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 published 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, systematic review.

I. INTRODUCTION
Anomaly detection is a significant problem that has been 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 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], [4], [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 banks or credit card transaction data, which might suggest fraud [7]. Furthermore, identifying an anomaly from an aviation detector may lead
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 in 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 an 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 discusses the related work, Section 3 contains the methodology used to do this research, Section 4 consists 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 subsections 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

![FIGURE 1. Examples of anomalies categories [63].](image-url)
compare 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 $t_1$ (winter), a temperature of 20°F is typical.
However, a temperature of 20°F at time $t_2$ (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.

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 com-
 parable to supervised learning, with the exception that
  the algorithm is not taught using a dataset. The rein-
 forcement 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 com-
  bines 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 outper-
  forms 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 applica-
tions. 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 detect-
tion applications and techniques. Cyber intrusion detection,
fraud detection, medical anomaly detection, industrial dam-
age detection, image processing detection, textual anomaly
detection, and sensor networks are all instances of cyber
intrusion detection. However, this anomaly comprehensive
survey lacks 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] 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 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
using 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 con-
ducted by [59], the authors only focused on anomaly detect-
tion 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 cer-
tain 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], to the contrary, 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 two primary production log datasets.

Many studies focused on anomalous intrusion detection. For example, in [19], the author 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 presented 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 system. 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) VIDEO ANOMALY DETECTION

In surveillance videos, 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 acts 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 shows 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 system relies 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 program, 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 DRL techniques. 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 DRL techniques, which is the primary motivation behind this research. Our systematic literature review is considerably different from those described in the earlier section, 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 a 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:

- Articles that written in English including scientific journals and conference proceedings.
- Articles on anomaly detection or its application.
- Articles which use the DRL technique to detect anomalies.
- Articles published from 2017 to 2022.

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

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

### 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 4 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 Section B.
3. Remove review papers.
4. Apply quality assessment rules to include the best-selected paper for this SLR.
5. 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.

### 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 that they have used for anomaly detection. 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, 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 as shown in TABLE 1. To address RQ3, we mention the DRL models used in each selected paper in TABLE 2. To address RQ4, we made a performance comparison of each DRL model discussed in selected papers are presented in TABLE 3.

### IV. RESULTS AND DISCUSSION

This section provides an overview of the chosen papers. In the following section, 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 implemented and discussed deep reinforcement learning and anomaly detection application. These research articles were published from 2017 to...
| 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]      |
### TABLE 1. (Continued.) Selected research papers.

| No. | Paper ID | Title |
|-----|----------|-------|
| P25 | 2022 Elsevier [51] | “Low latency cyberattack detection in smart grids with deep reinforcement learning” |
| P26 | 2022 Elsevier [52] | “Deep-attack over the deep reinforcement learning” |
| P27 | 2022 Springer [53] | “Intrusion Detection Framework Using an Improved Deep Reinforcement Learning Technique for IoT Network” |
| P28 | 2022 IEEE [54] | Deep Reinforcement Learning based Intrusion Detection System with Feature Selections Method and Optimal Hyper-parameter in IoT Environment” |
| P29 | 2022 IEEE [55] | “Controlled Sensing and Anomaly Detection Via Soft Actor-Critic Reinforcement Learning” |
| P30 | 2022 IEEE [56] | “Double Deep Q-Learning With Prioritized Experience Replay for Anomaly Detection in Smart Environments” |
| P31 | 2022 MDPI [57] | “Deep Q-Learning Based Reinforcement Learning Approach for Network Intrusion Detection” |
| P32 | 2022 Hindawi [58] | “A Hidden Attack Sequences Detection Method Based on Dynamic Reward Deep Deterministic Policy Gradient” |

### 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 |

2022, which is relatively recent. The list of chosen papers for this SLR is given in Table 1.

### A. ANOMALY DETECTION APPLICATION (RQ1)

In this section, we address Research Question 1 (RQ1), 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 2 lists these applications and mentions the paper discussing them.

