SSURJET

Sir Syed University Research Journal of Engineering & Technology
2023, Vol. 13, No. 2,
https://doi.org/10.33317/ssurj.615

Comparative Analysis of Anomaly Detection Techniques Using Generative Adversarial Network

Shah Noor[1], Ahthasham Sajid[1], Imranullah Khan[2], Junaid Javaid[1], and Iqra Tabasusum[1]

[1]*Department of Computer Science, Faculty of ICT, BUITEMS, Quetta, Baluchistan, Pakistan*
[2]*College of Electrical Engineering/International College, Hebei University, Baoding, China*

Correspondence Author: Imranullah Khan (khan@hrbeu.edu.cn)

Received January 05, 2023; Revised May 2, 2023; Accepted May 25, 2023

Abstract

Anomaly detection in a piece of data is a challenging task. Researchers use different approaches to classify data as anomalous. These include traditional, supervised, unsupervised, and semi-supervised techniques. A more recently introduced technique is Generative Adversarial Network (GAN), which is a deep learning-based technique. However, it is difficult to choose one anomaly detection algorithm over another because each algorithm stands out with its own performance. Therefore, this paper aims to provide a structured and comprehensive understanding of machine-learning-based anomaly detection techniques. This paper surveys the existing literature on machine-learning-based algorithms for anomaly detection. This paper places a special emphasis on Generative Adversarial Network-based algorithms for anomaly detection since it is the most widely used machine-learning-based algorithm for anomaly detection.

Index Terms: *Anomaly Detection Techniques, Deep Learning, Generative Adversarial Network, Intrusion Detection System, Neural Networks.*

I. BACKGROUND

Anomalies are patterns in data or information that do not behave normally [1]. Anomalies are also known as outliers, peculiarities, contaminants, aberrations, surprises, discordant observations, and exceptions in various application areas. The procedure of discovering an anomaly in data is known as anomaly detection. An anomaly can be explained as the point in a specified time in which system performance is quite different from normal behavior. Therefore, the main objective of anomaly detection is to identify the time stamp where there is a chance of anomaly occurrence. Anomaly detection is also known as outlier identification, novelty identification, exception mining, and deviation identification.

An anomaly in data may change the information, hide the real information, or provide incorrect information. This may lead to various problems in different forms in our daily lives. For example, anomalous MRI pictures may lead to the occurrence of malignant tumors [2]; credit card transaction record anomalies could designate identity theft [3], and in a network, abnormal traffic patterns may make a computer more vulnerable to malicious attacks.

Anomaly detection is used in various areas, such as credit card fraud identification, insurance, health, and medical risk, image processing, astronomical data, cyber-security intrusion identification, sensor networks, safety-critical systems fault identification, and financial surveillance activities [4].

This paper is structured as follows: Section II includes research contribution, types of anomaly, anomaly detection challenges, and three main types of anomaly detection techniques. Section III includes a brief introduction to the Generative Adversarial Network framework, a critical analysis of different GAN approaches, and previous research on GAN. Section IV represents other machine-learning methods for anomaly identification. Section V contains the conclusion of the paper.

II. INTRODUCTION

The main responsibility of anomaly detection is to recognize whether testing data corresponds to the normal data distribution, where the abnormal points in data distributions are known as anomalies. Since the 19th century, anomaly detection methods have been used in research in the statistics community [5], but unfortunately, anomaly detection is still a challenging task. Therefore, the main objective of this paper is to review different Generative Adversarial Network-based approaches used for anomaly detection.

A. Research Contribution

This survey paper provides a structured and comprehensive overview of different anomaly detection techniques and also presents their pros and cons. In recent years, various deep learning-based anomaly identification methods have been developed with lower computational power needs. Hence, this paper aims to review different machine learning techniques, with a significant focus on Generative Adversarial Networks (GAN), which is a deep learning-based technology.

B. Types of Anomaly

An anomaly can be categorized into three types as illustrated in figure I:

a) Point Anomaly:

Point anomaly can be defined as an individual data point that deviates from the remaining data points in a dataset. Because of simplicity, point anomalies are more focused in the research area. For example, if the daily spending of a man is two hundred dollars and on a specific day he spends three hundred dollars, then this situation can be classified as a point anomaly [4].

b) Contextual Anomaly:

Contextual anomalies are also known as conditional anomalies. Conditional anomalies are observations or events that can be considered anomalous in a specific context. Contextual and behavioral attributes define the conditional anomaly. Contextual features describe the context or environment like in time series data time indicates the position and location of the sample. Whereas behavioral features indicate the non-contextual attributes of a sample like indicators that determine whether the sample is anomalous or not in a specific context [6].

c) Pattern Anomaly:

Pattern anomaly is also known as a collective anomaly. A pattern anomaly is the collection or group of alike data that act anomalously with reference to the entire dataset. Individual data may not be considered anomalous in the pattern anomaly. A Group of observations that act anomalous is considered a pattern anomaly.

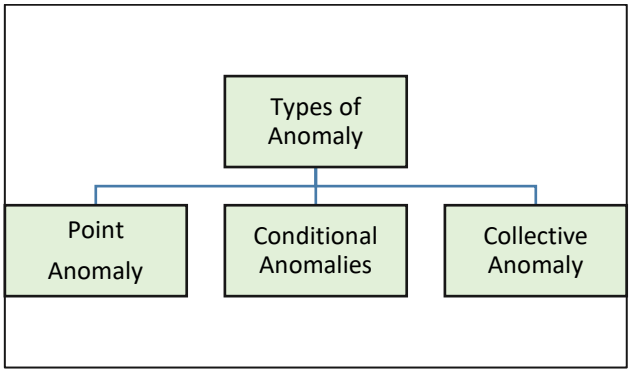

Figure I: Types of Anomalies

C. Challenges of Anomaly Detection

The Challenges of Anomaly Detection are as follows:

1. Traditional algorithms usually cannot properly capture complex structures in datasets, and that's why their performance is lower when identifying anomalies in image or sequence datasets.
2. Generally, large-scale anomaly identification methods are required, because it becomes difficult for traditional algorithms to scale large datasets for anomaly identification.
3. The availability of publicly large-label datasets is the main issue in anomaly detection.
4. In different data domains, the boundary between anomalous and normal behavior is constantly evolving, and at times it is not even properly defined. This may lead to many challenges for both traditional and deep learning-based algorithms.

