IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS. 2020.Doi Number

多年来，异常检测一直用于检测和分析数据中的异常元素。 已经开发了各种技术来检测异常。 然而，最方便的是机器学习，它表现良好但对于大规模未标记数据集仍有局限性。 基于深度强化学习 (DRL) 的技术优于现有的监督或无监督和其他异常检测替代技术。 本研究提出了系统文献综述 (SLR)，它分析了在其应用程序中检测异常的 DRL 模型。 该 SLR 旨在分析异常检测应用程序的 DRL 框架、提出的 DRL 方法，以及它们与替代方法的性能比较。 在这篇综述中，我们确定了 2017 年至 2022 年的 32 篇研究文章，讨论了用于各种异常检测应用的 DRL 技术。 在分析了选定的研究文章之后，本文介绍了在选定的研究文章中发现的 13 种不同的异常检测应用。 我们确定了用于异常检测实验的 50 个不同的数据集，并展示了在选定论文中用于检测异常的 17 个不同的 DRL 模型。 最后，我们分析了这些 DRL 模型的性能并对其进行了审查。 此外，我们观察到使用 DRL 框架检测异常是一个很有前途的研究领域，并表明 DRL 在其他模型缺乏的异常检测方面表现出更好的性能。 因此，我们根据这篇综述为研究人员提供建议和指南。

Deep Reinforcement Learning for Anomaly Detection: A Systematic Review

Kinza Arshad[1], Rao Faizan Ali[1], Amgad Muneer[2,3], Izzatdin Abdul Aziz[2,3], Sheraz Naseer[1], Nabeel Sabir Khan[1], Shakirah Mohd Taib [2,3]

[1]Department of Computer Science, University of Management and Technology, Lahore 54728, Pakistan
[2]Department of Computer and Information Sciences, Universiti Teknologi PETRONAS,Seri Iskandar 32160
[3]Centre for Research in Data Science (CERDAS), Universiti Teknologi PETRONAS,Seri Iskandar 32610

Corresponding author: Amgad Muneer (muneeramgad@gmail.com).

This work was supported by the Centre for Research in Data Science (CERDAS), Under Cost Centre (015LC0-353), Universiti Teknologi PETRONAS.

ABSTRACT Anomaly detection has been used to detect and analyze anomalous elements from data for years. Various techniques have been developed to detect anomalies. However, the most convenient one is Machine learning which is performing well but still has limitations for large-scale unlabeled datasets. Deep Reinforcement Learning (DRL) based techniques outperform the existing supervised or unsupervised and other alternative techniques for anomaly detection. This study presents a Systematic Literature Review (SLR), which analyzes DRL models that detect anomalies in their application. This SLR aims to analyze the DRL frameworks for anomaly detection applications, proposed DRL methods, and their performance comparisons against alternative methods. In this review, we have identified 32 research articles from 2017-2022 that discuss DRL techniques for various anomaly detection applications. After analyzing the selected research articles, this paper presents 13 different applications of anomaly detection found in the selected research articles. We identified 50 different datasets applied in experiments on anomaly detection and demonstrated 17 distinct DRL models used in the selected papers to detect anomalies. Finally, we analyzed the performance of these DRL models and reviewed them. Additionally, we observed that detecting anomalies using DRL frameworks is a promising area of research and showed that DRL had shown better performance for anomaly detection where other models lack. Therefore, we provide researchers with recommendations and guidelines based on this review.

INDEX TERMS Anomaly detection, Deep Reinforcement Learning.

I. INTRODUCTION

Detecting anomalies is a significant problem that is being researched for decades. To identify anomalies for various purposes, a variety of techniques have been proposed and employed. The challenge of detecting patterns in data that do not match to predicted behavior is known as anomaly detection [1], [2]. Anomaly detection is commonly applied in a wide range of different applications. Anomaly detection is also employed in cyber security intrusion detection, network intrusion detection [3]–[5], anomaly detection in videos to detect any unusual activity like road crimes or robberies etc., fault detection, streaming, and hyperspectral imaging, among other applications. The relevance of identifying anomalies in many application areas arises from the possibility of unprotected data, which might include valuable, relevant, and essential data. For example, detecting an anomalous network traffic pattern may reveal an intrusion from a hacked machine [6]. It is also used in medical applications. Another instance is identifying abnormalities in bank or credit card transaction data, which might suggest fraud [7]. Furthermore, identifying an anomaly from an aviation detector may lead to discovering a defect in one or more of the airplane's systems.

Many techniques have been used for anomaly detection. Statistical anomaly detection techniques are some of the oldest algorithms used to detect anomalies [8]. They use a statistical model to calculate and detect unusual patterns from the data. Machine Learning (ML) has been a trendy technique for anomaly detection. It is the most conventional and popular approach to detecting anomalies. ML has been successful to some extent. They include a supervised

model, which uses labeled data, unsupervised, which uses un-labelled data and semi-supervised learning methods, which use a small labeled and large set of unlabeled datasets to detect anomalies. It simply builds models that separate the ordinary and anomalous classes [9]. The agent (ML algorithm) learns the input-output mapping (model) using labeled training data in supervised learning. A supervised learning method generalizes across training cases to predict data labels. Labels are not always correct. In the process sector, the subject matter expert is often an unreliable and noisy sensor measuring a process's present status (temperature, pressure, etc.). The supervised learning agent cannot defeat the subject matter expert since it copies the expert's labeling behavior. The agent's performance limit is called the Bayes error rate and is commonly used unsupervised learning, *e.g.*, similarity-based data separation. Segregating data depending on data set components is one example. Unsupervised learning aims to reduce dimensions, extract features and clustering. Semi-supervised learning combines supervised and unsupervised approaches. Manually labeling data sets is costly in the process industry, but many applications, like defect detection, require them. Semi-supervised learning can be used to learn from labeled data and unlabeled data. Semi-supervised learning cannot outperform the supervisor. Older approaches can just reduce expenses while failing to increase modern capabilities.

Reinforcement learning (RL) is a sub-domain of ML that does not need labeled data. Unlike supervised ML, it uses an intelligent agent to make optimal decisions by maximizing rewards to achieve the goal [10]. RL is similar to dynamic programming. Deep Reinforcement Learning (DRL) combines deep learning and reinforcement learning. DRL incorporates the DL to a solution which helps the agent in RL to make an optimal decision from unstructured data and solve the problem of manual engineering of the state space in RL. DRL algorithms can perform well for huge-scale datasets and are helpful in diverse applications, including anomaly detection, video games, robotics, transportation, NLP, healthcare, computer vision and finance [11].

Anomaly detection is an important application of Deep Reinforcement Learning (DRL). DRL combines the ability of deep learning with the decision-making ability of Reinforcement learning [12]. It solves the critical yet largely unsolved problem of detecting anomalous data. DRL approach actively seeks novel classes of anomalies that lie beyond the scope of the label dataset. It outperforms the other model to detect anomalies in massive volume datasets, which is practically hard to handle in alternative unsupervised problems [13].

The primary objective of this research is to conduct a systematic review that represents a comprehensive study of proposed frameworks of DRL for anomaly detection and its applications. In addition, this review presents DRL models, and their performance compared to alternative models, and suggests DRL models for various anomaly detection applications. This review also represents all anomaly datasets that have been used in the research articles that are selected for review in this SLR.

The remaining part of this paper consists of the following sections: Section 2 includes the related work, Section 3 contain the methodology used to do this research, Section 4 consist of results and discussion, and Section 5 addresses limitation, conclusion and suggested future work.

II. LITERATURE REVIEW

Anomaly detection is a critical topic that has already been researched and implemented in various disciplines. Many anomaly detection systems have been adapted to specific purposes but are much more generic. The following sub-sections address the concept of anomaly detection and DRL with an investigation of the prior works, anomaly detection types, methods, and applications.

A. Anomaly Detection

Anomaly detection is the process of identifying anomalous patterns that do not conform to expected behavior; these anomalous patterns are commonly known as anomalies and outliers [62]. Anomaly detection has been applied to various fields of study, including data breaches, identity theft, networking, manufacturing, video surveillance, and IoT anomaly detection.

Solid knowledge of the nature of anomalies is essential for the development of anomaly detection systems. Anomalies are divided into three classes:

- **Point Anomalies**: A data point-based anomaly is an instance of data that is regarded as an aberration compared to the rest of the data. This sort of anomaly is the simplest and is typically the focus of most of the research on anomaly identification. This category is shown in Figure 1(a), which depicts the discharge capacity data collected from a lithium-ion battery and the anomaly locations.
- **Contextual Anomalies**: A context-based anomaly is an instance of data that is considered anomalous if it is anomalous in a particular context but not in another. Figure 1(b) illustrates a temperature time series that depicts the average monthly temperature for a region. At time *t1* (winter), a temperature of 20°F is typical. However, a temperature of 20°F at time *t2* (summer) may be anomalous.
- **Collective Anomalies**: This category specifies that a group of data instances are out of the ordinary relative to the overall dataset. Figure 1(c) illustrates an ECG output, and the highlighted zone is an anomaly set since the human ECG output should not remain below for an extended period.