As shown in Table 2, 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 frequency of each application discussed in the selected papers is tabulated based on the anomaly application category. The results show that DRL techniques work well with applications listed in TABLE 2.
| Paper ID | Databases                | Datasets for Anomaly detection                      | Frequency |
|---------|--------------------------|-----------------------------------------------------|-----------|
| NB15    | Analysis                 |                                                     |           |
|         | DoS                      |                                                     |           |
|         | Exploits                 |                                                     |           |
|         | Fuzzers                  |                                                     |           |
|         | Generic                  |                                                     |           |
| P1      | Recon                    |                                                     |           |
|         | Thyroid                  | Hypothyroi                                          | 12        |
|         |                          | Subnormal                                           |           |
|         | HAR                      | Downstairs                                          |           |
|         |                          | Upstairs                                            |           |
|         | Covertype                | Cottonwood                                          |           |
|         |                          | Douglas-fir                                         |           |
|         |                          | Annthyroid                                          |           |
|         |                          | Arrhythmia                                          |           |
|         |                          | Breastw                                             |           |
|         |                          | Cardio                                              |           |
|         |                          | Glass                                               |           |
|         |                          | Ionosphere                                          |           |
|         |                          | Letter                                              |           |
| P2      | Lympho                   |                                                     | 24        |
|         |                          | Mammography                                         |           |
|         |                          | 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         |
| P9      | NSL-KDD                  |                                                     | 2         |
| P10     |                          |                                                     | 1         |
| P11     | NSL-KDD                  |                                                     | 2         |
|         |                          | UNSW-NB15                                           |           |
Figure 4 shows 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

### TABLE 3. (Continued.) Anomaly detection datasets utilized in selected research articles.

| ID  | Dataset Description                                      | Count |
|-----|-----------------------------------------------------------|-------|
| 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 | Software 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                      | 2     |
| P31 | Local position data of person for fall detection          |       |
| P32 | NSL-KDD                                                   | 1     |
|     | 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 papers, which is one of the primary goals of this review. Table 4 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 helps 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:

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

To make reinforcement learning effective in 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 propensity to exaggerate some values. Therefore, provided the prediction error is maintained at 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: Double Deep Q-learning. It is possible to choose exaggerated 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.

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 train the neural net regulators effectively. According to the results, the best technique, an asynchronous actor-critic variation, exceeds the most significant algorithms currently available. 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.
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)                               | Video anomaly detection       |
|          | Deep Q learning Network (DQN)                                                      |                               |
| P8       | Double Deep Q-Network (DDQN)                                                       | Cloud intrusion detection     |
|          | DRL Adaptive IDS                                                                   |                               |
| P9       | Deep Q learning Network (DQN)                                                      | Intrusion detection           |
|          | Double Deep Q-Network (DDQN)                                                       |                               |
|          | Policy Gradient (PG)                                                               |                               |
|          | Actor-Critic (AC)                                                                  |                               |
| 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 intrusion 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        |

