Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Adaptive Multi-Agent Deep Reinforcement Learning for Distributed Zero Trust Architectures in Global Financial Ecosystems

**Ankur Mahida**

ahmahida13@gmail.com

Barclays (United Kingdom)

**Research Article**

**Keywords:** Financial Infrastructures, Cyber-Attacks, Zero Trust Architecture, Dynamical Trust Measurement, Dynamical Threat Abatement, Multi-Agent Deep Reinforcement Learning, Zero Trust Policy

**Posted Date:** August 13th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10406334/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Adaptive Multi-Agent Deep Reinforcement Learning for Distributed Zero Trust Architectures in Global Financial Ecosystems

Ankur Mahida

Site Reliability Engineer, Barclays

Whippany, New Jersey

ahmahida13@gmail.com

# Adaptive Multi-Agent Deep Reinforcement Learning for Distributed Zero Trust Architectures in Global Financial Ecosystems


## Abstract

Digitalization of global financial systems has been one of the main exposures of financial infrastructures to highly sophisticated cyber attacks. Based on the industry reports released recently, it is stated that close to 25% of global internet attacks target financial organizations and the average data breach cost per attack is about 5.9 million USD. However, although Zero Trust Architecture (ZTA) emerges as a promising security approach that guarantees continuous authentication and stringent control of access for all users and devices, conventional designs of ZTA rely on centralized policy enforcement architectures. This research proposes Adaptive Multi-Agent Deep Reinforcement Learning-Zero Trust Architecture (AMADRL-ZTA). In this regard, AMADRL-ZTA combines multi-agent deep reinforcement learning and distributed Zero Trust policy enforcement, where autonomous agents continuously scan network traffic, evaluate the level of trust, and change access control policies. The experimentally derived results show that AMADRL-ZTA has a 96.8% threat detection capability, 38% reduction in response time, and a near 31% improvement in terms of policy changes in comparison to the existing centralized Zero Trust architecture and the rule-based intrusion detection system. It is evident that the proposed framework has a high chance of improving the cyber resilience of global financial infrastructures and enabling secure digital transactions.



**Keywords:** Financial Infrastructures, Cyber-Attacks, Zero Trust Architecture, Dynamical Trust Measurement, Dynamical Threat Abatement, Multi-Agent Deep Reinforcement Learning, Zero Trust Policy.


## 1. Introduction

The high rate of digitalization of the world financial ecosystems has made them highly vulnerable to advanced cyber attacks. Interrelated digital systems, cloud computing, and real-time transaction systems on which financial institutions are highly dependent have become key victims of cyber-attacks. It has been reported that financial systems have a considerable share of global cyber incidences, and thus complex security measures are required. The old models of perimeter-based security are not adequate to handle the current threats that utilize the distributed network designs. This has led to an increasing demand of adaptive and intelligent security structures that can perform their operations in ever-changing environments. The development of cyber threats has led to the need to constantly monitor, make decisions in real-time, and have automated threat mitigation policies and strategies. Such needs have resulted in the investigation of sophisticated artificial intelligence methods in cybersecurity [1], [2].

Zero Trust Architecture (ZTA) has also become one of the emerging paradigms that can overcome these challenges by removing implicit trust in the network environments. In contrast to

old form of security models, ZTA is strict with its identity verification of all the users and devices that may be seeking to gain access to the resources. This will make access decisions grounded on dynamic authentication and authorization instead of having a fixed trust. ZTA adoption is now one of the most popular topics of protection of sensitive financial data and transactions. The current ZTA designs, however, tend to be centralized, restricting the performance of the distributed system in terms of scalability and adaptability. The centralization of control systems can create bottlenecks and delays in the implementation of policies thereby making the entire system less effective. Hence, the need to improve ZTA with distributed and intelligent decision-making is there [2], [17].

The nature of the contemporary financial network needs its security solutions to be able to work independently in geographically dispersed environment. Distributed systems are made up of a number of interacting components each having its security requirements and constraints. It is hard to manage the trust and apply security policy within such settings. The nature of cyber threats requires more than the use of the static rule-based system since attackers have continued to improve their tactics. This has seen the incorporation of machine learning methodology to cybersecurity in order to facilitate adaptive threat detection and reaction. The machine learning models have the capability to process high amounts of data on the network and detect network patterns that may indicate malicious activities. Nevertheless, the conventional machine learning methods are incapable of making consecutive decisions in dynamic settings [6], [15].