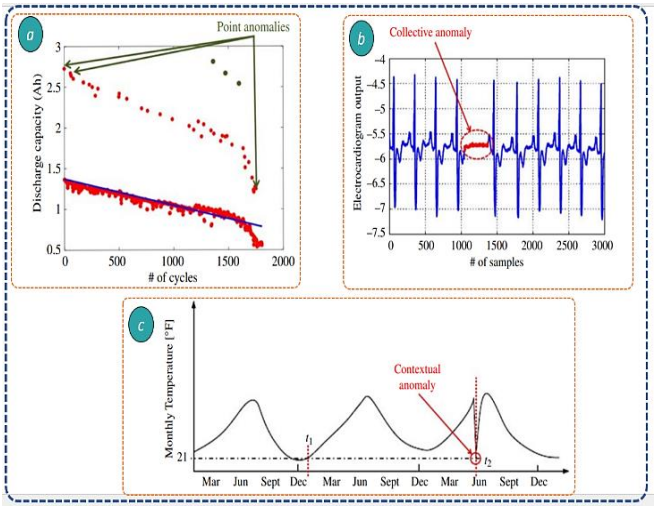

FIGURE 1. Examples of Anomalies Categories [63]

ML-based anomaly detection is becoming more prevalent, and this technique is used to construct a model that differentiates between normal and abnormal classes [59]. Based on the data function, anomalous approaches can be categorized into three types. These are the three categories:

- **Supervised Anomaly Detection**: requires all dataset instances to be labelled "normal" and "anomalous" This method is essentially a type of binary classification task [64].
- **Semi-Supervised Anomaly Detection**: requires only "normal" cases in a dataset to be labelled. In this method, the model will predict only normal occurrences [65].
- **Unsupervised Anomaly Detection**: requires no labelling of cases. In these methods, the model attempts to predict which instances are "normal" and which are "abnormal" [66], [67].
- **Reinforcement Learning:** is a learning model comparable to supervised learning, with the exception that the algorithm is not taught using a dataset. The reinforcement learning paradigm acquires knowledge from external feedback provided by a thinking entity or the environment [68]. Anomaly detection is an important application of deep reinforcement learning. DRL combines the ability of deep learning with the decision-making ability of RL [12]. It solves the critical yet largely unsolved problem of detecting anomalous data. DRL approach actively seeks novel classes of anomalies that lie beyond the scope of the label dataset. It outperforms the other model to detect anomalies in massive volume datasets, which is practically hard to handle in alternative unsupervised problems [13].

The authors in [1] for instance, gave a comprehensive overview of anomaly detection approaches and their applications. A detailed comprehensive review of several machine learning and non-machine learning algorithms, including statistical and spectral detection methods, was conducted. In addition, the review covers a variety of anomaly detection applications and techniques. Cyber intrusion detection, fraud detection, medical anomaly detection, industrial damage detection, image processing detection, textual anomaly detection, and sensor networks are all instances of cyber intrusion detection. However, this anomaly comprehensive survey lack of discussing the recent and powerful algorithms in detecting anomaly and does not focus on DRL. The same researchers also published a survey [8] of discrete patterns of anomaly detection. This researcher gave a thorough and well-organized review of the available research on identifying anomalies in symbolic patterns.

Nevertheless, the limitations of the survey in [8] were involved classical methods, and anomaly detection-based-DRL was not discussed. The authors in [14] also gave an overview of ML and statistical anomaly detection methods. Additionally, the authors compared the benefits and drawbacks of each technique. Thus, DRL-based anomaly detection is still a hot and popular area with praise from academia and industry's massive interests.

Agrawal and Agrawal [7], on the other hand, offered a survey on anomaly detection using data mining approaches. The methods in [7] survey still have limitations for large-scale unlabeled datasets and do not perform well. The author in [9] presented an SLR of anomaly detection using machine techniques. This SLR includes a comprehensive research of supervised, unsupervised, and semi-supervised methods for anomaly detection. They compared all model's performance-wise and made a recommendation for the researcher of this domain. Moreover, they represented all anomalous datasets used the papers they used in their SLR. This SLR also did not focus on the methods and applications of DRL in the anomaly detection domain.

Similarly goes to the systematic literature review conducted by [59], the authors only focused on anomaly detection using ML methods in smart shirts. The SLR in [59] does not include or discuss the DRL methods for anomaly detection; instead, it explores only classical ML methods targeting smart shirt anomaly detection. A different survey was conducted by authors in [60] for dynamically varying environments using RL algorithms. The survey in [60] presents the various categories of RL-based MDP, decision rules and policies and value function. It does not explain the hybridization of DNNs and RL, their benefits, performance, and challenges in the field of anomaly detection.

Numerous studies aimed at identifying anomalies in certain areas and applications like [15], in which the researchers gave an overview of broad clustering-based fraud detection approaches and evaluated them from various viewpoints. The author gave several frameworks and classification techniques for anomaly detection in automated surveillance in [16]. The authors looked at research papers based on the issue, scope, technique, and

strategy. Furthermore, the researcher in [17] presented an overview of the most used anomaly detection approaches in the area of geochemical data analysis, including fractal models, compositional data analysis, and machine learning (ML). However, the author mainly emphasizes on ML algorithms. In [18], on the other hand, looked at the models for log-based anomaly detection. The authors looked at six different anomaly detection algorithms and ranked them. The authors also compared the accuracy and efficiency of 2 primary production log datasets.

Many studies focused on anomalous intrusion detection. In [19], the author, for example, published thorough research on anomalous intrusion detection approaches such as statistics, ML, NNs, and data mining. The author in [20] also looked at intrusion detection, although their emphasis was on ML approaches. They wrote a review of ML approaches for solving intrusion detection issues that were published between 2000 and 2007. Furthermore, the authors examined similar studies based on classifier design types, datasets, and other criteria. In [21], they conducted a comprehensive analysis of anomaly detection and intrusion detection strategies, while in [22], they examined ML and data mining approaches for cyber intrusion detection. They described each approach and discussed the difficulties of using ML and data mining for cyber security. Finally, the researcher in [23] showed how to enhance the effectiveness of detecting abnormalities in network intrusion systems by combining several ML approaches with particle swarm optimization.

Identifying network abnormalities have long been a focus of study [24], [25]. As a result, several surveys have been conducted on the subject. In [26], detailed research on network anomaly detection was published for contrast. They defined the types of assaults that IDS are most likely to experience and then explained and evaluated several anomaly detection approaches' efficiencies. The authors also examined the techniques used by network security. The authors in [6] comprehensively analyzed very well distance-based, density-based, and supervised and unsupervised learning approaches in network anomaly detection. In [27], on the other hand, emphasized on DL approaches, including machine-based DNN, DRNN and ML for network anomaly detection systems. Furthermore, the article provides studies that show how deep learning algorithms may be used to analyze network traffic data.

B. DRL FOR ANOMALY DETECTION IN DIFFERENT DOMAINS

1) VEDIO ANOMALY DETECTION

In surveillance video, the primary action is frequently identified as commonplace, unproblematic behavior. A smart video surveillance system's more critical and challenging task is to locate and detect anomalous actions that are predicted to occur with a lower likelihood than regular activity [32]. Public security was greatly enhanced by smart video surveillance, which used computer vision algorithms to analyze and comprehend the longer video stream. Abnormal activity detection is a crucial component of smart video surveillance because it automatically determines and recognizes anomalies when watching a constantly changing scene and takes action when necessary to deal with emergencies. Due to numerous efforts to flag violent activity in surveillance videos, anomaly detection systems have seen a lot of progress in recent years in helping to resolve security issues [34], [61]. The introduction of deep reinforcement learning showing a significant impact on recognition of area and action from the video.

2) NETWORK INTRUSION DETECTION

One of the most essential security protection techniques used today to keep an eye on computer networks or systems for network-based threats or harmful assaults that might impair system functionality is Network Intrusion Detection Systems [38,40]. A misuse-based network intrusion anomaly-based systems rely on a large database of malicious activity. Furthermore, this system has a slow processing speed and is vulnerable to zero-day attacks. An anomaly-based IDS system uses atypical traffic patterns to spot computer system threats that are concealed. Reinforcement learning (RL) is another machine learning technique that has promise in a variety of applications, including robots and gaming. Recently, several articles have examined the effects of RL in NIDS applications; however, less research has examined the effects of RL on the NIDS problem with unbalanced dataset [43, 49].

3) NETWORK INTRUSION IN IOT

An intrusion detection system (IDS) is consistently regarded as one of the effective tools for protecting the Internet of Things (IoT) network's critical data. IoT devices are more susceptible to security assaults due to the ongoing expansion of interconnected Internet of Things (IoT) devices, which has greatly increased network traffic, complexity, and the constantly shifting Internet environment. To secure the IoT environment, a strong and sophisticated intrusion detection system (IDS) based on cutting-edge machine learning techniques is needed. Reinforcement learning (RL) is one of the best ways to protect the Internet of Things (IoT) from hostile environment learning, incorporating environmental behavior into the learning process. The RL maximizes the overall benefit by engaging the agent with the environment. The data set is created by the agents, who then utilize it to train their models. Using a strategic selection of pertinent features, the RL agent recognizes and categorizes various attacks. Exploring the surroundings and getting positive or negative feedback helps the agent perform better. The agent learns certain attack behaviors after gathering feedback from the environment, at which point it creates a strategy to safeguard IoT against intrusion [53].

4) CYBER ATTACK INTRUSION DETECTION

Cybersecurity is the collection of procedures and techniques created to defend against attacks, unauthorized access, alteration, and damage of computers, networks, programmes, and data. Network security systems and computer (host) security systems make up cyber security systems. Each of these has a firewall, antivirus programme, and intrusion detection system, at the very least (IDS). IDSs assist in finding, determining, and identifying information systems' unlawful use, duplicate, change, and destruction. Attacks from outside the company (external intrusions) and internal intrusions are among the security lapses. In recent few research, DRL has been used to defend systems against network intrusion attacks and solve the problem [51,58].

5) INTRUSION DETECTION IN CLOUD

Cloud computing offers a very adaptable and scalable platform for compensation on-demand access to computing power, data storage, and infrastructure components. Due to its dispersed structure, cloud computing is a prime target for hackers who frequently use new techniques to take advantage of its flaws [35]. There are several innovative assaults and ongoing modifications to attack patterns in the present cloud environment, which makes it more challenging to identify breaches. The current systems require regular updates via retraining with a fresh dataset together with an old dataset to remain viable in such situations, which is not always practicable given the computing cost and resources required. Based on the specific attack types that have been directed at it, a context suggests a certain sort of cloud network. As a result, there is a need for a low-cost IDS that automatically picks up on and adjusts to any changes in attack patterns in the environment while requiring the least amount of human involvement. In this regard, a cloud IDS architecture based on deep reinforcement learning is adaptable and maintains a balance between accuracy and FPR. We now give a succinct history of reinforcement learning (RL) [37].