These challenges make the anomaly detection process difficult to resolve. The majority of existing anomaly identification methods address a particular formulation of the issue, whereas formulation is influenced by different factors such as the nature or type of data, label dataset availability, anomaly types to be identified, and so on. Frequently these factors are resolved by the application domain where anomalies require to be identified.

D. Types of Anomaly Detection Techniques

Anomaly detection can be classified into three main techniques based on the way historical data is processed: unsupervised, semi-supervised, and supervised. Each of these three techniques has several different types.

The first technique is statistical process control which uses univariate or multivariate analysis for tracking and managing the quality of the manufacturing process [7]. Statistical process control usually detects changes in the variance process and mean process. Statistical process control uses CUSUM control and Shewhart control charts methods as the univariate approach for identifying mean shifts [8]. Unfortunately, multivariate processes need identically distributed and self-standing assumptions where the assumption is usually contravened in reality [9].

The second technique is supervised machine learning, where supervised machine learning uses a predictive classification approach for normal and anomalous class datasets. Supervised models are trained using labeled datasets and then data is automatically classified into corresponding classes [10]. Different supervised machine-learning approaches are used for anomaly detection. These include the Bayes classifier approach [11], neural network method [12], multivariate regression [13], support vector data description approach [11], Fisher discriminant analysis [14], support vector machine [15], and tree-structured learning approach [12]. Supervised machine learning models rely more on the accessibility of label training datasets. Anomalies data are uncommon so attaining correct anomaly-labeled class data is a difficult task.

The third technique is the unsupervised learning approach which is also called the undirected classification training approach because for training it does not need any labeled classes' dataset. Unsupervised learning can manage a great number of process data. That is why in different industrial procedures, the unsupervised learning approach is used for anomaly detection. Principal component analysis [16] and partial least [17] are unsupervised methods used for anomaly detection where both of these use a multivariate data analysis approach. Unfortunately, these approaches are suitable only for highly correlated data, and they also need the data to track multivariate Gaussian distribution [18].

Deep learning performs various high-dimensional machine learning tasks and leaves behind manual feature engineering. Deep learning is used in different areas like image processing or classification [19], speech identification [20], and natural image processing [21]. Generative Adversarial Networks attain state-of-the-art results in high-dimensional generative modeling.

III. GENERATIVE ADVERSARIAL NETWORKS

Recently, another unsupervised approach Generative Adversarial Network (GAN) has been used for anomaly detection. It uncovers and learns the patterns or regularities from a given dataset. Using the results of this learning process new examples can be generated from the real data distribution.

The Generative Adversarial Network is the unsupervised machine learning method that automatically identifies and learns regularities as well as patterns from the given dataset [22]. It then uses the results of these learning processes to generate new samples that look like the actual dataset. Generative Adversarial Networks use two sub-models for generating new samples: one of them is a generator used for generating new samples, and the other one is a discriminator used for classifying samples as either fake or real. Both these models are trained using turning on or off the parameter simultaneously until the discriminator model gets fooled and the generator creates plausible examples. As discussed earlier, Generative Adversarial Networks uncover and learn patterns of data. So once the generator is able to generate normal samples, abnormalities that appear in samples can be detected. The structure of the Generative Adversarial Network is illustrated in figure II.

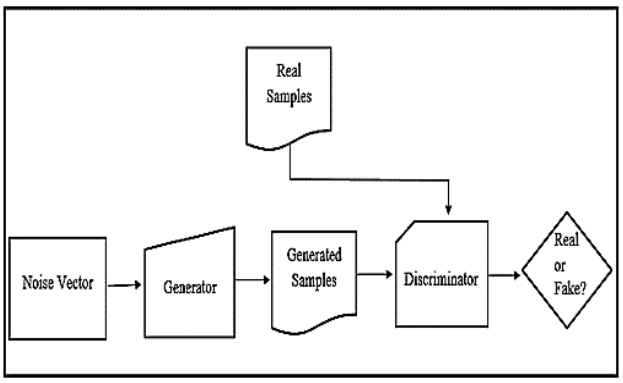

Figure II: GAN Structure

IV. PREVIOUS RESEARCH ON GAN

Anomaly detection has a long history. Multiple types of research have been carried out to detect anomalous or unusual patterns in the dataset. Recent Generative Adversarial Networks achieve higher performance as compared to other previous deep generative approaches.

Schlegl et al [23] proposed a deep convolutional Generative Adversarial Network for classifying anomalies in imaging data that can be used as a candidate for disease markers. Using a healthy dataset author trained the model and achieved 89% accuracy. The proposed model can identify different anomalies like HRF and retinal fluid in data.

Zenati et al [24] train the generative adversarial model using the KDD99 network intrusion dataset and MNIST image dataset for anomaly detection where they achieve state-of-art results. During the training process simultaneously this model learns an encoder for making this model efficient in test time. For the training model using the MNIST dataset author used 80% normal data and tested this model using the remaining 20% data that included both normal and anomalous data. Whereas for the KDD99 dataset author used 50% of the data for training and the remaining 50% for testing.

Yotam et al [25] proposed a Multi Discriminator Generative Adversarial Network for anomaly identification. The author uses two discriminators, where one of them is used for confirming that the generated sample is valid and the other discriminator works as an auto-encoder used for anomaly classifier. Both of these discriminators use different cost functions. That model is evaluated using ten datasets of different features and domains.

The authors in [26] propose a Generative Adversarial Network-based approach for complex multi-process cyber-physical systems to identify cyber-attack-based anomalies. For the base of the framework, the author uses LSTM and recurrent neural networks. GAN is trained using a normal multivariate series of data. The model is evaluated using a secure water treatment testbed dataset.