Deep Reinforcement Learning (DRL) has become a strong approach towards decision-making in uncertain and complex environments. DRL is an integration based learning method which involves deep learning as well as reinforcement learning to allow the agents to learn the best policy when interacting with the environment. This method is especially applicable in cybersecurity implementation where decisions will have to be made on the fly depending on changing threats. Systems that are built on DRL can constantly improve themselves through learning and experience, and are thus very flexible. Sequential decision-making processes can be modeled, which enables DRL to solve the problems that cannot be solved successfully using traditional approaches. Consequently, DRL has extensively found application in intrusion detection and network security practices [4], [13].

Multi-agent systems have also received some interest in addition to DRA because of their capability to address the task of distributed decision-making. In a multi-agent, the agents act in an autonomous manner as they do in a joint manner to reach a shared goal. The strategy fits well in the case of distributed financial ecosystems, where various elements on the network have to coordinate with one another to guarantee safety. Multi-agent reinforcement learning helps agents to learn both cooperative and competitive strategies to improve the overall system performance. Multi-agent systems are decentralized and thus, are more scalable and resilient to failures. The system is able to react to localized threats with greater efficiency since it is able to spread decision-making among various agents [14], [19].

In spite of these developments, there is still a difficult task to integrate DRA and multi-agent systems into the Zero Trust frameworks. The current solutions tend to use either centralized ZTA or standalone reinforcement learning models without using their potential together. Frameworks that utilize multi-agent DRL to evaluate trust dynamics, as well as enforce policies, in a distributed space, are lacking. Additionally, the financial systems are complex and need the models, which would be able to adapt to different network conditions and threats. To resolve these issues, a new solution involving the integration of the benefits of ZTA, DRL, and multi-agent systems is required. This type of integration has a potential of boosting financial infrastructures security and resilience [18], [17].

The other serious issue of financial cybersecurity is the dynamism of the evaluation of trust. Conventional trust models are based on a set of rules and fixed parameter, which is inadequate in reflecting the changing behavior of users and devices. Dynamic trust evaluation should be followed with constant examination of network activities and behavioral pattern analysis. This may be realized by smart systems which can learn and respond to emerging threats. Reinforcement learning can offer an appropriate model of dynamic trust management allowing agents to revise their policies depending on the observed results. This will make sure that the current state of the network is always in line with trust decisions [3], [8].

Moreover, the growing use of cloud computing and Internet of Things (IoT) technologies made the financial systems attack surface larger. Such technologies pose new points of vulnerability that can be exploited by the attackers. To provide security in these types of environments, scalable and flexible solutions are necessary to manage the various network conditions. These issues can be effectively solved by multi-agent systems based on DRL, which can redistribute the load of the computations, and allow making decisions in a local way. This minimizes the vulnerability of single points of failures and increases the general resiliency of the system. The introduction of the more sophisticated AI methods and the cybersecurity systems is the key to solving the new challenges [9], [7].

The suggested study presents an Adaptive Multi-Agent Deep Reinforcement Learning-based Zero Trust Architecture (AMADRL-ZTA) to deal with the shortcomings of the current systems. The framework uses various smart agents to track network activities, assess trust levels and apply security policies in a dynamic manner. The proposed model allows detecting the threat in real-time and responding to it using adaptive mechanisms by integrating DRL with distributed ZTA. The system will be able to work in a large scale financial setting and scale well as it is also designed to be robust. The decision-making process that is based on learning improves the system in responding to the complex and dynamic threats [18], [19].

To sum up, the increased sophistication of financial ecosystems around the world requires more sophisticated security solutions than the existing methods. Combining the concepts of Zero Trust with multi-agent deep reinforcement learning opens an opportunity of improving cybersecurity. The suggested AMADRL-ZTA model should mitigate the issue of scalability, adaptability, and

dynamic threat management in distributed settings. The system is based on smart agents and dynamic learning methods which provide better threat detection and response. The study will be of great importance in the creation of future cybersecurity solutions to the financial systems that will remain resilient and safe in terms of digital activities [1], [2].

### 1.1 Hypothesis

1. Multi-agent deep reinforcement learning with Zero Trust Architecture enhances the accuracy of threat detection over centralized systems of the past.
2. The distributed agent-based policy enforcement minimizes the response time in other dynamic financial network environments.
3. The reinforcement-based adaptive trust evaluation improves decision making in the case of uncertain threats.
4. Distributed financial systems should encompass multi-agent collaboration to enhance its scalability and resilience in large-scale applications.
5. The proposed AMADRL-ZTA framework has better dynamic threat mitigation than the rule-based intrusion detection systems.

### 1.2 Research Contributions

1. Suggests a new AMADRL-ZTA system that combines multi-agent deep reinforcement learning and Zero Trust Architecture.
2. Presents a trust evaluation model that is dynamic and adapts its policies to reinforcement learning.
3. Creates a multi agent system distributed to monitor and eliminate threats in real-time.
4. Increases the efficiency of policy enforcement by using decentralized decision making policies.
5. Detection and response time to threats is significantly higher than traditional methods.
6. Delivers a flexible and secure framework that can be used in worldwide financial

environments.