Although some literature reviews are available, none of the studies has addressed these methods appropriately. However, to the best of our knowledge, this study is among the first SLR on Anomaly detection using Deep reinforcement learning techniques, which is the primary motivation behind this research. Our systematic literature review is different from those described above, as we present extensive research on detecting anomalies using DRL techniques. Our SLR includes:

- Various DRL models for anomaly detection.
- Performance comparison of those with alternative techniques.
- Applications of anomaly detection that are used in the research articles selected for this SLR.
- Represent all anomaly datasets used in the research articles selected for this SLR.
- This SLR covers research articles from 2017-2022.

III. METHODOLOGY

This research follows the Kitchenham and Charters methodology [28] to conduct this Systematic Literature Review. Planning, conducting, and reporting the research are all process parts. Each level has several stages. The planning step is broken down into six sections. The first step is to come up with research questions that are relevant to the review's goals. After determining the appropriate search keywords, the second stage is to devise a search strategy for gathering research articles on the issue that answers the research questions. The research selection processes, which comprise exclusion and inclusion criteria, are identified in the third step. In the fourth stage, there is laying up an extraction approach to address the previously stated research topics. Finally, the data must be synthesized in the fifth stage. The following subsections illustrate how we implemented the review procedure.

A. RESEARCH QUESTION

In this SLR, we aim to present a comprehensive study of DRL models for anomaly detection, which includes an examination of DRL models and their performance from 2017-2022. Research questions raised for this purpose are:

1. **RQ1: What anomaly detection applications are discussed using DRL techniques?**

RQ1 aims to discuss the application of anomaly detection that is used in this SLR using DRL.

2. **RQ2: What anomalous datasets are used for anomaly detection using DRL techniques?**

RQ2 aims to present various anomalous datasets that are used in the papers selected for this SLR.

3. **RQ3: What algorithms of DRL are used to detect anomalies?**

The purpose of RQ3 is to mention precisely which DRL algorithm is proposed for detecting anomalies in this research.

4. **RQ4: What is the performance of the DRL model compared with the alternative method?**

RQ4 focuses on the model's performance, which includes estimation, and prediction accuracy to detect anomalies using DRL and their performance with other alternative models.

B. SEARCH STRATEGY

The search scope is defined and restricted to computer science, social science, information systems, and information security (behavioral aspect). This research focuses on automated and manual search techniques to get as many research papers as feasible to meet the study's goals. As previously mentioned, a manual search procedure was also carried out using search engines and reference lists of similar publications. To conduct this SLR, the procedure that we followed is listed below:

1. First, we identified the search terms by analyzing RQs.

2. Then we defined new relative terms like synonyms, *i.e.* intrusion.
3. We used AND and OR operators to search for the required topic.
4. The keywords we searched for this SLR are related to Deep Reinforcement learning AND anomaly detection.

We used the following libraries that we used in this SLR to collect research papers which include conference and journal papers:

- IEEE Explorer
- Springer
- Elsevier
- ACM Digital Library

1) INCLUSION AND EXCLUSION CRITERIA

Inclusion criteria to select a paper for this SLR are given below:

- Only include English language papers.
- Only conference papers or journal papers can be included.
- Only include a paper on anomaly detection or its application.
- Only include papers which use the DRL technique to detect anomalies.
- Only consider papers published from 2017 to 2022.

Exclusion criteria to reject a paper for this SLR are given below:

- Papers with no clear publication information are excluded.
- Papers related to DRL but do not mention anomaly detection are excluded.
- Papers related to anomaly detection but do not discuss DRL are excluded.
- Review papers are excluded.

C. STUDY SELECTION

To conduct this SLR, we collected 46 papers based on search terms discussed earlier. After observing them using selection criteria, we discarded 3 review papers and 6 unrelated papers which do not define the inclusion criteria and 5 duplicate articles. After this filtration, we finally selected 32 papers to observe and review for this SLR. These filtration steps to select paper are given below:

1. Remove duplicate research papers collected from different digital libraries.
2. Apply the inclusion and exclusion criteria discussed above.
3. Remove review papers.
4. Apply quality assessment rules to include the best-selected paper for this SLR.

Search related articles from references of selected papers and repeat the steps above. Figure 2 shows the study selection criteria utilized in this SLR, and Figure 3 illustrates the identified 32 research articles written from 2017-2022 that discuss DRL techniques for various applications of anomaly detection.

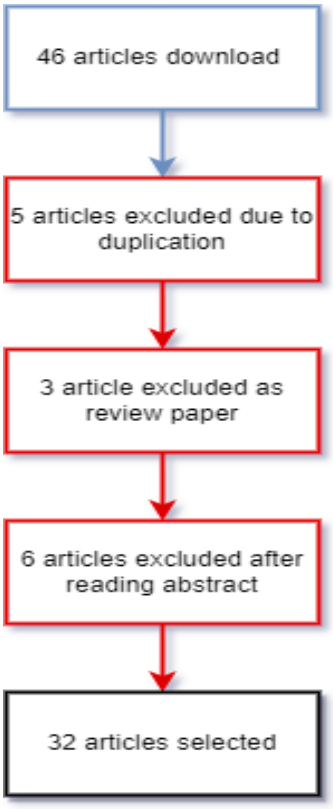

FIGURE 2. Study selection

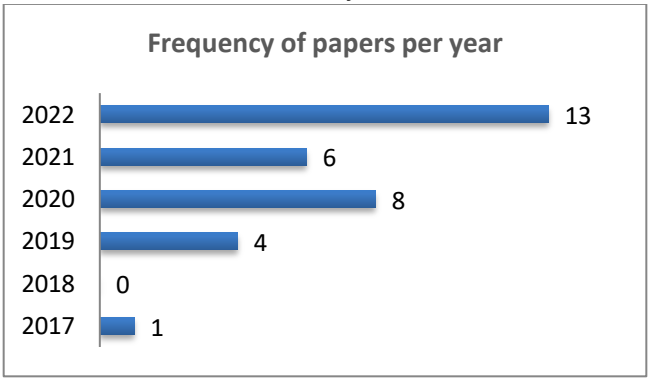

FIGURE 3. Frequency of the papers per year

D. DATA EXTRACTION STRATEGY

In this SLR, we aim to present the various DRL techniques for anomaly detection and specify their application. We also aim to present the different anomalous datasets they have used for the anomaly dataset. For this purpose, the information we extracted from the selected papers includes the title of the research paper, year of publication, type of anomaly detection application, DRL models they proposed to detect an anomaly, dataset they used and performance of the DRL model. All of these are included in RQs.

E. SYNTHESIS OF EXTRACTED DATA

In completing this SLR, we employed several techniques to collect knowledge to address the RQs by synthesizing the

information from the chosen publications. To answer the RQ1, we identified all anomaly detection applications from selected papers and represented them in a tabular form mentioning paper ID. To answer RQ2, we extracted all the datasets from all selected papers and represented them in a tabular form mentioning paper ID. To address RQ3, we mention the DRL models used in each selected paper in a table. To address RQ4, we made a performance comparison of each DRL model discussed in selected models to its alternative in Table I.

IV. RESULTS AND DISCUSSION

This section provides an overview of the review's chosen papers. In the following parts, the outcomes of each study topic are discussed in depth. The results of each research question are detailed in the following four sections. A total of 32 papers were chosen for this SLR which implement and discuss deep reinforcement learning and anomaly detection application. These research articles were published from 2017 to 2022, which is relatively recent. The list of chosen papers for this SLR is given in Table I.

TABLE I. Selected research papers

ID	TITLE	YEAR	SOURCE	REFS. NO.
P1	"Deep Reinforcement Learning for Unknown Anomaly Detection"	2020	IEEE	[13]
P2	"Meta-AAD: Active Anomaly Detection with Deep Reinforcement Learning"	2020	IEEE	[29]
P3	"Deep Actor-Critic Reinforcement Learning for Anomaly Detection"	2019	IEEE	[30]
P4	"Learning of Binocular Fixations using Anomaly Detection with Deep Reinforcement Learning"	2017	IEEE	[31]
P5	"Deep Reinforcement Learning for Real-world Anomaly Detection in Surveillance Videos"	2019	IEEE	[32]
P6	"Towards Adaptive Anomaly Detection in Buildings with Deep Reinforcement Learning"	2019	ACM DL	[33]
P7	"Intelligent video anomaly detection and classification using faster RCNN with deep reinforcement learning model"	2021	Elsevier	[34]
P8	"Robust Adaptive Cloud Intrusion Detection System Using Advanced Deep Reinforcement Learning"	2020	Springer	[35]
P9	"Application of deep reinforcement learning to intrusion detection for supervised problems"	2020	Elsevier	[36]
P10	"Deep Reinforcement Learning based Intrusion Detection System for Cloud Infrastructure"	2020	IEEE	[37]
P11	"A Deep Reinforcement Learning Approach for Anomaly Network Intrusion Detection System"	2020	IEEE	[38]
P12	"A Deep Reinforcement Learning Based Intrusion Detection System (DRL-IDS) for Securing Wireless Sensor Networks and Internet of Things"	2020	Springer	[39]
P13	"Designing online network intrusion detection using deep auto-encoder Q-learning"	2019	Elsevier	[40]
P14	"Abnormal flow detection in industrial control network based on deep reinforcement learning."	2021	Elsevier	[12]
P15	"Network Intrusion Detection Systems Using Adversarial Reinforcement Learning with Deep Q-network"	2020	IEEE	[41]
P16	"A Dynamic Deep Reinforcement Learning-Bayesian Framework for Anomaly Detection"	2021	IEEE	[42]
P17	"Deep Q-Learning based Reinforcement Learning Approach for Network Intrusion Detection"	2021	IEEE	[43]
P18	"Intrusion Detection Framework Using an Improved Deep Reinforcement Learning Technique for IoT Network"	2021	Springer	[44]
P19	"Deep-Reinforcement –Learning-Based Intrusion Detection in Aerial Computing Networks"	2021	IEEE	[45]