Tharindu et al [27] use a bidirectional GAN approach for industrial systems anomaly identification. Where BIGAN shows better accuracy as compared to other traditional GAN-based approaches. BIGAN work on an instant anomaly score calculation scheme so this method is most suitable for large-scale production environment.

Fei Dong et al [28] proposed a GAN-based model for video anomaly identification. The author used two discriminators and one generator for the proposed framework. The generator is used to predict future video frames and one of the discriminators distinguishes input frames as generated or original frames. Other discriminators discriminate the optical frame as true or false. The model is evaluated using ShanghaiTech, UCSD Ped2, and CUHK Avenue datasets. Training data only includes normal data and testing data includes both normal and anomalous data.

The technique demonstrated in [29] uses Sequence Generative Adversarial, Auto-encoder, and Gated Recurrent Unit to resolve the problem of imbalanced log messages. Sequence Generative Adversarial Network oversampled the negative logs then auto-encoder works as the feed-forward network that is used for extracting important features and information from resulting data. A Gated recurrent unit was used for anomaly identification. The model is tested using OpenStack and BGL datasets.

The authors in [30] used two Generative Adversarial Network architectures one of them is AnoGAN and the other one is ALAD for network anomaly classification.

AnoGAN work is based on standard Generative Adversarial Networks and ALAD architecture construct using bi-directional Generative Adversarial Networks with numerous enhancements that provide fast detection and stabilize the process of GAN training. The model is developed using a large number of hidden layers. For the evaluation and training process, the author used both synthetic traffic and realistic traffic captures that were generated using simulation platforms.

Bashar et al [31] propose the TAnoGan method for time series anomaly detection, where data points are accessible in small numbers. The generator learns the normal distribution of the dataset and mapping is used to map the sequence of data to latent space. TAnoGan detects anomalies in small datasets more effectively. The model is tested using 46 real-world time series datasets.

The technique demonstrated in [32] uses a Generative Adversarial Network for learning the normal behavior of firewall time series data and then the author applies various anomaly classification techniques for anomaly detection. Propose method used to identify an anomaly in network traffic logged using the firewall. Where author also uses two different encoding approaches namely binary and embedding encoding for data.

In 2021 Laya et al [33] worked on the combination of Generative Adversarial Networks and auto-encoders to classify anomalies in image datasets. The author used SVHN, MNIST, and CIFAR10 as natural datasets and acute lymphoblastic leukemia medical datasets for the evaluation process. While decreasing the inference time auto-encoder helps the GAN model to improve its previous results. Laya et al also use a small dataset for the training model where it performs well.

The technique proposed in [34] used a long-short-term memory network-based GAN named LogGan for system logs anomaly identification. Based on patterns, LogGan identifies log-level anomalies. LogGan uses permutation event modeling for distinctive unusual upcoming events which are based on temporal system logs. For evaluating the model author used two real-world datasets. The author achieved effective results for log-level anomaly identification.

Chen et al [35] use the GAN method for anomaly identification in industrial control systems. For learning latent data distribution encoder–decoder–encoder method based dual Generative Adversarial Network is used. To effectively learn the marginal distribution of the training data, a parameter-free dynamic method is proposed. In the last optimized anomaly, the score simplifies whether an example is anomalous or not by using data of marginal distribution and learned normal distribution.

Patil [36] uses principal component analysis and BiGAN algorithms together for network traffic anomaly detection. PCA is used for feature extraction and dimensional reduction. BiGAN is used for network traffic anomaly identification. The method was tested using the KDDCUP-99 dataset and the result was compared with other algorithms also.

Table I: Critical Analysis of State-of-Art Techniques

S. No.	Author	Year	Techniques	Methodology	Dataset	Pros	Cons	Accuracy Achieved
1.	Schlegl et al. [23]	2017	Deep Convolutional Generative Adversarial Network	Proposed AnoGAN model for detecting an anomaly in imaging data.	Real medical image dataset	As compared to previous work proposed model achieves good accuracy	The model required more improvement for better results	89%
2.	Zenati et al. [24]	2018	Generative Adversarial Network	Work on the BiGAN model for classifying anomalies in imaging datasets and network intrusion datasets	MNIST, KDD99	This model is greatly competitive as it shows good results in the KDD99 dataset	High inference time	92%
3.	Yotam et al. [25]	2018	Multi-Discriminator Generative Adversarial Network	Proposed MDGAN for anomaly classification. It uses two discriminators.	OpenML and Outlier Detection DataSets	Improve the performance through the different dataset	High model training time for performing well	95%
4.	Dan Li et al. [26]	2019	Generative Adversarial Network	Use GAN-AD for cyber-physical systems	SWaT Dataset	A high identification rate with a low false-positive rate	Treat all variables equally in one plain framework.	94%
5.	FEI DONG et al. [28]	2020	Generative Adversarial Network	Propose GAN based model for video anomaly detection	UCSD Ped2, CUHK Avenue, ShanghaiTech	Attain larger gaps which result in a greater identification rate	The model required enhancement for better prediction anomaly in the video dataset.	73%
6.	Farzad et al. [34]	2019	Sequence Generative Adversarial Network, auto-encoder, gated recurrent unit	Proposed GAN-based model for solving imbalanced log message problems.	OpenStack dataset, BGL dataset	Even if data is imbalanced model provides state-of-the-art results	----------	98%
7.	Tram et al. [30]	2020	Generative Adversarial Network	Use the GAN for network anomaly identification	UNSW-NB15, CICIDS2017, Stratosphere IPS	Compared to other deep learning approaches this approach shows state of art results	The model needs enhancement for better results	90%
8.	Sandeep et al. [32]	2021	Generative Adversarial Network	Used Generative Adversarial Network for classification of	Installed FortiGate firewall for obtaining data	Apply on rules of time series reconstruction.	Not work properly for continuously updating data.	99.7%

				anomalous log messages				
9.	Laya et al. [33]	2021	Generative Adversarial Network with auto-encoder	Proposed GAN-based model combination with auto-encoder and new scoring function for identifying anomaly	MNIST, CIFAR10, SVHN, Acute Lymphoblastic Leukemia dataset	Even on a small dataset model performed quite well	The model required improvement for further learning discriminative representations	97%
10.	Bin Xia1 et al. [29]	2021	GAN with Long short-term memory network	Proposed LSTM-based Generative Adversarial Network for detecting the log-level anomaly	HDFD, BGL Real-world dataset	Show effective results for the task of log-level anomaly identification.	For the training model, temporal information and signature data were only used	98%

Accuracy Comparison of Existing Techniques after the extensive literature review and critical analysis has also been presented in figure III.