7. Shows the usefulness of the suggested model due to detailed experimental assessment.

## 2. Literature Survey

Zero Trust Architecture (ZTA) was initially developed to remove implicit trust in the network environment and implement strict mechanisms of access control. The initial attempts focused on perennial authentication and verification of both users and devices irrespective of their location within the network [1]. This design was a great way of enhancing security over the old-fashioned models of perimeters. The framework was further standardized by the use of formal guidelines that described how the framework should be implemented in the contemporary systems. Nevertheless, some ZTA models at the beginning were highly centralized and unable to adapt to dynamic environments. Such constraints limit their applicability in distributed financial ecosystems such that real time decision-making is important. Thus, the development of ZTA by introducing the intelligent and decentralized mechanisms has turned into an active research field [2].

The most recent developments have investigated how machine learning methods may be incorporated into the security systems. Intrusion detection systems that use machine learning have proved that they can be able to detect patterns and anomalies in network traffic. These systems are able to identify known and unknown attacks using massive datasets [6]. Nevertheless, the conventional machine learning models are very much dependent on the fixed training data, and are ineffective in adapting to the changing threats. This is a weakness that renders them ineffective in dynamic environments like financial networks. In addition, these models usually fail to make sequential decisions that are needed in real-time threat mitigation. This has led to the development of more sophisticated methods like deep learning and reinforcement learning being studied by researchers [15].

Deep learning has played a prominent role in enhancing the effectiveness of cybersecurity systems in that automatic features can be extracted out of complex datasets. Intrusion detection and traffic analysis has been extensively done using neural networks especially convolutional and recurrent architectures [8]. These models are able to describe spatial and temporal profiles of network data. Deep learning models are computationally expensive and need massive volumes of labeled data to train except in their benefits. Also, they are not capable of dynamically adapting to emerging threats without retraining. This has prompted the consideration of the technique of reinforcement learning that can learn directly through the interaction with the environment [13].

Deep Reinforcement Learning (DRL) is now a promising technology in adaptive decision-making in cybersecurity. DRL allows the agents to acquire the best policies after trying and failing in an environment. Such a solution is especially applicable to dynamic situations whereby pre-set rules do not suffice [4]. Applications of DRL have been done successfully in network intrusion detection, resource allocation and traffic management. DRL is appropriate in changing

threat environments due to its capacity to keep learning and changing. Nevertheless, the majority of the currently existing DRR-based systems are centralized and are not capable of effectively responding to the needs of distributed systems [18].

Multi-agent reinforcement learning (MARL) builds up the features of DRL by allowing several agents to work and decide on distributed settings. The agents in MARL systems are able to exchange information and to synchronize their activities towards reaching shared goals [14]. This has been demonstrated to increase scalability and resilience in high-scale networks. MARL was used in some other fields, such as network optimization and distributed control systems. Decentralization of MARL enables it to be applied to complicated systems like financial ecosystems. Nonetheless, other issues that pose a challenge in MARL systems include coordination, overhead of communication and convergence [19].

Multi-agent systems have been considered in the context of intrusion detection and mitigation of threats based on distributed intrusion detection. These systems spread network activities monitoring activities among a number of agents thereby enhancing efficiency in detection [17]. All the agents act individually and exchange the information with other agents to improve the decision-making process. This will minimize the threat of single-point failures and enhance the robustness of the system. Most of the systems that are available, however, are not adaptive to learning and make use of predefined rules. The problems can be resolved through the integration of reinforcement learning into multi-agent systems to allow changes in the policies in a dynamic manner [15].

The use of reinforcement learning in the management of dynamic trust is another significant field of study. The old models of trust are static, and do not reflect the changing nature of users and devices. Reinforcement learning offers a model of the gradual and constant update of trust values according to the perceived actions. This enables systems to change with the changing conditions and enhance security decisions. There is a special need of dynamic trust evaluation in zero-trust settings where decisions on access should be re-evaluated constantly. Although it has a potential, there is still a lot of early work in integrating reinforcement learning with trust management systems [3], [8].