P20	"DeepAir: Deep Reinforcement Learning for Adaptive Intrusion Response in Software-Defined Networks"	2022	IEEE	[46]
P21	"A Deep Reinforcement Learning based Intrusion Detection Strategy for Smart Vehicular Networks"	2022	IEEE	[47]
P22	"Intrusion Detection System for Industrial Internet of Things Based on Deep Reinforcement Learning"	2022	Hindawi	[48]
P23	"Deep Q-Learning Based Reinforcement Learning Approach for Network Intrusion Detection"	2022	MDPI	[49]
P24	"Temporal Detection of Anomalies via Actor-Critic Based Controlled Sensing"	2022	IEEE	[50]
P25	"Low latency cyberattack detection in smart grids with deep reinforcement learning"	2022	Elsevier	[51]
P26	"Deep-attack over the deep reinforcement learning"	2022	Elsevier	[52]
P27	"Intrusion Detection Framework Using an Improved Deep Reinforcement Learning Technique for IoT Network"	2022	Springer	[53]
P28	Deep Reinforcement Learning based Intrusion Detection System with Feature Selections Method and Optimal Hyper-parameter in IoT Environment"	2022	IEEE	[54]
P29	"Controlled Sensing and Anomaly Detection Via Soft Actor-Critic Reinforcement Learning"	2022	IEEE	[55]
P30	"Double Deep Q-Learning With Prioritized Experience Replay for Anomaly Detection in Smart Environments"	2022	IEEE	[56]
P31	"Deep Q-Learning Based Reinforcement Learning Approach for Network Intrusion Detection"	2022	MDPI	[57]
P32	"A Hidden Attack Sequences Detection Method Based on Dynamic Reward Deep Deterministic Policy Gradient"	2022	Hindawi	[58]

A. ANOMALY DETECTION APPLICATION (RQ1)

In this section, we address Research Question 1 (QR1), which discusses anomaly detection and its applications that are implemented using DRL techniques. Anomaly detection may be applied in a wide range of applications. In this research, we found 13 different applications in the anomaly detection-based-DRL publications gathered from the literature. Table II lists these applications and mentions the paper discussing them.

As shown in Table II, our selected articles discuss general anomaly detection, network anomaly detection, intrusion detection, network intrusion detection, cloud intrusion detection, video anomaly detection, building anomaly detection, wireless network security, and the internet of things (IoT). In addition, the table also mentions the frequency of each application discussed in the selected papers and the paper number from Table I. The results show that DRL techniques work well with applications listed in the table above.

Now, if we can observe from Figure 3, which show the percentage of each anomaly detection application from the selected papers, general anomaly detection and application related to intrusion detection, which includes network and cloud intrusion detection, are the most popular applications which have been used for detection using deep reinforcement learning techniques. DRL outperforms other popular techniques like ML and another statistical model for the anomaly detection application, which requires extensive unlabeled data or signal data like in network, wireless signals or cloud intrusion detection. DRL is also popular and performs well for video anomaly detection because the video dataset is high dimensional and contains raw and unlabeled anomalies, which has been a problem for other models.

TABLE 2. Anomaly detection application among selected research articles.

No.	Application	Frequency	Paper ID
1	Anomaly detection	6	P1, P2, P4, P16, P29, P32
2	Network anomaly detection	2	P3, P20
3	Intrusion detection	6	P12, P9, P18, P19, P21, P22
4	Network intrusion detection	6	P11, P13, P15, P17, P23, P33
5	Cloud intrusion detection	2	P8, P10
6	Video anomaly detection	2	P5, P7
7	Building anomaly detection	1	P6
8	IoT network intrusion detection	3	P12, P27, P28
9	Wireless network security	1	P12
10	Industrial network control	1	P14
11	Cyber-attack detection	2	P25, P32
12	Temporal anomaly detection	1	P24
13	Adversarial attack detection	1	P26

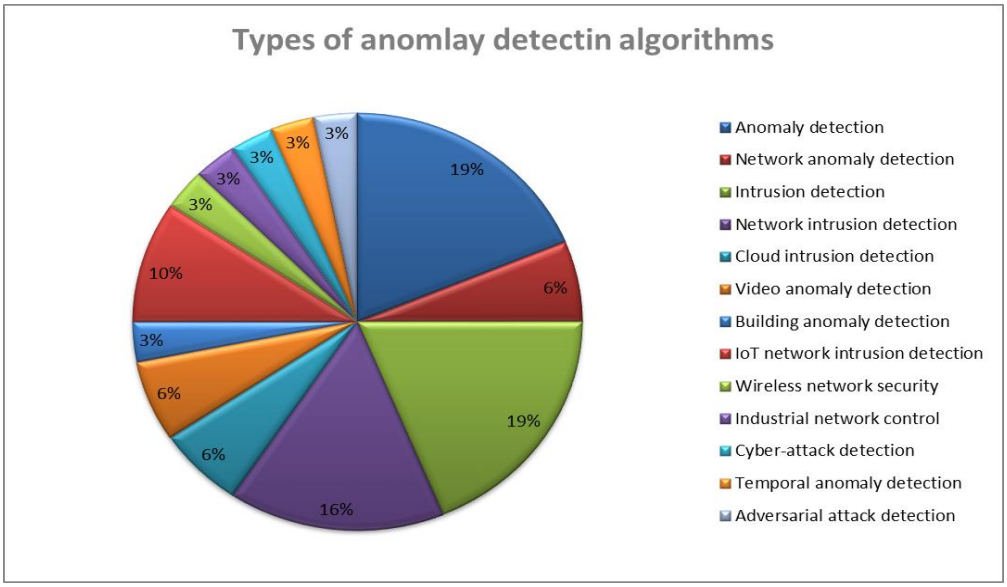

FIGURE 4. Percentage of anomaly detection types in selected papers

B. ANOMALY DATASETS (RQ2)

This section addresses RQ2, which aims to represent all the datasets used for anomaly detection using DRL. Various datasets exist depending on which application of anomaly detection you are dealing with. We have presented 48 different datasets utilized by each selected research paper for this SLR, as given in Table III.

Table III shows that the authors in P1 have used four databases for anomaly detection named NB15, Thyroid, HAR, and Cover type. These databases include 12 different anomalous datasets. In P2, the author has used 24 different datasets used for anomaly detection. Datasets used in P1 and P2 can be used for general anomaly detection models. P3, P11, P12, P13, P14 and P15 have used different network anomaly datasets, which can be used for models built for other network anomaly detection. ISOT-CID, NSL-KDD, AWID, and UNSW-NB15 are the Intrusion detection datasets in P8, P9, P10, P11, P15, and P17 used in ML to detect network intrusions or attacks. P4 used Gazebo's hand-designed objects dataset to detect the position and shape of objects in robotics using anomaly detection. In P5 and P7, they used video datasets named UCF and UCSD, respectively. P3 used Connected and automated vehicles (CAV) sensor data to detect anomalies using the DRL model. MedbIoT is a dataset containing traces of the internet of things (IoT) used in P18 and P19 used aerial computing network data in their research for anomaly detection using DRL. In UCF-anomaly-detection-dataset, it is about 1900 untrimmed and 128 hours long real-world surveillance videos containing 13 cases of real video anomalies. About UCSD, it is an anomaly detection video dataset that was acquired with a camera mounted on walkways used to detect anomalous pedestrian motion patterns. In P6, it is a building-specific anomaly detection dataset used to detect anomalies for building and checking the performance of the parameters from all sensors.

TABLE 3. **Anomaly detection datasets utilized in selected research articles**

Paper ID	Databases	Datasets for Anomaly detection	Frequency
	NB15	Analysis	
		DoS	
		Exploits	
		Fuzzers	
		Generic	
P1		Recon	12
	Thyroid	Hypothyroi	
		Subnormal	
	HAR	Downstairs	
		Upstairs	
	Covertype	Cottonwood	
		Douglas-fir	
		Annthyroid	
		Arrhythmia	
		Breastw	
		Cardio	
		Glass	
		Ionosphere	
		Letter	
P2		Lympho	
		Mammography	24
		Mnist	
		Musk	
		Optdigits	
		Pendigits	
		Pima	
		Satellite	
		Satimage	
		Shuttle	
		Speech	
		Thyroid	
		Vertebral	
		Vowels	
		Wbc	
		Wine	
		Yeast	
P3		Network sensor data	1
P4		Gazebo hand-designed objects	1
P5		UCF-Anomaly-Detection	1
P6		Building Signal data	1
P7		UCSD anomaly	1
P8		ISOT-CID	2
		NSL-KDD	
P9		NSL-KDD	2
		AWID	
P10		UNSW-NB15	1

P11	NSL-KDD UNSW-NB15	2
P12	IDS	1
P13	Online Network data	1
P14	Industrial Network data	1
P15	NSL-KDD	1
P16	CAV sensor data	1
P17	NSL-KDD	1
P18	MedBIoT	1
P19	Aerial computing Network data	1
P20	Sotfware denied network data	1
P21	Vehicular network data	1
P22	Industrial IoT network attack data	1
P23	NSL-KDD	1
P24	Temporal anomaly data	1
P25	Smart grids data	1
P26	Adversarial attack data	1
P27	MedBIoT dataset	1
P28	IoT network data	1
P29	IoT network data	1
P30	Sensory data for occupancy detection Local position data of person for fall detection	2
P31	NSL-KDD	1
P32	Network data traffic	1