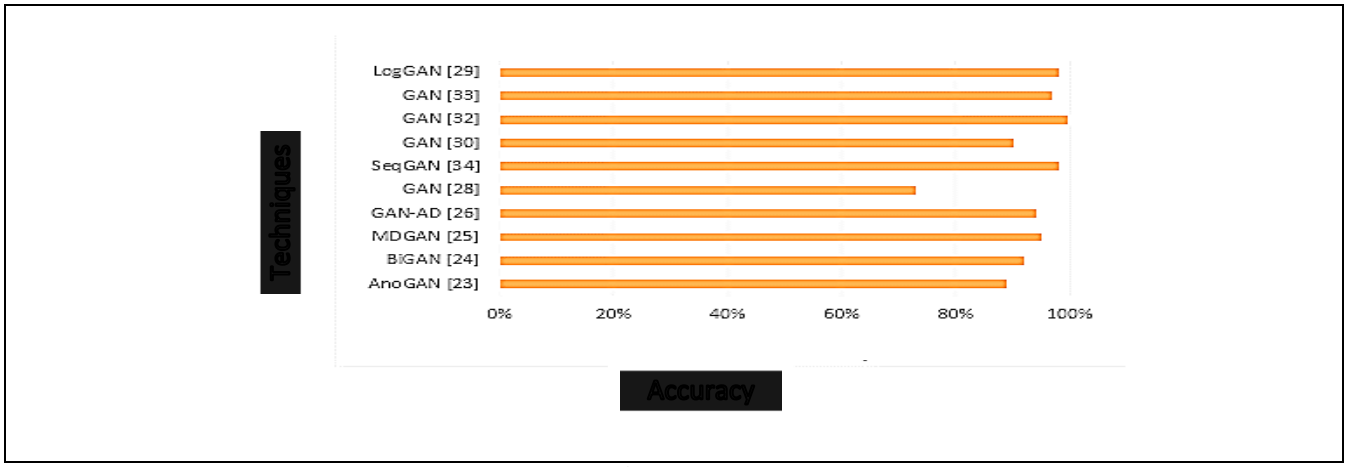

Figure III: State-of-Art Techniques Vs. Accuracy

Figure III previously describe the accuracy achieved against each of the existing technique to cop anomaly detection by other researchers till yet as per table II.

V. OTHER MACHINE LEARNING TECHNIQUES FOR ANOMALY DETECTION

Anomaly detection methods depend on input data labels and input data types. It can be either unlabeled or labeled data, problem statements, or preferred output. Figure IV represents different machine learning algorithms that are used for anomaly detection.

A. Classification-Based Anomaly Detection

In the classification-based anomaly detection method, a classifier is used to split the space regions among anomalous data and normal data by observing data instances in a certain feature space. Neural networks, Bayes networks, rule-based, and support vector machine techniques are included in classification-based anomaly identification techniques. Classification-based anomaly identification has three ways of processing data: one class, two classes, and three classes. In one class only one label class is considered.

In two classes, two label classes are considered based on attribute space. Multiclass considers more than two classes based on attribute space.

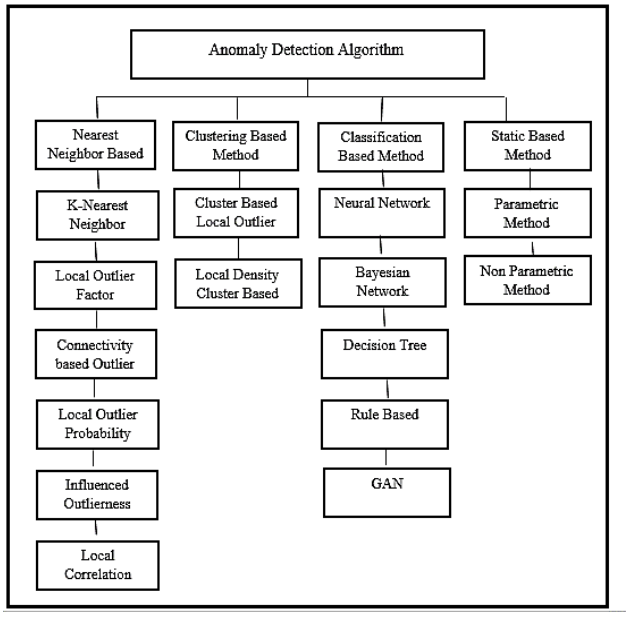

Figure IV: Machine Learning Techniques

a) Neural Networks:

The ability of a neural network to classify data has been applied to the identification of network anomalies. Although neural networks have been used in a variety of application domains, including voice and image processing, they are computationally intensive. A neural

network has been combined with various methods, such as a statistical method and its variations, for the detection of network anomalies. A multi-layer perceptron with a gravitational search algorithm has also been used [37] and one hidden layer for a network intrusion detection system where the dataset consists of labeled flow data.

b) Bayes Network Based:

Bayesian network evaluates the posterior probability of a specific event using a few given observations. Amor et al [38] propose a network intrusion detection system on the KDD'99 dataset using level three granularity Bayes networks and also show a comparison between decision trees and Bayes Network.

c) Support Vector Machine:

The Support Vector Machine's fundamental idea is to create a hyperplane that increases the distance, among the positive and negative classes. Zhang et al [39] use various types of unlabeled qualitative data (KPIs) and quantitative datasets to perform a one-class support vector machine for network intrusion detection system

d) Rule-Based Method:

This method learns rules that summarize a system's normal performance. Duffield et al [40] present a real-time rule-based method for detecting anomalies in internet protocol packet flows in a communication network.