for each potential state. Except for 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, batches, and for each iteration, a new mini-batch is created because 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,
| Paper ID | Anomaly detection application | Model                                                                 | Dataset                          | Performance Comparison                                      |
|---------|-------------------------------|----------------------------------------------------------------------|----------------------------------|-------------------------------------------------------------|
| P1      | General Anomaly detection     | DPLAN                                                                | NB15                             | 23%-98% relative AUC-PR improvement                         |
|         |                               |                                                                      | Thyroid                          |                                                             |
|         |                               |                                                                      | HAR                              |                                                             |
|         |                               |                                                                      | Cover type                       |                                                             |
|         |                               |                                                                      | 24 different anomaly datasets    | Outperform alternative unsupervised, SSDO, AAD and FIF models. |
| P2      | General Anomaly detection     | Meta-ADD                                                              |                                  |                                                             |
|         |                               |                                                                      |                                  |                                                             |
| 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     |                                                             |
|         | Video anomaly detection       | Deep Q Learning Network (DQN)                                       |                                  |                                                             |
|         | Building anomaly detection    | Deep Deterministic Policy Gradient (DDPG)                            | Building Signal data             | Up to 3x better than alternative models                     |
| P7      | Video anomaly detection       | IVADC-FDRL                                                            | UCSD anomaly                     | 98.50% accuracy                                             |
|         | Cloud intrusion detection     | Deep Q Learning Network (DQN)                                       |                                  | 94.80% accuracy over other models                          |
|         |                               | DRL Adaptive IDS                                                     | ISOT-CID                         | Outperform state-of-the-art                                |
|         |                               |                                                                      | NSL-KDD                          |                                                             |
| P9      | Intrusion detection           | Deep Q Learning Network (DQN)                                       | NSL-KDD                          | 87.87% accuracy                                             |
|         |                               | Double Deep Q-Network (DDQN)                                        | AWID                             | 89.78% accuracy                                             |
|         |                               | Policy Gradient (PG)                                                 |                                  | 78.73% accuracy                                             |
|         |                               | Actor-Critic (AC)                                                    |                                  | 80.78% accuracy                                             |
|         | Cloud intrusion detection     | DRL-adaptive cloud IDS                                              | UNSW-NB15                        | Outperform all existing works                              |
| P10     |                               |                                                                      |                                  |                                                             |
| P11     | Network intrusion detection   | DRL-NIDS                                                              | NSL-KDD                          | 91.4% accuracy                                              |
|         |                               |                                                                      | UNSW-NB15                        | 91.8% accuracy                                              |
|         |                               |                                                                      | Real-time campus network traffic data | 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                                         | Online Network data              | Outperform state-of-the-art                                |
|         |                               | (DAEQ-N)                                                             |                                  |                                                             |
### TABLE 5. (Continued.) Performance analysis of DRL algorithms.

| P14 | Industrial network control | Deep Reinforcement Network (DRNN) | Neural Network data | Industrial Network data | 98.06% accuracy |
|-----|-----------------------------|-----------------------------------|---------------------|-------------------------|-----------------|
| P15 | Network intrusion detection | Adversarial/Multi-Agent Reinforcement Deep Q-Learning | 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)      | Software 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 Local position data of person for fall detection | 92.6% accuracy |
| 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 |
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 using well-tailed 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), an active anomaly
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.

V. CONCLUSION

A. THEORETICAL IMPLICATIONS

First, this study is among the first systematic literature review on Anomaly detection using Deep reinforcement learning (DRL) techniques. Although some literature reviews are available, none of the studies has addressed these methods appropriately. Our systematic literature review focuses on presenting extensive research on detecting anomalies using DRL techniques, datasets used and the performance of each DRL model. 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, DRL has outperformed Deep learning and Machine learning in many ways. DRL models combine artificial neural networks with reinforcement learning to help the agent learns 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.

In anomaly detection study, 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 which limited number of research articles because it is a new technique that was started in 2017. Therefore, 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 exclude 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 with only 32 papers published from 2017 to 2022. Therefore, we highly recommend that other researchers 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 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 4, 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 models from the selected papers.

**ACKNOWLEDGMENT**

The authors would like to thank the Center for Research in Data Science (CeRDaS) and Institute of Autonomous Systems and Autonomous Facilities (IAS), UTP, Malaysia for the support and paying the APC.