C. TYPES OF DEEP REINFORCEMENT LEARNING TECHNIQUES (RQ3)

This section addresses the RQ3 in which we aim to specify the DRL algorithms that have been used to detect anomalies utilized in the selected paper which is one of the primary goals of this review. Table IV represents 17 Deep reinforcement learning algorithms used for anomaly detection from 2017 to 2022, along with their application. DRL models combine artificial neural networks with reinforcement learning to help the agent learn to achieve the goal. Deep Q learning, Actor critic, deep policy gradients, and neural networks with RL are popular algorithms used for different anomaly detection in the selected papers are explained in the following subsections:

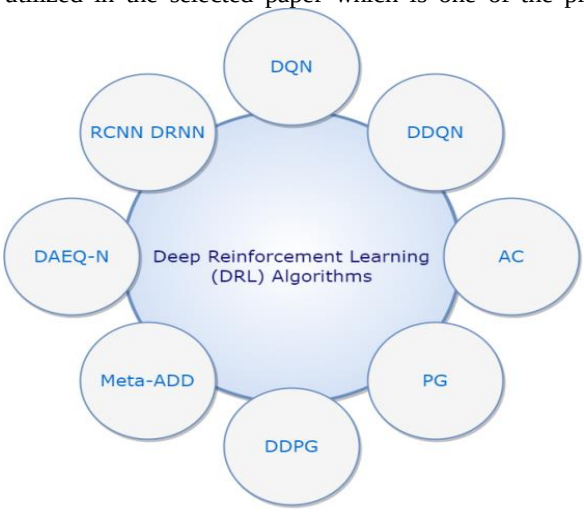

FIGURE 5. Deep Reinforcement Learning Algorithms

1) DEEP Q-NETWORK (DQN) AND DOUBLE DEEP Q-NETWORK (DDQN)

To make reinforcement learning effective in an extensive features and complex situations like video games and automation, DQN is an RL algorithm that combines Q-Learning with DNNs. DQN, however, has several drawbacks that DDQN resolves. When attempting to approximate the state-action value function, it corrects for the DQN algorithm's sporadic values, leading to too optimistic value estimations, because the max operator in both Q-learning (DQL) and DDQN picks and analyzes an action using the same values. By breaking down the target's optimum operation into action selection and action assessment, Double Deep Q-learning aims to reduce overestimation. Double DQN varies from DQN solely during the Q-value update phase.

propensity to exaggerate some values. Therefore, provided the prediction error is maintained to a low, the DQN can be trained. Despite being efficient, the Deep Q Learning method is known to have serious problems, such as overestimating action values in some circumstances. Researchers developed an enhanced technique to address these issues, known as Double Deep Q-learning. It is possible to choose exaggerated

2) ACTOR-CRITIC (AC)

Adapting to a new one, this is a simple and compact framework for deep RL. The actor-critic technique optimizes deep neural network integrators via concurrent gradient descent. Depending on concurrent versions of four common RL algorithms, the study was carried out. The findings demonstrate that concurrent actor-learners stabilize learning and enable all four techniques to effectively train the neural net regulators. The best technique, an asynchronous actor-critic variation, exceeds the most significant algorithms currently available, according to the results. According to research, a concurrent actor-critic also works well on a wide range of persistent motor control issues.

3) POLICY GRADIENT (PG)

The foundation of policy gradient is the training of a policy function, which specifies the course of action to be followed for each potential state. Except the last layer, which uses softmax activation to create a probabilistic model for the action, a basic NN with a few layers and ReLU activation for all layers approximates the policy function. The technique shown employs generalized trajectories that consist of a list of pairs generated by a state and the ground-truth label that goes with it. A small batch of n trajectories includes this generic trajectory. The algorithm's training iterations process every mini-trajectory, batch's and for each iteration, a new mini-batch is created as a result of the process. To use the states and the policy equation, the algorithm first predicts the actions. All the states in a trajectory are subject to action prediction, which results in a list of anticipated actions. The probability distribution of the actions specified by the policy function was sampled to produce these projected actions. The phrase "Prob. Distribution Sampler" is used to describe this.

4) DEEP DETERMINISTIC POLICY GRADIENT (DDPG)

Deep-RL algorithms that are actor-critical, off-policy, and sample-efficient are DDPG. With deterministic policy and off-policy updating utilizing a replay buffer, DDPG is a mix of DQN and QAC. It employs deterministic policy as a rough action space Q-value maximizes. It uses target networks, a postponed update, and Gaussian noise for stochastic actions in discovery. A few weaknesses and instability in DDPG can be attributed partly to an overestimation bias in critic updates. Because of its sensitivity to hyper-parameter settings, it is well known to be challenging to tune. These problems can be solved with the use of well-tailored code baselines that include many cutting-edge methods.

5) META POLICY ACTIVE LEARNING (META-ADD)

Deep Reinforcement Learning is used in Active Anomaly Detection with Meta-Policy (Meta-AAD), which is an active anomaly detection method. Meta-AAD may be a universal framework for active anomaly detection since it may intrinsically optimize short-term and long-term incentives. It is a brand-new methodology that develops a query decision metapolicy. In particular, Meta-AAD makes use of deep reinforcement learning to train the meta-policy to choose the best example to particularly optimize the quantity of anomalies found during the querying procedure. Since a learned meta-policy may be applied immediately to any fresh datasets without additional adjustment, Meta-AAD is simple to implement. It can acquire a meta-policy that explicitly maximizes the quantity of anomalies found. More precisely, we model active anomaly detection as a Markov decision process and use deep reinforcement learning to train the meta-policy to choose the best example in each loop.

6) DEEP AUTOENCODER Q-NETWORK (DAEQ-N)

This model framework is built on an unorthodox approach to experience replay that is comparable to recently published ground-breaking research. The incentives in our suggested model are determined by adding up all the discrepancies between encoding and decoding. We employ an auto-encoder to our advantage for this. Because tiny changes in the weights can generate bigger changes in the state distribution, the average total reward in RL tends to fluctuate dramatically. A typical deep neural network has oscillating average reward graphs. However, throughout training, we try to develop the average total reward. Depending on the auto-encoder, we have a very high probability of making consistent, steady improvements.

7) RCNN DRNN

Deep reinforcement learning (DRL) uses deep neural networks to achieve specified objectives while aiming to train an autonomous agent to interact with a given environment (DNN). Recurrent neural network (RNN) based DRL has proven to perform better than other approaches because RNNs are more adept at capturing the time dynamics of the environment and delivering the right agent responses. Besides their exceptional performance, RNNs' internal environmental comprehension and long-term memory are also little understood. For deep learning professionals, it is crucial to reveal these specifics in order to comprehend and enhance DRLs. However, doing so is problematic since these models contain intricate data transformations.

TABLE 4. Deep reinforcement learning techniques from the selected articles

Paper ID	Proposed DRL models	Anomaly Detection Application
P1	Deep Q-learning with Partially Labeled Anomalies (DPLAN)	General Anomaly detection
P2	Active Anomaly Detection with Meta-Policy (Meta-ADD)	General Anomaly detection
P3	Actor-critic (AC)	Network anomaly detection
P4	Deep Reinforcement Neural Network (DRNN)	General Anomaly detection
P5	Deep Q Learning Network (DQN)	Video anomaly detection
P6	Deep Deterministic Policy Gradient (DDPG)	Building anomaly detection
P7	RCNN with deep reinforcement learning model (RNN DRL) Deep Q learning Network (DQN)	Video anomaly detection
P8	Double Deep Q-Network (DDQN) DRL Adaptive IDS	Cloud intrusion detection
P9	Deep Q learning Network (DQN) Double Deep Q-Network (DDQN) Policy Gradient (PG) Actor-Critic (AC)	Intrusion detection
P10	DRL- adaptive cloud IDS	Cloud intrusion detection
P11	DRL-NIDS	Network intrusion detection
P12	DRL-IDS	Intrusion detection, Internet of Things (IoT), Wireless network security
P13	Deep auto-encoder Q-network (DAEQ-N)	Network intrusion detection
P14	Deep Reinforcement Neural Network (DRNN)	Industrial network control
P15	Adversarial/Multi-Agent Reinforcement Deep Q-Learning Network (AE-DQN)	Network intrusion detection
P16	POMDP model	Anomaly detection
P17	Deep Q learning (DQL) model	Network intrusion detection
P18	DRL-IDS	IoT Network intrusion detection
P19	DRL-IDS	Aerial computing intrusion detection
P20	Double deep Q network (DDQN)	Network anomaly detection
P21	Deep Q learning (DQN) model	Intrusion detection
P22	DRL-IDS	Intrusion detection
P23	Deep Q Learning (DQL)	Network intrusion detection
P24	Actor-Critic	Temporal anomaly detection
P25	Deep Q Network (DQN)	Cyber-attack detection
P26	DRL	Adversarial attack detection
P27	DRL-IDS	IoT network intrusion detection
P28	DRL-IDS	IoT network intrusion detection
P29	Actor critic	Anomaly detection
P30	Double Deep Q-Learning (DDQL) network	Anomaly detection
P31	Deep Q-Learning (DQL) network	Network intrusion detection
P32	Deep Deterministic Policy Gradient (DDPG)	Cyber-attack detection

D. PERFORMANCE ANALYSIS FOR DEEP REINFORCEMENT LEARNING ALGORITHMS (RQ4)

In this section, we address the RQ4, which is concerned with the performance of the DRL model and its comparison with other alternative models utilized in the papers we selected for this review. Table V shows all the DRL models, specifying the application of anomaly detection, mentioning the dataset, and showing the models' performance with their accuracy. Some papers mention accuracy of the model. Others evaluated the models based on comparison with state-of-the-art. As we can see from the table, the NSL-KDD dataset is used by papers P8, P9, P11 and P15. In P11, Deep

reinforcement learning for network intrusion detection system DRL-NIDS proved to be the better DRL algorithm with 91.4% accuracy over other models to detect network anomalies from NSL-KDD dataset. Both P10 and P11 used the UNSW-NB15 dataset, but in P11, DRL-NIDS performed better with 91.8% accuracy. Concerning the application type of anomaly detection, P5 and P6 performed video anomaly detection on real-time large video datasets. For comparison, IVADC-FDRL and Deep Q learning Network (DQN) model in P7 performed better with up to 98% accuracy over another model for video anomaly detection. Each proposed model with DRL reviewed in selected papers from 2017-2021 has outperformed other competitors or alternative models. DRL algorithms have performed better for anomaly detection, whereas other techniques lack.