B. Nearest Neighbor-Based Anomaly Detection

Nearest Neighbor is the simplest anomaly identification technique where it assumes normal data samples occur in their nearest neighborhoods whereas anomalous instances occur far from their nearest neighbors. For the nearest neighbor method, it is required to define similarity or distance measures among two data instances. K Nearest neighbor and relative density classes are included in the nearest neighbor anomaly detection method.

a) K Nearest Neighbor:

This method computes the similarities with nearby data points. The score-based non-parametric adaptive anomaly detection technique was proposed [41] and implemented on the banana dataset using a KNN Graph on n-point nominal data.

b) Relative Density:

For each data point, this approach computes the neighborhood density.

C. Clustering-Based Anomaly Detection

Clustering work is based on the definition of pairwise similarity or distance function for grouping similar data instances into clusters. Clustering is a semi-supervised or unsupervised method. Clustering-based anomaly identification is classified into three different groups. In the first category, normal data samples belong to clusters whereas anomalous data samples do not belong to the cluster. In the second category, normal sample data are adjacent to their nearest cluster centroid while anomalous sample data are away from their nearest cluster centroid. In the last category, normal sample data belong to large or dense clusters whereas anomalous data samples belong to small or sparse clusters. Density-based cluster is a subcategory of clustering-based anomaly identification. For example, Kiss et al [42] propose a K-mean clustering-based method for identifying cyber-attacks that cause an anomaly in Networked Critical Infrastructure.

D. Statistical-Based Anomaly Detection

Statistical anomaly identification work is based on the following key assumption: normal data samples will be in the high probability region of the stochastic model, while anomalous data instances will be in the low probability region. This method is further divided into two categories: parametric techniques and non-parametric techniques.

a) Parametric Technique:

Parametric distribution is used for data generating with probability density function and parameters in time series anomaly detection, such as ARIMA, ARMA, and linear regression. An example of a parametric-based method is the regression model. Yip et al [43] represent network intrusion classification in grids through gathering energy utilization meter readings from the Sustainable Energy Authority of Ireland's SM dataset.

b) Non-Parametric Technique:

In non-parametric methods, the model is determined by real-time datasets rather than defined theoretically apriority. Compared to the parametric approach, this method makes fewer data-related assumptions. Smrithy et al [44] apply a non-parametric approach to big data streams, detecting anomalous online access requests to illegal shared resources at runtime.

E. Information-Theoretic Based Anomaly Detection

The work of information-theoretic anomaly detection is based on the following key assumption: anomalous data instances in the data set are irregularities in the information content. The information-theoretic method analyzes the information content of the dataset using various approaches like relative entropy, Kolmogorov complexity, and entropy. For identifying anomalies in univariate datasets, Lee et al [45] use information-theoretic methods like conditional entropy, information gain, and relative conditional entropy. Using the MAWILab dataset, Callegari et al [46] represent an intrusion identification system in a network based on the change in entropy measures.

F. Spectral-Based Anomaly Detection

Spectral techniques use a combination of features for finding approximate data that captures the bulk of the inconsistency in the data. This technique is based on the following supposition: data can be transformed into a lower-dimensional subspace while anomalous and normal data samples appear remarkably diverse. Spectral techniques can work in both unsupervised and semi-supervised settings. For identifying score-based abnormalities over wireless sensor networks using graph-based filtering, Egilmez and Ortega [47] use PCA, ROCs, and an auto-regressive model.

G. Graph-Based Anomaly Detection

Graph-based anomaly detection work is based on two assumptions: first, irregular entities or nodes are separated from the rest of the graph, and second, irregular entities or nodes share a network with the remaining graph. Graph-based anomaly detection is categorized based on dynamic and static graphs. Akoglu et al [48] represent graph-based anomaly detection in various fields like finance, law enforcement, security, and health care. Vasseur et al [49] use graph-based anomaly detection to identify the cause of anomalous events through calculating correlations.

Table II: Anomaly Detection Techniques, Applications, and Algorithms

S. No.	Category	Sub Categories	Application	Algorithms	Pros	Cons
1.	Classification	Neural Network	Network Intrusion Classification	Auto-encoder and Boltzmann machines [50]	Powerful computing algorithms for differentiating amongst data instances that belong to dissimilar classes	For training, these algorithms depend upon on availability of precise label datasets.
				GSA, multilayer perceptron [37]		
				ICSVM and deep belief networks [51]		
			SHM System	DTrees, byes net [38]		
			Dynamic Network	SMOTE, LSTM, AM [52]		
				LSTM, deep belief network, stack auto encoder [53]		
			5G Network for Cyber Security	Deep neural networks [54]		
		Support Vector Machine	Network Intrusion Detection	CCCP or ramp-OCSVM algorithms [55]		
				GSS [56]		
			Wireless Sensor Network	Clustering and OCSVM [57]		
				doOCSVM [58]		
		Rule Based	Telephone Call Anomaly	Dynamic rule-based [40]		
			Disease Outbreak	Rule-based [59]		
			Communication Networks	KNN graph [60]		
		Bayes Network Based	Social Networks	One class support vector machine [61]		
			Disease Outbreak Identification	Bayesian network [62]		
			Network Intrusion Identification	Bayes network [63]		
2.	Nearest neighbor	K Nearest Neighbor	Adaptive Anomaly Classification	Self-organizing based K nearest neighbor [6, 64]	From nature, these algorithms are unsupervised so it does not require label datasets. Amendment to various data types.	Testing computation is slow. These algorithms missed anomaly data instances as it have not sufficient neighbors.
			Steel Plant	KNN or genetic [65]		
			Network	Local Outlier Factor, traditional regression model [66]		
		Relative Density	Dam Protection Monitoring	Local Outlier Factor [67]		
3.	Clustering	Density-Based Clusters	Networked Critical Infrastructure	Different clustering views [68]	Fast testing process. These algorithms are unsupervised in nature.	Anomalous data instances may lie in normal instances clusters because anomalous data may not make major clusters.
			Cloud Environment	HSE algorithms [69]		
			Multi-View Grouping	Artificial neural network with fuzzy C-mean [70]		
4.	Information-Theoretic	Information-Theoretic Methods	Intrusion Identification	Based on entropy [71]	Unsupervised in nature. Regarding distributions of the dataset no assumptions.	Results rely on the selection of measures.
			Modern Vehicle Controller Area Network	Based on entropy [72]		
5.	Statistical	Parametric Methods	Sensors Network	Regression, support vector machine [73]	Robust distribution may work as unsupervised anomaly detection	Distribution may be false for high-dimensional real-world datasets.
			Energy Consumption	Parametric technique [43]		
6.	Spectral	Spectral-Based Anomaly Detection	Intrusion Identification	Auto-encoder [74]	Unsupervised in nature and applicable for high dimensional datasets.	It works only when normal or anomalous data instances are separated in inferior dimensions.
			Nonlinear Reduction	Auto-encoder [75]		
			Web Application	Autoregressive, PCA [47]		
7.	Graph	Graph-Based Anomaly Detection	Intrusion Identification	Correlations graph-based [76]	Applicable for time series data.	Work only for pairwise connection.
			Mobile Communication Networks	Graph-based [49]		

