521-023-08398-z.

Syed, A., Shaik, H., 2023. A deep exposition of GAN and its applications. Int. J. Eng. Technol. Manag. Sci. 7 (2), 32–37. https://doi.org/10.46647/ijetms.2023.v07i02.005.

Tharwat, A., 2018. Classification assessment methods. Appl. Comput.d Inform. 17 (1), 168–192. https://doi.org/10.1016/j.aci.2018.08.003.

Torres, P., Catania, C., Garcia, S., Garino, C.G., 2016. An analysis of recurrent neural networks for botnet detection behavior. In: 2016 IEEE Biennial Congress of Argentina, ARGENCON, 2016, pp. 1–6. https://doi.org/10.1109/ARGENCON.2016.7585247.

Umer, M., Saleem, Y., Saleem, M., Aman, N., 2021. A GAN based malware adversaries detection model. In: 2021 15th International Conference on Open Source Systems and Technologies, ICOSST 2021 - Proceedings, pp. 1–9. https://doi.org/10.1109/ICOSST53930.2021.9683863.

Wang, H, Pang, G, Shen, C, Ma, C, 2021. Unsupervised representation learning by predicting random distances. In: IJCAI Int. Jt. Conf. Artif. Intell.. Janua, pp. 2950–2956. https://doi.org/10.24963/ijcai.2020/408.

Wang, H., Wang, Y., Guo, Y., 2021a. A novel approach of unknown network attack detection based on zero-shot learning. In: Proceedings of 2021 IEEE International Conference on Data Science and Computer Application, ICDSCA 2021, pp. 312–318. https://doi.org/10.1109/ICDSCA53499.2021.9650182.

Wang, S., Balarezo, J.F., Kandeepan, S., Al-Hourani, A., Chavez, K.G., Rubinstein, B., 2021b. Machine learning in network anomaly detection: a survey. IEEe Access. 9, 152379–152396. https://doi.org/10.1109/ACCESS.2021.3126834.

Wang, S., Chen, H., Ding, L., Sui, H., Ding, J., 2023. GAN-SR anomaly detection model based on imbalanced data. IEICe Trans. Inf. Syst. (7), 1209–1218. https://doi.org/10.1587/transinf.2022EDP7187. *E106.D*.

Wang, X., Du, Y., Lin, S., Cui, P., Shen, Y., Yang, Y., 2020a. adVAE: a self-adversarial variational autoencoder with Gaussian anomaly prior knowledge for anomaly detection. Knowl. Syst. 190, 105187 https://doi.org/10.1016/j.knosys.2019.105187.

Wang, X., Han, Y., Leung, V.C.M., Niyato, D., Yan, X., Chen, X., 2020b. Convergence of edge computing and deep learning: a comprehensive survey. IEEE Commun. Surv. Tutor. 22 (2), 869–904. https://doi.org/10.1109/COMST.2020.2970550.

Wang, Z., Wang, P., Zhou, X., Li, S., Zhang, M., 2019. FLOWGAN:unbalanced network encrypted traffic identification method based on GAN. In: Proceedings - 2019 IEEE International Conference on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking, ISPA/BDCloud/SustainCom/SocialCom 2019, pp. 975–983. https://doi.org/10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00141.

Xia, X., Pan, X., Li, N., He, X., Ma, L., Zhang, X., Ding, N., 2022. GAN-based anomaly detection: a review. Neurocomputing 493, 497–535. https://doi.org/10.1016/j.neucom.2021.12.093.

Xiong, L., Póczos, B., Schneider, J., 2011. Group anomaly detection using flexible genre models. In: Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011, pp. 1–9.

Xu, Z., Huang, X., Zhao, Y., Dong, Y., B, J.L, 2022. Contrastive Attributed Network Anomaly Detection with Data Augmentation Zhiming, 1. Springer International Publishing. https://doi.org/10.1007/978-3-031-05936-0.

Yaacoub, J.P.A., Salman, O., Noura, H.N., Kaaniche, N., Chehab, A., Malli, M., 2020. Cyber-physical systems security: limitations, issues and future trends. Microprocess. Microsyst. 77, 103201 https://doi.org/10.1016/j.micpro.2020.103201.

Yang, Y., Li, Y., Zhang, W., Qin, F., Zhu, P., Wang, C.X., 2019. Generative-adversarial-network-based wireless channel modeling: challenges and opportunities. IEEE Commun. Mag. 57 (3), 22–27. https://doi.org/10.1109/MCOM.2019.1800635.

Yuan, D., Ota, K., Dong, M., Zhu, X., Wu, T., Zhang, L., Ma, J., 2020. Intrusion detection for smart home security based on data augmentation with edge computing. In: IEEE International Conference on Communications, pp. 0–5. https://doi.org/10.1109/ICC40277.2020.9148632, 2020-June.