According to our SLR, researchers utilize a variety of evaluation metrics to assess the performance of various RL models. Table 6 provides a collection of commonly employed evaluation metrics.

TABLE 5. Performance analysis of DRL algorithms

Paper ID	Anomaly detection application	Model	Dataset	Performance Comparison
P1	General Anomaly detection	DPLAN	NB15 Thyroid HAR Cover type	23%-98% relative AUC-PR improvement
P2	General Anomaly detection	Meta-ADD	24 different anomaly datasets	Outperform alternative unsupervised, SSDO, AAD and FIF models.
P3	Network anomaly detection	Actor-critic	Network sensor data	Outperform Chernoff test
P4	General Anomaly detection	Deep Reinforcement Neural Network (DRNN)	Gazebo hand-designed objects	
P5	Video anomaly detection	Deep Q Learning Network (DQN)	UCF-Anomaly-Detection	78.20% accuracy, outperforming alternative models
P6	Building anomaly detection	Deep Deterministic Policy Gradient (DDPG)	Building Signal data	Up to 3x better than alternative models
P7	Video anomaly detection	IVADC-FDRL Deep Q Learning Network (DQN)	UCSD anomaly	98.50% accuracy 94.80% accuracy over other models
P8	Cloud intrusion detection	Double Deep Q-Network (DDQN) DRL Adaptive IDS	ISOT-CID NSL-KDD	Outperform state-of-the-art
P9	Intrusion detection	Deep Q Learning Network (DQN) Double Deep Q-Network (DDQN) Policy Gradient (PG) Actor-Critic (AC)	NSL-KDD AWID	87.87% accuracy 89.78% accuracy 78.73% accuracy 80.78% accuracy Outperform alternative models
P10	Cloud intrusion detection	DRL- adaptive cloud IDS	UNSW-NB15	83.30% Outperform all existing works
P11	Network intrusion detection	DRL-NIDS	NSL-KDD UNSW-NB15 Real-time campus network traffic data	91.4% accuracy 91.8% accuracy 97.95% accuracy

P12	Intrusion detection, Internet of Things (IoT), Wireless network security	DRL-IDS	IDS	Outperform state-of-the-art
P13	Network intrusion detection	Deep auto-encoder Q-network (DAEQ-N)	Online Network data	Outperform state-of-the-art (DNN)
P14	Industrial network control	Deep Reinforcement Neural Network (DRNN)	Industrial Network data	98.06% accuracy
P15	Network intrusion detection	Adversarial/Multi-Agent Reinforcement Deep Q-Learning Network (AE-DQN)	NSL-KDD	80% accuracy
P16	Anomaly detection	POMDP model	CAV sensor data	Outperform state-of-the-art
P17	Network intrusion detection	Deep Q learning (DQL) model	NSL-KDD	Outperform state-of-the-art ML approaches
P18	IoT Network Intrusion detection	DRL-IDS	MedBIoT	96.99% accuracy
P19	Aerial computing Intrusion detection	DRL-IDS	Aerial computing Network data	Outperform state-of-the-art
P20	Network anomaly detection	Double deep Q network (DDQN)	Sotfware denied network data	Outperform existing solutions i.e. GATE (by 75%) and GTAC-IRS (by 80%), respectively
P21	Intrusion detection	Deep Q learning (DQN) model	Vehicular network data	Outperform state-of-the-art
P22	Intrusion detection	DRL-IDS	Industrial IoT network attack data	90% accuracy
P23	Network intrusion detection	Deep Q Learning (DQL)	NSL-KDD	Accuracy of 90%, outperforming existing ML algorithms
P24	Temporal anomaly detection	Actor Critic	Temporal anomaly data	Outperform state-of-the-art
P25	Cyber-attack detection	Deep Q Network (DQN)	Smart grids data	Outperform state-of-the-art
P26	Adversarial attack detection	DRL	Adversarial attack data	Outperform state-of-the-art
P27	IoT network intrusion detection	DRL-IDS	MedBIoT dataset	96.99% accuracy
P28	IoT network intrusion detection	DRL-IDS	IoT network data	Outperform state-of-the-art
P29	Anomaly detection	Actor critic	IoT network data	Outperform state-of-the-art
P30	Anomaly detection	Double Deep Q-Learning (DDQL) network	Sensory data for occupancy detection	92.6% accuracy

			Local position data of person for fall detection	
P31	Network intrusion detection	Deep Q-Learning (DQL) network	NSL-KDD	94% accuracy
P32	Cyber-attack detection	Deep Deterministic Policy Gradient (DDPG)	Network data traffic	97.64% accuracy

TABLE 6: Commonly Employed RL Performance Metrics

No	Evaluation Metric	Description
1	Dispersion across Time (DT)	Instead of depending on longer-term trends, Dispersion over Time (DT) is calculated by isolating higher-frequency variability. Detrending is utilized to prevent the measures from being impacted by a positive trend. Detrending is a statistical approach that includes removing the effects of accumulating data sets from a trend to simply display the absolute changes in values and find possible repeating patterns. The final DT metric is the interquartile range (IQR) within a sliding window along the detrended training curve.
2	Risk across Runs (RR)	CVaR is applied to the cumulative results of all training runs. This measure provides information about the performance of the poorest runs.
3	Dispersion across Runs (DR)	It is determined by calculating training runs' variance or standard deviation at a series of evaluation points. First, the training data are low-pass filtered to remove high-frequency fluctuation between runs. IQR replaces the variance or standard deviation
4	Long-term Risk across Time (LRT)	This indicator aids in tracking performance relative to the highest peak to date and can be used to identify significant dips that occur over extended periods of time (drawdown). CVaR is applied to the drawdown time for this metric.
5	Short-term Risk across Time (SRT)	This measure indicates the worst-case projected decline in performance from one evaluation point to the next throughout training. To accomplish this, CVaR is employed for the performance variations between evaluation points. The SRT is determined as follows: a) Calculate the differences between two training run time points. b) Normalize the differences based on the distance between time points to ensure evaluation frequency invariance. c) Determine the α-quantile and the distribution of these differences. d) Compute the distribution's expected value below the α-quantile.
6	Risk across Fixed-Policy Rollouts (RF)	This measure is comparable to the Dispersion among Fixed-Policy Rollouts metric, with the exception that this metric does not include the total number of rollouts CVaR is applied to rollout results.
7	Dispersion across Fixed-Policy Rollouts (DF)	To construct this measure, the IQR is computed using the rollout performance. This aids in evaluating a fixed policy to determine the performance variation when the same policy is implemented several times.

V. CONCLUSION

A. THEORETICAL IMPLICATIONS

First, this study is among the first systematic literature review on Anomaly detection using Deep reinforcement learning techniques. Although some literature reviews are available, none of the studies has addressed these methods appropriately. Our systematic literature review is totally different from other anomaly detection studies, as we present extensive research on detecting anomalies using DRL techniques. Our SLR provides a review of various DRL models for anomaly detection, performance comparison of those with alternative techniques, applications of anomaly detections that are used in this research articles selected for this SLR, and we represented all anomaly detection datasets that are used in this research articles which are selected for this SLR covered from 2017 to 2022.

In recent years, Deep reinforcement learning (DRL) has outperformed Deep learning and Machine learning in many ways. DRL models combine artificial neural networks with reinforcement learning to help the agent learn to achieve the goal. As far as our topic, Anomaly detection, is concerned, the main techniques used in DRL are Deep Q learning, policy gradient, deep auto-encoder Q learning, double deep Q learning, policy gradient, and actor-critic models. These models of DRL have outperformed the other deep/machine learning techniques for detecting an anomaly in various applications. This research shows that a deep Q network can be used if the researcher is dealing with intrusion or video anomaly data. Deep policy gradient techniques have been used for building anomaly detection. The actor-critic has been used for intrusion detection.

If we talk about anomaly datasets, there are various datasets that DRL has covered, like Network anomaly, industrial network anomaly, wireless network anomaly, network intrusion, cloud intrusion, general anomaly, video anomaly, building anomaly, signal anomaly and unknown anomaly detection. We have shown that one DRL technique uses a different application and dataset of anomaly yet outperforms other models. Therefore, DRL proves to perform best for all applications of anomaly detection.

B. LIMITATIONS

This research is about anomaly detection from Deep Reinforcement Learning. There exists a very limited number of research articles about this topic because the first paper we could find about it was in 2017, so our Systematic literature review starts from 2017 to 2022. This SLR is also limited to journal and conference papers that have used only DRL framework for anomaly detection. Our scope is limited; we did not include several other anomaly detection methods to meet the selection criteria requirement. We believe this systematic literature review would have been improved by increasing the scope and sources.

C. FUTURE AVENUE FOR RESEARCHERS

This review presents DRL models for anomaly detection; as we know, this is the SLR of only 32 papers published from 2017 to 2022. Therefore, we recommend that other researchers to conduct more research on deep reinforcement learning for anomaly detection to gain evidence about the performance of DRL for anomaly detection. RL is an emerging field and has many scopes. Moreover, we observed that there are limited anomaly detection applications that have been used for DRL. Possible future avenues for other researchers to explore DRL techniques for other anomaly detection applications not listed in this SLR.