VI. CONCLUSION

Anomaly is nonconforming patterns in data that do not look or work like normal behavior. Because of anomalies, real data may change, hide, or provide improper information which may cause various problems. Different traditional or machine-learning approaches are used for anomaly detection. However, Generative Adversarial Networks achieve a high performance rate. This survey discussed different approaches of Generative Adversarial Networks for anomaly detection. This will assist not only in understanding techniques but also in highlighting their pros and cons. Future work could be on this thorough survey for the researchers to investigate highlighted limitations further in the current techniques; and design and deploy a more efficient anomaly detection technique which can beat the current achieved accuracy till yet.

Acknowledgment

I would like to thank Almighty Allah first and then all the co-authors for their full support and hard work that we have completed this research study successfully.

Authors Contributions

The Idea and survey guidance was suggested for this research work by Ahthasham Sajid and Imranullah Khan. The write-up and English proofreading were done by Shahnoor. The revision of the draft was improved by Junaid Javed and Iqra Tabassum, and plagiarism was improved by Ahthasham Sajid.

Conflict of Interest

All the researchers declare no conflict of interest.

Data Availability Statement

The testing data is available in this paper.

Funding

This research received no external funding.

References

[1] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR), 41*(3), 1-58.

[2] Spence, C., Parra, L., & Sajda, P. (2001, December). Detection, synthesis, and compression in mammographic image analysis with a hierarchical image probability model. In *Proceedings IEEE workshop on mathematical methods in biomedical image analysis (MMBIA 2001)* (pp. 3-10). IEEE.

[3] Aleskerov, E., Freisleben, B., & Rao, B. (1997, March). Cardwatch: A neural network based database mining system for credit card fraud detection. In *Proceedings of the IEEE/IAFE 1997 computational intelligence for financial engineering (CIFEr)* (pp. 220-226). IEEE.

[4] Ahmed, M., Mahmood, A. N., & Islam, M. R. (2016). A survey of anomaly detection techniques in financial domain. *Future Generation Computer Systems, 55*, 278-288.

[5] Chandola, V., Banerjee, A., & Kumar, V. (2010). Anomaly detection for discrete sequences: A survey. *IEEE transactions on knowledge and data engineering, 24*(5), 823-839.

[6] Song, X., Wu, M., Jermaine, C., & Ranka, S. (2007). Conditional anomaly detection. *IEEE Transactions on knowledge and Data Engineering, 19*(5), 631-645.

[7] Ye, N., Vilbert, S., & Chen, Q. (2003). Computer intrusion detection through EWMA for autocorrelated and uncorrelated data. *IEEE transactions on reliability, 52(*1), 75-82.

[8] Ryan, T. P. (2011). *Statistical methods for quality improvement*. John Wiley & Sons.

[9] Ye, N., & Chen, Q. (2001). An anomaly detection technique based on a chi-square statistic for detecting intrusions into information systems. *Quality and reliability engineering international, 17*(2), 105-112.

[10] Koturwar, P., Girase, S., & Mukhopadhyay, D. (2015). A survey of classification techniques in the area of big data. *arXiv preprint arXiv:1503.07477*.

[11] Xiao, F., Zhao, Y., Wen, J., & Wang, S. (2014). Bayesian network based FDD strategy for variable air volume terminals. *Automation in Construction, 41*, 106-118.

[12] Li, D., Zhou, Y., Hu, G., & Spanos, C. J. (2016). Fault detection and diagnosis for building cooling system with a tree-structured learning method. *Energy and Buildings, 127*, 540-551.

[13] Mustafaraj, G., Chen, J., & Lowry, G. (2010). Development of room temperature and relative humidity linear parametric models for an open office using BMS data. *Energy and Buildings, 42*(3), 348-356.

[14] Jaikumar, P., Gacic, A., Andrews, B., & Dambier, M. (2011, May). Detection of anomalous events from unlabeled sensor data in smart building environments. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2268-2271). IEEE.

[15] Mulumba, T., Afshari, A., Yan, K., Shen, W., & Norford, L. K. (2015). Robust model-based fault diagnosis for air handling units. *Energy and Buildings, 86*, 698-707.

[16] Li, S., & Wen, J. (2014). A model-based fault detection and diagnostic methodology based on PCA method and wavelet transform. *Energy and Buildings, 68*, 63-71.

[17] He, X., Wang, Z., Liu, Y., & Zhou, D. H. (2013). Least-squares fault detection and diagnosis for networked sensing systems using a direct state estimation approach. *IEEE Transactions on Industrial Informatics, 9*(3), 1670-1679.

[18] Dai, X., & Gao, Z. (2013). From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis. *IEEE Transactions on Industrial Informatics, 9*(4), 2226-2238.