**REFERENCES**

[1] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” ACM Comput. Surv., vol. 41, no. 3, pp. 1–58, Jul. 2009.
[2] M. Injadat, F. Salo, A. B. Nassif, A. Essex, and A. Shami, “Bayesian optimization with machine learning algorithms towards anomaly detection,” in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 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, Jan. 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 Proc. Int. Conf. Comput., Netw. Commun. (ICNC), Feb. 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, Apr. 2011.
[7] S. Agrawal and J. Agrawal, “Survey on anomaly detection using data mining techniques,” Proc. Comput. Sci., vol. 60, pp. 708–713, Jan. 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, May 2012.
[9] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, “Machine learning for anomaly detection: A systematic review,” IEEE Access, vol. 9, pp. 78658–78700, 2021.
[10] L. P. Kaebbling, M. L. Littman, and A. W. Moore, “Reinforcement learning: A survey,” J. Artif. Intell. Res., vol. 4, no. 1, pp. 237–285, Jan. 1996.
[11] K. Arulkumaran, M. P. Deisenroth, M. Branlard, and A. A. Bharath, “A brief survey of deep reinforcement learning,” 2017, arXiv:1708.05866.
[12] F. Wang, J. Guo, Z. Wang, H. Wang, J. Cheng, C. Wang, M. Yuan, J. Kurths, X. Luo, and Y. Gao, “Abnormal flow detection in industrial control network based on deep reinforcement learning,” Appl. Math. Comput., vol. 409, Nov. 2021, Art. no. 126379.
[13] A. Munee, S. M. Taib, S. M. Fati, A. O. Balogun, and I. A. Aziz, “A Hybrid deep learning-based unsupervised anomaly detection in high dimensional data,” Comput. Mathe. Continu., vol. 70, no. 3, pp. 6073–6088, 2022.
[14] V. J. 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, Feb. 2016.

[16] A. A. Sodemann, M. Ross, and B. J. Borghetti, “A review of anomaly detection in automated surveillance,” IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 42, no. 6, pp. 1257–1272, Nov. 2012.

[17] R. Zuo, “Machine learning of mineralization-related geochemical anomalies: A review of potential methods,” Natural Resour. Rev., vol. 26, no. 4, pp. 457–464, Oct. 2017.

[18] S. He, J. Zhu, P. He, and M. R. Lyu, “Experience report: System log analysis for anomaly detection,” in Proc. IEEE 27th Int. Symp. Softw. Rel. Eng. (ISSRE), Oct. 2016, pp. 207–218.

[19] Y. Yu, “A survey of anomaly invasion detection techniques,” J. Comput. Sci. Colleges, vol. 28, no. 1, pp. 9–17, Oct. 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, Aug. 2007.

[22] A. L. Bukrz and E. Guven, “A survey of data mining and machine learning methods for cyber security intrusion detection,” IEEE Commun. Surveys Tuts., vol. 18, no. 2, pp. 1153–1176, 2nd Quart., 2016.

[23] K. Slaypute, 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 Proc. Int. Conf. Frontiers Infell. Comput. Appl. (FICATA), 2013, pp. 441–452.

[24] V. Sharma, R. Kumar, W.-H. Cheng, M. Atiquzzaman, K. Sripanivasa, 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, Nov. 2018.

[25] P. Zhao, Y. Zhang, M. Wu, S. C. H. Hoi, M. Tan, and J. Huang, “Adaptive cost-sensitive online classification,” IEEE Trans. Knowl. Data Eng., vol. 31, no. 2, pp. 214–228, Feb. 2019.

[26] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, “Network anomaly detection: Methods, systems and tools,” IEEE Commun. Surveys Tuts., vol. 16, no. 1, pp. 303–336, 1st Quart., 2014.

[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,” Cluster Comput., vol. 22, pp. 949–961, Jan. 2019.

[28] B. Kitchenham and S. Charters, “Guidelines for performing systematic literature reviews in software engineering,” Dept. Comput. Sci., Univ. Durham, Durham, U.K., Tech. Rep. Version 2.3, 2007.

[29] T. V. Phan and T. Bauschert, “DeepAir: Deep reinforcement learning for intrusion detection in automated surveillance,” in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 2019, pp. 1–6.

[30] S. K. Arshad et al.: Deep Reinforcement Learning for Anomaly Detection: A Systematic Review

[31] M. Shabaz, “Intrusion detection system for industrial Internet of Things based on deep reinforcement learning,” in Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS), May 2022, pp. 1–6.

[32] G. P. Gupta, “Intrusion detection framework using an improved deep reinforcement learning technique for IoT network,” in Soft Computing for Security Applications. Cham, Switzerland: Springer, 2022, pp. 765–779.

[33] Y. Li, T. Han, and R. Li, “Deep-reinforcement-learning-based intrusion detection in aircraft communications,” IEEE Netw., vol. 35, no. 4, pp. 66–72, Jul. 2021.