Zaman, M., Lung, C.H., 2018. Evaluation of machine learning techniques for network intrusion detection. In: IEEE/IFIP Network Operations and Management Symposium: Cognitive Management in a Cyber World, NOMS 2018, pp. 1–5. https://doi.org/10.1109/NOMS.2018.8406212.

Zavrak, S., Iskefiyeli, M., 2020. Anomaly-based intrusion detection from network flow features using variational autoencoder. IEEe Access. 8, 108346–108358. https://doi.org/10.1109/ACCESS.2020.3001350.

Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., Saeed, J., 2020. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. J. Appl. Sci. Technol. Trends 1 (2), 56–70. https://doi.org/10.38094/jastt1224.

Zenati, H., Foo, C.S., Lecouat, B., Manek, G., & Chandrasekhar, V.R. (2018). *Efficient GAN-based anomaly detection*. https://arxiv.org/abs/1802.06222.

Zenati, H., Romain, M., Foo, C.S., Lecouat, B., Chandrasekhar, V., 2018b. Adversarially learned anomaly detection. In: Proceedings - IEEE International Conference on Data Mining, ICDM, pp. 727–736. https://doi.org/10.1109/ICDM.2018.00088.

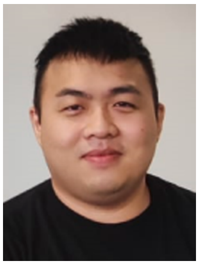
Zhao, B., Xiao, X., Gan, G., Zhang, B., Xia, S., 2020. Maintaining discrimination and fairness in class incremental learning. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 13205–13214. https://doi.org/10.1109/CVPR42600.2020.01322.

Zhao, G., Zhang, C., Zheng, L., 2017. Intrusion detection using deep belief network and probabilistic neural network. In: Proceedings - 2017 IEEE International Conference on Computational Science and Engineering and IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, CSE and EUC 2017, 1, pp. 639–642. https://doi.org/10.1109/CSE-EUC.2017.119.

Zhao, M., Zhang, Y., 2022. GAN-based deep neural networks for graph representation learning. Eng. Rep. 4 (11) https://doi.org/10.1002/eng2.12517.

Zhou, C., Paffenroth, R.C., 2017. Anomaly detection with robust deep autoencoders. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Part F1296*, pp. 665–674. https://doi.org/10.1145/3097983.3098052.

Zong, B., Song, Q., Min, M.R., Cheng, W., Lumezanu, C., Cho, D., Chen, H., 2018. Deep autoencoding Gaussian mixture model. In: ICLR, pp. 1–19.



**Willone Lim** holds a Master's degree in ICT from Swinburne University of Technology and is currently dedicated to his PhD studies in Computer Science at the same institution. His-research passions span across diverse domains, encompassing digital well-being, educational psychology, information security, digital forensics, and the application of AI and machine learning in the field of Cybersecurity.

**Kelvin Yong** received his PhD and BEng(Hons) in Electrical and Electronic Engineering from Swinburne University of Technology in 2016 and 2010 respectively. His-research interests include the area of cybersecurity, specifically information security and cyber threat detection with the focus on the study of improvement of detection schemes leading to effective tools to detect cybersecurity attacks and threats.

**Lau Bee Theng** is currently a professor in ICT and the Associate Dean, Research and Development. Her research interests are:

1. Artificial intelligence in activity recognition, natural scene text recognition, audio event detection, road accidents recognition, wafer surface defect detection, financial risks recognition, aesthetic preference of design objects
2. Smart/creative technologies for the healthcare and well being of people with visual impairment, autism, cerebral palsy, intellectually giftedness, paediatric cancer etc.
3. Alternative learning methods in education and STEM education
4. Rural informatics for bridging the digital divide
5. Impact study of digital devices on students

She has been actively contributing to her research areas with various edited books, peer reviewed journals, conference proceedings, master by research and PhD completions, and funded research projects on assistive technologies for special children, facial expression recognition-based communication, social skills acquisition with animations, real time behavior recognition, smart technologies for the visually impaired, creative art therapies for Autism, STEM education etc. In addition, she also involves actively in community services working with special education schools, NGOs for the people with visual impairment and other disabilities. She is a Senior Member of IEEE and Association of Computing Machinery, Professional Technologist (Malaysia Board of Technologist) and Certified Tester (International Software Testing Qualifications Board). She has involved actively in national and international professional bodies in the executive committees, education, professional and humanitarian projects, winner of IEEE Malaysia Outstanding Volunteer 2018 and IEEE Asia Pacific Outstanding HTA Volunteer 2021.

**Colin Tan** received his BEng (Hons) in Electronics and Computer Engineering, MSc and PhD in Computer Science from Universiti Malaysia Sarawak (UNIMAS). Throughout his research journey, Colin has published a number of articles in top-tier ISI indexed journals. His-research interests include information security, anti-phishing, machine learning, feature selection and image processing.