[19] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84-90.

[20] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

[21] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in neural information processing systems, 27*.

[22] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). *Generative adversarial networks*. *Communications of the ACM, 63*(11), 139-144.

[23] Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., & Langs, G. (2017, May). Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International conference on information processing in medical imaging* (pp. 146-157). Cham: Springer International Publishing.

[24] Zenati, H., Foo, C. S., Lecouat, B., Manek, G., & Chandrasekhar, V. R. (2018). Efficient gan-based anomaly detection. *arXiv preprint arXiv:1802.06222*.

[25] Intrator, Y., Katz, G., & Shabtai, A. (2018). Mdgan: Boosting anomaly detection using\\multi-discriminator generative adversarial networks. *arXiv preprint arXiv:1810.05221*.

[26] Li, D., Chen, D., Goh, J., & Ng, S. K. (2018). Anomaly detection with generative adversarial networks for multivariate time series. *arXiv preprint arXiv:1809.04758*.

[27] Kumarage, T., Ranathunga, S., Kuruppu, C., De Silva, N., & Ranawaka, M. (2019, July). Generative adversarial networks (GAN) based anomaly detection in industrial software systems. In *2019 Moratuwa Engineering Research Conference (MERCon)* (pp. 43-48). IEEE.

[28] Dong, F., Zhang, Y., & Nie, X. (2020). Dual discriminator generative adversarial network for video anomaly detection. *IEEE Access*, *8*, 88170-88176.

[29] Xia, B., Bai, Y., Yin, J., Li, Y., & Xu, J. (2021). Loggan: a log-level generative adversarial network for anomaly detection using permutation event modeling. *Information Systems Frontiers*, *23*, 285-298.

[30] Truong-Huu, T., Dheenadhayalan, N., Pratim Kundu, P., Ramnath, V., Liao, J., Teo, S. G., & Praveen Kadiyala, S. (2020, October). An empirical study on unsupervised network anomaly detection using generative adversarial networks. In *Proceedings of the 1st ACM Workshop on Security and Privacy on Artificial Intelligence* (pp. 20-29).

[31] Bashar, M. A., & Nayak, R. (2020, December). TAnoGAN: Time series anomaly detection with generative adversarial networks. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1778-1785). IEEE.

[32] Kulyadi, S. P., Mohandas, P., Kumar, S. K. S., Raman, M. S., & Vasan, V. S. (2021, July). Anomaly detection using generative adversarial networks on firewall log message data. In *2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)* (pp. 1-6). IEEE.

[33] Sevyeri, L. R., & Fevens, T. (2021). On the effectiveness of generative adversarial network on anomaly detection. *arXiv preprint arXiv:2112.15541*.

[34] Farzad, A., & Gulliver, T. A. (2019). Oversampling log messages using a sequence generative adversarial network for anomaly detection and classification. *arXiv preprint arXiv:1912.04747*.

[35] Chen, L., Li, Y., Deng, X., Liu, Z., Lv, M., & Zhang, H. (2022). Dual auto-encoder GAN-based anomaly detection for industrial control system. *Applied Sciences*, *12*(10), 4986.

[36] Patil, R., Biradar, R., Ravi, V., Biradar, P., & Ghosh, U. (2022). Network traffic anomaly detection using PCA and BiGAN. *Internet Technology Letters*, *5*(1), e235.

[37] Jadidi, Z., Muthukkumarasamy, V., Sithirasenan, E., & Sheikhan, M. (2013, July). Flow-based anomaly detection using neural network optimized with GSA algorithm. In *2013 IEEE 33rd international conference on distributed computing systems workshops* (pp. 76-81). IEEE.

[38] Amor, N. B., Benferhat, S., & Elouedi, Z. (2004, March). Naive bayes vs decision trees in intrusion detection systems. In *Proceedings of the 2004 ACM symposium on Applied computing* (pp. 420-424).

[39] Zhang, R., Zhang, S., Lan, Y., & Jiang, J. (2008, March). Network anomaly detection using one class support vector machine. In *Proceedings of the International MultiConference of Engineers and Computer Scientists* (Vol. 1).

[40] Duffield, N., Haffner, P., Krishnamurthy, B., & Ringberg, H. A. (2016). Systems and methods for rule-based anomaly detection on IP network flow. *U.S. Patent No. 9,258,217*. Washington, DC: U.S. Patent and Trademark Office.

[41] Zhao, M., & Saligrama, V. (2009). Anomaly detection with score functions based on nearest neighbor graphs. *Advances in neural information processing systems*, *22*.

[42] Kiss, I., Genge, B., Haller, P., & Sebestyén, G. (2014, September). Data clustering-based anomaly detection in industrial control systems. In *2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP)* (pp. 275-281). IEEE.

[43] Yip, S. C., Wong, K., Hew, W. P., Gan, M. T., Phan, R. C. W., & Tan, S. W. (2017). Detection of energy theft and defective smart meters in smart grids using linear regression. *International Journal of Electrical Power & Energy Systems*, *91*, 230-240.

[44] Smrithy, G. S., Munirathinam, S., & Balakrishnan, R. (2016, December). Online anomaly detection using non-parametric technique for big data streams in cloud collaborative environment. In *2016 IEEE International Conference on Big Data (Big Data)* (pp. 1950-1955). IEEE.

[45] Lee, W., & Xiang, D. (2000, May). Information-theoretic measures for anomaly detection. In *Proceedings 2001 IEEE Symposium on Security and Privacy. S&P 2001* (pp. 130-143). IEEE.

[46] Callegari, C., Giordano, S., & Pagano, M. (2017). An information-theoretic method for the detection of anomalies in network traffic. *Computers & Security*, *70*, 351-365.

[47] Egilmez, H. E., & Ortega, A. (2014, May). Spectral anomaly detection using graph-based filtering for wireless sensor networks. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1085-1089). IEEE.

[48] Akoglu, L., Tong, H., & Koutra, D. (2015). Graph based anomaly detection and description: a survey. *Data mining and knowledge discovery*, *29*(3), 626-688.