[34] T. V. Phan and T. Bauschert, “DeepAir: Deep reinforcement learning for adaptive intrusion response using software-defined networks,” IEEE Trans. Netw. Service Manage., vol. 19, no. 3, pp. 2207–2218, Sep. 2022.

[35] Z. Wang, D. Jiang, Z. Lv, and H. Song, “A deep reinforcement learning-based intrusion detection strategy for smart vehicular networks,” in Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS), May 2022, pp. 1–6.

[36] S. Tharewal, M. W. Ashfaque, S. S. Banu, P. Uma, S. M. Hassen, and M. Shabaz, “Intrusion detection system for industrial Internet of Things based on deep reinforcement learning,” Wireless Commun. Mobile Comput., vol. 2022, pp. 1–8, Mar. 2022.

[37] H. Alaviavadeh, H. Alaviavadeh, and J. Jang-Jaccard, “Deep Q-learning based reinforcement learning approach for network intrusion detection,” Computers, vol. 11, no. 3, p. 41, Mar. 2022.

[38] S. Bakhshad, V. Ponnusamy, R. Annur, M. Waqasyz, H. Alasmary, and W. Bai, “A hidden attack sequences detection method based on actor-critic reinforcement learning technique for IoT network,” in Soft Computing for Security Applications. Singapore: Springer, 2022, pp. 765–779.

[39] S. Bakhshad, V. Ponnusamy, R. Annur, M. Waqasyz, H. Alasmary, and W. Bai, “A hidden attack sequences detection method based on actor-critic reinforcement learning technique for IoT network,” in Soft Computing for Security Applications. Singapore: Springer, 2022, pp. 765–779.

[40] G. Joseph, M. C. Gursoy, and P. K. Varshney, “Temporal detection of anomalies via actor-critic based controlled sensing,” in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 2021, pp. 1–6.

[41] Y. Li, T. Han, and R. Li, “Deep-reinforcement-learning-based intrusion detection in aircraft communications,” IEEE Netw., vol. 35, no. 4, pp. 66–72, Jul. 2021.

[42] Y. Li, Q. Pan, and E. Cambria, “Deep-attack over the deep reinforcement learning-Bayesian framework for anomaly detection,” IEEE Access, vol. 10, no. 1, pp. 5629–5641, Feb. 2022.

[43] S. Tharewal, M. W. Ashfaque, S. S. Banu, P. Uma, S. M. Hassen, and M. Shabaz, “Intrusion detection system for industrial Internet of Things based on deep reinforcement learning,” in Wireless Commun. Mobile Comput., vol. 2022, pp. 1–8, Mar. 2022.

[44] G. P. Gupta, “Intrusion detection framework using an improved deep reinforcement learning technique for IoT network,” in Soft Computing for Security Applications. Singapore: Springer, 2022, pp. 765–779.

[45] S. Bakhshad, V. Ponnusamy, R. Annur, M. Waqasyz, H. Alasmary, and S. Tux, “Deep reinforcement learning based intrusion detection system with feature selection method and optimal hyper-parameter in IoT environment,” in Proc. Int. Conf. Comput. Inf., Telecommun. Syst. (CITS), Jul. 2022, pp. 1–7.

[46] C. Zhong, M. C. Gursoy, and S. Velipasalar, “Actor-critic deep reinforcement learning framework for anomaly detection,” in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 2021, pp. 1–6.

[47] G. P. Gupta, “Intrusion detection framework using an improved deep reinforcement learning technique for IoT network,” in Soft Computing for Security Applications. Singapore: Springer, 2022, pp. 765–779.

[48] S. Bakhshad, V. Ponnusamy, R. Annur, M. Waqasyz, H. Alasmary, and S. Tux, “Deep reinforcement learning based intrusion detection system with feature selection method and optimal hyper-parameter in IoT environment,” in Proc. Int. Conf. Comput. Inf., Telecommun. Syst. (CITS), Jul. 2022, pp. 1–7.

[49] G. P. Gupta, “Controlled retraining and early detection of an anomaly using self-actor-critic reinforcement learning,” in Proc. IEEE Conf. Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2022, pp. 4198–4202.