Anomaly detection can be applied to a wide range of applications. We found 13 different applications in this SLR. Most of the research on DRL is about a network or intrusion-type anomaly detection. Researchers can experiment with DRL techniques for other anomaly detection applications, *e.g.*, video anomaly, wireless anomaly, or anomaly detection in the industry, as DRL has performed well for these applications.

There are various anomaly datasets available in the literature. Most of the anomaly data found in the research articles identified in this SLR using DRL techniques includes medical and network datasets, and DRL techniques have outperformed the ML and DL techniques. Another future avenue for researchers is to work on another anomaly dataset with these DRL techniques to prove to be better than other techniques for each type of anomaly data.

As we can see from table IV, Deep Q learning, actor-critic, policy gradient and reinforcement learning with RNN are the most valuable techniques of deep reinforcement learning, so researcher should explore their more variant, *e.g.*, double deep Q learning, Q network with autoencoders, meta policy and combine DL models with RL techniques and experiment on different applications to gain more evidence on DRL with anomaly detection.

VI. CLOSING REMARKS

This systematic literature review presents anomaly detection through Deep reinforcement learning (DRL). We collected a total of 32 research papers that used the DRL framework for anomaly detection from 2017 to 2022. We reviewed and analyzed these papers from these four perspectives: the type of anomaly detection application, the anomaly detection dataset, the proposed DRL techniques, and the DRL model performance over other alternative models.

For RQ1, we observed 13 different applications of anomaly detection that have been used in with DRL in selected papers. We have observed that the most popular anomaly detection with applications with DRL includes network intrusion detection, video anomaly detection, and general anomaly detection. In RQ2, we identified 50 different anomaly detection datasets from different specific anomaly detection applications. Most datasets are real-time datasets, and some are public datasets. As for RQ3, we

demonstrated 17 different DRL models that have been used for anomaly detection in the selected papers from 2017 to 2021. The most popular DRL methods are Deep Q learning, Actor critic, deep policy gradient, and neural networks with RL. Finally, for RQ4, we presented a performance comparison of the DRL technique with the alternative model from the selected papers.

ACKNOWLEDGEMENT

We wish to acknowledge the tremendous support from the Department of Computer and Information Sciences (CISD), UTP, Malaysia for all academic support and facilities.

REFERENCES

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Comput. Surv. CSUR, vol. 41, no. 3, pp. 1–58, 2009.
[2] M. Injadat, F. Salo, A. B. Nassif, A. Essex, and A. Shami, "Bayesian optimization with machine learning algorithms towards anomaly detection," in 2018 IEEE global communications conference (GLOBECOM), 2018, pp. 1–6
[3] F. Salo, A. B. Nassif, and A. Essex, "Dimensionality reduction with IG-PCA and ensemble classifier for network intrusion detection," Comput. Netw., vol. 148, pp. 164–175, 2019.
[4] F. Salo, M. Injadat, A. B. Nassif, A. Shami, and A. Essex, "Data mining techniques in intrusion detection systems: A systematic literature review," IEEE Access, vol. 6, pp. 56046–56058, 2018.
[5] F. Salo, M. Injadat, A. Moubayed, A. B. Nassif, and A. Essex, "Clustering enabled classification using ensemble feature selection for intrusion detection," in 2019 International Conference on Computing, Networking and Communications (ICNC), 2019, pp. 276–281.
[6] P. Gogoi, D. K. Bhattacharyya, B. Borah, and J. K. Kalita, "A survey of outlier detection methods in network anomaly identification," Comput. J., vol. 54, no. 4, pp. 570–588, 2011.
[7] S. Agrawal and J. Agrawal, "Survey on anomaly detection using data mining techniques," Procedia Comput. Sci., vol. 60, pp. 708–713, 2015.
[8] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection for discrete sequences: A survey," IEEE Trans. Knowl. Data Eng., vol. 24, no. 5, pp. 823–839, 2010.
[9] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine Learning for Anomaly Detection: A Systematic Review," IEEE Access, 2021.
[10] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," J. Artif. Intell. Res., vol. 4, pp. 237–285, 1996.
[11] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "A brief survey of deep reinforcement learning," ArXiv Prepr. ArXiv170805866, 2017.
[12] W. Wang et al., "Abnormal flow detection in industrial control network based on deep reinforcement learning," Appl. Math. Comput., vol. 409, p. 126379, 2021.
[13] Muneer, A., Taib, S.M., Fati, S.M., Balogun, A.O. and Aziz, I.A., 2022. A Hybrid deep learning-based unsupervised anomaly detection in high dimensional data. *Computers, Materials and Continua*, *70*(3), pp.6073-6088.
[14] V. Hodge and J. Austin, "A survey of outlier detection methodologies," Artif. Intell. Rev., vol. 22, no. 2, pp. 85–126, 2004.
[15] M. Ahmed, A. N. Mahmood, and M. R. Islam, "A survey of anomaly detection techniques in financial domain," Future Gener. Comput. Syst., vol. 55, pp. 278–288, 2016.
[16] A. A. Sodemann, M. P. Ross, and B. J. Borghetti, "A review of anomaly detection in automated surveillance," IEEE Trans. Syst. Man Cybern. Part C Appl. Rev., vol. 42, no. 6, pp. 1257–1272, 2012.
[17] R. Zuo, "Machine learning of mineralization-related geochemical anomalies: A review of potential methods," Nat. Resour. Res., vol. 26, no. 4, pp. 457–464, 2017.
[18] S. He, J. Zhu, P. He, and M. R. Lyu, "Experience report: System log analysis for anomaly detection," in 2016 IEEE 27th International Symposium on Software Reliability Engineering (ISSRE), 2016, pp. 207–218.
[19] Y. Yu, "A survey of anomaly intrusion detection techniques," J. Comput. Sci. Coll., vol. 28, no. 1, pp. 9–17, 2012.
[20] C.-F. Tsai, Y.-F. Hsu, C.-Y. Lin, and W.-Y. Lin, "Intrusion detection by machine learning: A review," Expert Syst. Appl., vol. 36, no. 10, pp. 11994–12000, 2009.
[21] A. Patcha and J.-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," Comput. Netw., vol. 51, no. 12, pp. 3448–3470, 2007.
[22] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," IEEE Commun. Surv. Tutor., vol. 18, no. 2, pp. 1153–1176, 2015.
[23] K. Satpute, S. Agrawal, J. Agrawal, and S. Sharma, "A survey on anomaly detection in network intrusion detection system using particle swarm optimization based machine learning techniques," in Proceedings of the international conference on frontiers of intelligent computing: theory and applications (FICTA), 2013, pp. 441–452.
[24] V. Sharma, R. Kumar, W.-H. Cheng, M. Atiquzzaman, K. Srinivasan, and A. Y. Zomaya, "NHAD: Neuro-fuzzy based Horizontal Anomaly Detection in online social networks," IEEE Trans. Knowl. Data Eng., vol. 30, no. 11, pp. 2171–2184, 2018.
[25] P. Zhao, Y. Zhang, M. Wu, S. C. Hoi, M. Tan, and J. Huang, "Adaptive cost-sensitive online classification," IEEE Trans. Knowl. Data Eng., vol. 31, no. 2, pp. 214–228, 2018.
[26] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: methods, systems and tools," Ieee Commun. Surv. Tutor., vol. 16, no. 1, pp. 303–336, 2013.
[27] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, and K. J. Kim, "A survey of deep learning-based network anomaly detection," Clust. Comput., vol. 22, no. 1, pp. 949–961, 2019.
[28] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," 2007.
[29] D. Zha, K.-H. Lai, M. Wan, and X. Hu, "Meta-AAD: Active Anomaly Detection with Deep Reinforcement Learning," in 2020 IEEE International Conference on Data Mining (ICDM), 2020, pp. 771–780.
[30] C. Zhong, M. C. Gursoy, and S. Velipasalar, "Deep actor-critic reinforcement learning for anomaly detection," in 2019 IEEE Global Communications Conference (GLOBECOM), 2019, pp. 1–6.
[31] F. de La Bourdonnaye, C. Teuliere, T. Chateau, and J. Triesch, "Learning of binocular fixations using anomaly detection with deep reinforcement learning," in 2017 International Joint Conference on Neural Networks (IJCNN), 2017, pp. 760–767.
[32] S. Aberkane and M. Elarbi, "Deep reinforcement learning for real-world anomaly detection in surveillance videos," in 2019 6th International Conference on Image and Signal Processing and their Applications (ISPA), 2019, pp. 1–5.
[33] T. Wu and J. Ortiz, "Towards adaptive anomaly detection in buildings with deep reinforcement learning," in Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, 2019, pp. 380–382.
[34] R. F. Mansour, J. Escorcia-Gutierrez, M. Gamarra, J. A. Villanueva, and N. Leal, "Intelligent video anomaly detection and classification using faster RCNN with deep reinforcement learning model," Image Vis. Comput., p. 104229, 2021.
[35] K. Sethi, R. Kumar, D. Mohanty, and P. Bera, "Robust Adaptive Cloud Intrusion Detection System Using Advanced Deep Reinforcement Learning," in International Conference on Security, Privacy, and Applied Cryptography Engineering, 2020, pp. 66–85.
[36] M. Lopez-Martin, B. Carro, and A. Sanchez-Esguevillas, "Application of deep reinforcement learning to intrusion detection for supervised problems," Expert Syst. Appl., vol. 141, p. 112963, 2020.
[37] K. Sethi, R. Kumar, N. Prajapati, and P. Bera, "Deep reinforcement learning based intrusion detection system for cloud infrastructure," in 2020 International Conference on COMmunication Systems & NETworkS (COMSNETS), 2020, pp. 1–6.