[49] Vasseur, J. P., Mermoud, G., & Mota, J. C. (2016). Event correlation in a network merging local graph models from distributed nodes . *U.S. Patent Application No. 14/605,916*.

[50] Van, N. T., & Thinh, T. N. (2017, July). An anomaly-based network intrusion detection system using deep learning. In *2017 international conference on system science and engineering (ICSSE)* (pp. 210-214). IEEE.

[51] Erfani, S. M., Rajasegarar, S., Karunasekera, S., & Leckie, C. (2016). High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognition*, *58*, 121-134.

[52] Maimó, L. F., Gómez, Á. L. P., Clemente, F. J. G., Pérez, M. G., & Pérez, G. M. (2018). A self-adaptive deep learning-based system for anomaly detection in 5G networks. *IEEE Access*, *6*, 7700-7712.

[53] Garcia, S., Grill, M., Stiborek, J., & Zunino, A. (2014). An empirical comparison of botnet detection methods. *Computers & Security*, *45*, 100-123.

[54] Malaiya, R. K., Kwon, D., Kim, J., Suh, S. C., Kim, H., & Kim, I. (2018, March). An Empirical Evaluation of Deep Learning for Network Anomaly Detection. In *2018 International Conference on Computing, Networking and Communications (ICNC)* (pp. 893-898). IEEE.

[55] Tian, Y., Mirzabagheri, M., Bamakan, S. M. H., Wang, H., & Qu, Q. (2018). Ramp loss one-class support vector machine; a robust and effective approach to anomaly detection problems. *Neurocomputing*, *310*, 223-235.

[56] Anil, S., & Remya, R. (2013, July). A hybrid method based on genetic algorithm, self-organised feature map, and support vector machine for better network anomaly detection. In *2013 Fourth international conference on computing, communications and networking technologies (ICCCNT)* (pp. 1-5). IEEE.

[57] Harrou, F., Dairi, A., Taghezouit, B., & Sun, Y. (2019). An unsupervised monitoring procedure for detecting anomalies in photovoltaic systems using a one-class support vector machine. *Solar Energy*, *179*, 48-58.

[58] Miao, X., Liu, Y., Zhao, H., & Li, C. (2018). Distributed online one-class support vector machine for anomaly detection over networks. *IEEE transactions on cybernetics*, *49*(4), 1475-1488.

[59] Gopal, R. K., & Meher, S. K. (2007, November). A rule-based approach for anomaly detection in subscriber usage pattern. In *Proceedings of World Academy of Science, Engineering and Technology* (pp. 396-399).

[60] Zhao, M., & Saligrama, V. (2009). Anomaly detection with score functions based on nearest neighbor graphs. *Advances in neural information processing systems*, *22*.

[61] Zhang, R., Zhang, S., Lan, Y., & Jiang, J. (2008, March). Network anomaly detection using one class support vector machine. In *Proceedings of the International MultiConference of Engineers and Computer Scientists* (Vol. 1).

[62] Mascaro, S., Nicholso, A. E., & Korb, K. B. (2014). Anomaly detection in vessel tracks using Bayesian networks. *International Journal of Approximate Reasoning*, *55*(1), 84-98.

[63] Valdes, A. D. J., Fong, M. W., & Porras, P. A. (2008). Prioritizing Bayes network alerts. *U.S. Patent No. 7,379,993*. Washington, DC: U.S. Patent and Trademark Office.

[64] Tian, J., Azarian, M. H., & Pecht, M. (2014). Anomaly detection using self-organizing maps-based k-nearest neighbor algorithm. In *PHM society European conference* (Vol. 2, No. 1).

[65] Su, M. Y. (2011). Real-time anomaly detection systems for Denial-of-Service attacks by weighted k-nearest-neighbor classifiers. *Expert Systems with Applications*, *38*(4), 3492-3498.

[66] Hu, J., Ma, F., & Wu, S. (2018). Anomaly identification of foundation uplift pressures of gravity dams based on DTW and LOF. *Structural control and health monitoring*, *25*(5), e2153.

[67] Song, B., & Suh, Y. (2019). Narrative texts-based anomaly detection using accident report documents: The case of chemical

process safety. *Journal of Loss Prevention in the Process Industries, 57*, 47-54.

[68] Marcos Alvarez, A., Yamada, M., Kimura, A., & Iwata, T. (2013, October). Clustering-based anomaly detection in multi-view data. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management* (pp. 1545-1548).

[69] Saeedi Emadi, H., & Mazinani, S. M. (2018). A novel anomaly detection algorithm using DBSCAN and SVM in wireless sensor networks. *Wireless Personal Communications, 98*, 2025-2035.

[70] Pandeeswari, N., & Kumar, G. (2016). Anomaly detection system in cloud environment using fuzzy clustering based ANN. *Mobile Networks and Applications, 21*, 494-505.

[71] Bronte, R., Shahriar, H., & Haddad, H. (2016, June). Information theoretic anomaly detection framework for web application. In *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 2, pp. 394-399). IEEE.

[72] Marchetti, M., Stabili, D., Guido, A., & Colajanni, M. (2016, September). Evaluation of anomaly detection for in-vehicle networks through information-theoretic algorithms. In *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)* (pp. 1-6). IEEE.

[73] Salem, O., Guerassimov, A., Mehaoua, A., Marcus, A., & Furht, B. (2014). Anomaly detection in medical wireless sensor networks using SVM and linear regression models. *International Journal of E-Health and Medical Communications (IJEHMC)*, 5(1), 20-45.

[74] Sakurada, M., & Yairi, T. (2014, December). Anomaly detection using autoencoders with nonlinear dimensionality reduction. In *Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis* (pp. 4-11).

[75] Zong, B., Song, Q., Min, M. R., Cheng, W., Lumezanu, C., Cho, D., & Chen, H. (2018, February). Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *International conference on learning representations*.

[76] Akoglu, L., & Faloutsos, C. (2010, December). Event detection in time series of mobile communication graphs. In *Army science conference* (Vol. 1, p. 141).