[50] D. Fahrmann, N. Jorek, N. Damer, F. Kirchbuchner, and A. Kuiper, “Double deep Q-learning with prioritized experience replay for anomaly detection in smart environments,” IEEE Access, vol. 10, pp. 60836–60848, 2022.

[51] H. Alaviavadeh, H. Alaviavadeh, and J. Jang-Jaccard, “Deep Q-learning based reinforcement learning approach for network intrusion detection,” Computers, vol. 11, no. 3, p. 41, Mar. 2022.

[52] L. Zhang, Z. Pan, Y. Pan, S. Guo, Y. Liu, S. Xia, Q. Zheng, H. Li, and W. Bai, “A hidden attack sequences detection method based on dynamic reward deep deterministic policy gradient,” Secur. Commun. Netw., vol. 2022, pp. 1–13, Jan. 2022.

[53] E. C. Nunes, “Machine learning based anomaly detection for smart shirt: A systematic review,” 2022, arXiv:2206.03300.
S. Alla and S. K. Adari, “Beginning Anomaly Detection Using Python-Based Deep Learning.” New York, NY, USA: Apress, 2019.

M. Kang, “Machine learning: Anomaly detection,” in Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet of Things, Hoboken, NJ, USA: Wiley, 2018, pp. 131–162.

L. Bommes, M. Hoffmann, C. Baehr-Lutz, T. Pickel, J. Hauch, C. Brahec, A. Maier, and I. M. Peters, “Anomaly detection in IR images of PV modules using supervised contrastive learning,” Photovolt. Res. Appl., vol. 30, no. 6, pp. 597–614, Jun. 2022.

S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, “GANomaly: Semi-supervised anomaly detection via adversarial training,” in Proc. Asian Conf. Comput. Vis., vol. 159, Nov. 2020, Art. no. 113577.

W. Gouda, S. Tahir, S. Alanazi, M. Almufareh, and G. Alwakid, “Unsupervised outlier detection in IoT using deep VAE,” Sensors, vol. 22, no. 17, p. 6617, Sep. 2022.

C. Zhong, M. C. Gursoy, and S. Velipasalar, “Deep actor-critic reinforcement learning for anomaly detection,” in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 2019, pp. 1–6.

KINZA ARSHAD received the bachelor’s degree in computer science from the University of Management and Technology, Lahore, Pakistan, in 2020, where she is currently pursuing the M.S. degree in data science. Her projects and research interests include 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, and the Ph.D. degree from University Technology PETRONAS, Malaysia. He has ten years of experience in teaching and research. He is currently an Assistant Professor with the University of Management and Technology. He is with various computer science positions in financial, consulting, academia, and government sectors.

AMGAD MUNEER received the B.Eng. degree (Hons.) in mechatronic engineering from the Asia Pacific University of Technology and Innovation (APU), in 2018, and the master’s degree in information technology from Universiti Teknologi PETRONAS, Malaysia, in 2022, where he is currently pursuing the Ph.D. degree. Currently, he is a Research Assistant II with the Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. He has authored several high-impact articles in well-reputed journals and conferences. His research interests lie in AI applications for cancer data sciences, manufacturing data analytics, the Internet of Things, medical imaging, and bioinformatics. He is a reviewer of many international impact-factor journals.

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 is 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.

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

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 an Assistant Professor with the University of Central Punjab, Lahore, Pakistan. 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 (Member, IEEE) received the bachelor’s degree in information technology from Universiti Utara Malaysia and the M.Comp. degree from the University of Tasmania, Australia. She is a Lecturer and a Researcher with the Centre for Research in Data Science (CeRDaS), Universiti Teknologi PETRONAS (UTP), Malaysia. She has more than 15 years working experience at Universiti Teknologi Petronas (UTP). Her research interests include data science, machine learning, knowledge discovery, and information retrieval, using Artificial Intelligence techniques. She is a member of international organization, such as Malaysia Board of Technologists (MBOT) and Association for Information Systems (AIS).

* * *