[38] Y.-F. Hsu and M. Matsuoka, "A Deep Reinforcement Learning Approach for Anomaly Network Intrusion Detection System," in 2020 IEEE 9th International Conference on Cloud Networking (CloudNet), 2020, pp. 1–6.
[39] H. Benaddi, K. Ibrahimi, A. Benslimane, and J. Qadir, "A Deep Reinforcement Learning Based Intrusion Detection System (DRL-IDS) for Securing Wireless Sensor Networks and Internet of Things," in International Wireless Internet Conference, 2019, pp. 73–87.
[40] C. Kim and J. Park, "Designing online network intrusion detection using deep auto-encoder Q-learning," Comput. Electr. Eng., vol. 79, p. 106460, 2019.
[41] E. Suwannalai and C. Polprasert, "Network Intrusion Detection Systems Using Adversarial Reinforcement Learning with Deep Q-network," in 2020 18th International Conference on ICT and Knowledge Engineering (ICT&KE), 2020, pp. 1–7.
[42] J. Watts, F. van Wyk, S. Rezaei, and Y. Wang, "A Dynamic Deep Reinforcement Learning-Bayesian Framework for Anomaly Detection".
[43] H. Alavizadeh, J. Jang-Jaccard, and H. Alavizadeh, "Deep Q-Learning based Reinforcement Learning Approach for Network Intrusion Detection," ArXiv Prepr. ArXiv211113978, 2021.
[44] G. P. Gupta and others, "Intrusion Detection Framework Using an Improved Deep Reinforcement Learning Technique for IoT Network," in Soft Computing for Security Applications, Springer, 2022, pp. 765–779.
[45] J. Tao, T. Han, and R. Li, "Deep-Reinforcement-Learning-Based Intrusion Detection in Aerial Computing Networks," IEEE Netw., vol. 35, no. 4, pp. 66–72, 2021.
[46] Phan, Trung V., and Thomas Bauschert. "Deepair: Deep reinforcement learning for adaptive intrusion response in software-defined networks." IEEE Transactions on Network and Service Management (2022).
[47] Wang, Zhihao, et al. "A Deep Reinforcement Learning based Intrusion Detection Strategy for Smart Vehicular Networks." IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2022.
[48] Tharewal, Sumegh, et al. "Intrusion detection system for industrial Internet of Things based on deep reinforcement learning." Wireless Communications and Mobile Computing 2022 (2022).
[49] Alavizadeh, Hooman, Hootan Alavizadeh, and Julian Jang-Jaccard. "Deep Q-Learning based Reinforcement Learning Approach for Network Intrusion Detection." Computers 11.3 (2022): 41.
[50] Joseph, Geethu, M. Cenk Gursoy, and Pramod K. Varshney. "Temporal Detection of Anomalies via Actor-Critic Based Controlled Sensing." 2021 IEEE Global Communications Conference (GLOBECOM). IEEE, 2021.
[51] Li, Yaze, and Jingxian Wu. "Low latency cyberattack detection in smart grids with deep reinforcement learning." International Journal of Electrical Power & Energy Systems 142 (2022): 108265.
[52] Li, Yang, Quan Pan, and Erik Cambria. "Deep-attack over the deep reinforcement learning." Knowledge-Based Systems (2022): 108965.
[53] Gupta, Govind P. "Intrusion Detection Framework Using an Improved Deep Reinforcement Learning Technique for IoT Network." Soft Computing for Security Applications. Springer, Singapore, 2022. 765-779.
[54] Bakhshad, Said, et al. "Deep Reinforcement Learning based Intrusion Detection System with Feature Selections Method and Optimal Hyper-parameter in IoT Environment." 2022 International Conference on Computer, Information and Telecommunication Systems (CITS). IEEE, 2022.
[55] Zhong, Chen, M. Cenk Gursoy, and Senem Velipasalar. "Controlled Sensing and Anomaly Detection Via Soft Actor-Critic Reinforcement Learning." ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.
[56] Fährmann, Daniel, et al. "Double Deep Q-Learning With Prioritized Experience Replay for Anomaly Detection in Smart Environments." IEEE Access 10 (2022): 60836-60848.
[57] Alavizadeh, Hooman, Hootan Alavizadeh, and Julian Jang-Jaccard. "Deep Q-Learning based Reinforcement Learning Approach for Network Intrusion Detection." Computers 11.3 (2022): 41.
[58] Zhang, Lei, et al. "A Hidden Attack Sequences Detection Method Based on Dynamic Reward Deep Deterministic Policy Gradient." Security and Communication Networks 2022 (2022).
[59] Nunes, Eduardo C. "Machine Learning based Anomaly Detection for Smart Shirt: A Systematic Review." *arXiv preprint arXiv:2203.03300* (2022).
[60] Padakandla, Sindhu. "A survey of reinforcement learning algorithms for dynamically varying environments." *ACM Computing Surveys (CSUR)* 54.6 (2021): 1-25.
[61] Pawar, Karishma, and Vahida Attar. "Deep learning approaches for video-based anomalous activity detection." *World Wide Web* 22.2 (2019): 571-601.
[62] Alla, Sridhar, and Suman Kalyan Adari. *Beginning anomaly detection using python-based deep learning*. New Jersey: Apress, 2019.
[63] Kang, Myeongsu. "Machine learning: Anomaly detection." *Prognostics and health management of electronics: fundamentals, machine learning, and the internet of things* (2018): 131-162.
[64] Bommes, Lukas, et al. "Anomaly detection in IR images of PV modules using supervised contrastive learning." *Progress in Photovoltaics: Research and Applications* 30.6 (2022): 597-614.
[65] Akcay, Samet, Amir Atapour-Abarghouei, and Toby P. Breckon. "Ganomaly: Semi-supervised anomaly detection via adversarial training." *Asian conference on computer vision*. Springer, Cham, 2018.
[66] Sovilj, Dušan, et al. "A comparative evaluation of unsupervised deep architectures for intrusion detection in sequential data streams." *Expert Systems with Applications* 159 (2020): 113577.
[67] Gouda, Walaa, et al. "Unsupervised Outlier Detection in IOT Using Deep VAE." *Sensors* 22.17 (2022): 6617.
[68] Zhong, Chen, M. Cenk Gursoy, and Senem Velipasalar. "Deep actor-critic reinforcement learning for anomaly detection." *2019 IEEE global communications conference (GLOBECOM)*. IEEE, 2019.

Kinza Arshad received the bachelor's degree in computer science from University of Management and Technology, Lahore, Pakistan, in 2020, and currently doing M.S. degree in Data Science from University of Management and Technology, Lahore, Pakistan. Her projects and research are mainly focused on Machine Learning, Deep Learning and Machine Translation.

RAO FAIZAN ALI received the bachelor's degree in computer science from COMSATS University Islamabad, Pakistan, and the M.Phil. degree in computer science from the University of Management and Technology, Lahore, Pakistan. He is currently pursuing the Ph.D. degree with University Technology PETRONAS, Malaysia. He has eight years of experience in teaching and research. He has been with various computer science positions in financial, consulting, academia, and government sectors. He is currently working as a Research Officer with the Department of Computer and information Sciences, University Technology Petronas, Perak, Malaysia.

Izzatdin Abdul Aziz received the Ph.D. degree in information technology from Deakin University, Australia, working on the domain of hydrocarbon exploration and cloud computing. He is currently a Researcher with the High-Performance Cloud Computing Centre (HPC3), Universiti Teknologi Petronas (UTP), where he focuses on solving complex upstream oil and gas (O&G) industry problems from the viewpoint of computer sciences. He also serves as the Deputy Head of the Computer and Information Sciences Department, UTP. He is working closely with O&G companies in delivering solutions for complex problems, such as offshore O&G pipeline corrosion rate prediction, O&G pipeline corrosion detection, securing data on clouds, designing and implementing Metocean prediction system, and bridging upstream and downstream oil and gas businesses through data analytics. He is also working on big data transmission, security, and optimization problems on high performance clouds.

AMGAD MUNEER received the B.Eng. degree (Hons.) in mechatronic engineering from the Asia Pacific University of Technology and Innovation (APU) in 2018 and master's degree in information technology from Universiti Teknologi PETRONAS in 2022, Malaysia. He has authored several ISI and Scopus journal articles/conference papers. His research interests include machine learning, image processing, the Internet of Things, machine vision, bioinformatics. He is a Reviewer in some international impact-factor journals.

SHERAZ NASEER received the M.S. degree in information security and the Ph.D. degree in computer science. He received the professional certifications of IT including, CISSP, CoBit, and ITIL. He has 15 years of experience in industry and academia. He is currently working as an Assistant Professor with the Department of Computer Science, University of Management and Technology, Lahore, Pakistan. His research interests include bioinformatics, data driven information security, and anomaly detection.

Nabeel Sabir Khan was born in Lahore, Pakistan. He received the M.C.S. and M.S. degrees from the University of Central Punjab, and the Ph.D. degree from the University of Management and Technology, Pakistan, in 2020. He is currently working as an Assistant Professor with the University of Management and Technology. He has more than 13 years of teaching experience. He is also the Regional Director of ACM-ICPC ASIA, Lahore. His research interests include theory of programming language, machine translation, and computer science education.

Shakirah Mohd Taib is a lecturer and researcher at Centre for Research in Data Science (CeRDaS) in Universiti Teknologi PETRONAS (UTP), Malaysia. She obtained a bachelor's degree in information technology from Universiti Utara Malaysia and Master of Computing from University of Tasmania, Australia. She has more than 15years working experience at Universiti Teknologi Petronas (UTP). Her area of specialization includes data science, machine learning, knowledge discovery and information retrieval using Artificial Intelligence techniques. Shakirah is a member of international organization such as IEEE, Malaysia Board of Technologists (MBOT) and Association for Information Systems (AIS).