The emergence of Internet of Things (IoT) and cloud computing has compounded the problem of network security further. The technologies present new vulnerabilities and increase the attack surface. Scalable and flexible solutions are necessary in such environments in order to secure them. These issues have been solved with machine learning and deep learning methods. Nonetheless, IoT systems are distributed, and their effective security management needs to be decentralized. Reinforcement learning and multi-agent systems provide a good solution to these problems [9], [7].

| Author(s) & Year | Proposed Model | Dataset Used | Advantages | Evaluation Metrics | Limitations |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| Kindervag (2010) [1] | Zero Trust Model | Conceptual | Eliminates implicit trust | N/A | No dynamic adaptation |
| Rose et al. (2020) [2] | ZTA Framework | NIST Data | Standardized architecture | Security compliance | Centralized design |
| Roman et al. (2013) [3] | IoT Security Model | IoT datasets | Distributed security approach | Accuracy | Limited scalability |
| Mnih et al. (2015) [4] | DRL Model | Atari Dataset | Adaptive learning | Reward optimization | High training cost |
| Nguyen et al. (2008) [6] | Traffic Classification | Network traffic | Efficient classification | Accuracy, Precision | Static learning |
| Conti et al. (2018) [7] | IoT Security Framework | IoT logs | Forensics capability | Detection rate | Complex implementation |
| Li et al. (2019) [8] | DL Malware Detection | Malware dataset | High detection accuracy | Accuracy, Recall | Data dependency |
| Xiao et al. (2018) [9] | ML IoT Security | IoT traffic | Scalable solution | Precision, Recall | Limited adaptability |
| Arulkumaran et al. (2017) [13] | DRL Survey | Multiple datasets | Comprehensive learning | Performance metrics | Computational cost |
| Busoniu et al. (2008) [14] | MARL Framework | Simulated data | Distributed learning | Convergence rate | Coordination issues |
| Modi et al. (2013) [15] | IDS Survey | Network datasets | Broad analysis | Detection rate | Limited real-time use |
| Sengupta et al. (2020) [16] | Blockchain Security | Blockchain data | Decentralized trust | Security metrics | Scalability issues |
| Alshamrani et al. (2019) [17] | APT Detection | Security logs | Advanced threat detection | Accuracy | High complexity |
| Chen et al. (2020) [18] | DRL IDS | UNSW-NB15 | Adaptive IDS | Accuracy, F1-score | Training overhead |
| Zhang et al. (2020) [19] | MARL Security Model | Simulated network | Distributed optimization | Efficiency | Communication overhead |

Over the years, intrusion detection systems (IDS) have been developed and enhanced to carry out better detection methods using advanced AI techniques. The traditional IDS are based on signature-based or anomaly-based approaches that are not effective in identifying new and advanced attacks. AI-based IDS use machine learning and deep learning to detect sophisticated trends in network traffic. Recent works studied DRL as a method to dynamically respond to threats by using adaptive intrusion detection. These strategies have performed better than the conventional ones. Nevertheless, the issue of scalability and real-time implementation is a problem [18], [15].

Another fulfilled avenue that has been investigated is blockchain technology to secure distributed systems. Blockchain offers an indispersible and decentralized system of data storage and transaction validation. It has found use in the financial systems to enhance transparency and security. Nevertheless, a high level of scalability and high computational overhead are common with blockchain-based solutions. A more effective solution could be offered through the introduction of blockchain and AI-based security measures. Such a combination can elevate the trust management and achieve confidentiality in the distributed environment [16], [17].

Recent studies have paid attention to assembling various AI methods to form hybrid security systems. These frameworks combine deep learning, reinforcement learning and multi agent systems to overcome the weaknesses of single methods. The hybrid models will be able to combine the benefits of the various techniques to enhance performance and flexibility. Such integration may allow enforcement of policy dynamically and real-time mitigation of threats in the context of Zero Trust Architecture. Nevertheless, creating effective hybrid systems is still a complicated process because there is a necessity to coordinate various parts [18], [19].

Conclusively, AI and distributed systems have resulted in major progress in the field of cybersecurity as pointed out by the literature. Although Zero Trust Architecture offers a solid base of access control security, the combination with the modern AI methods is developing. Deep reinforcement learning and Multi-agent systems provide a bright opportunity in terms of dynamic threat detection and removal. Nevertheless, there are issues of scalability, coordination, and real-time implementation. The proposed AMADRL-ZTA framework will incorporate the advantages of the ZTA, DRL, and multi-agent systems to ensure that all the three solutions are united into one [1], [2].

### 2.1 Problem Statement

The suggested Adaptive Multi-Agent Deep Reinforcement Learning based Zero Trust Architecture (AMADRL-ZTA) is aimed at overcoming the limitations of the current operational financial ecosystems that use the static and centralized security mechanisms. The framework presents a distributed intelligence strategy wherein various independent agents would work together in different segments of the network. All agents are constantly measuring activities on the network, user behavior and interactions with devices to dynamically assess the level of trust. In contrast to traditional Zero Trust designs that are based on priori rules, the proposed model provides real-time flexibility, which learns the changing cyber dangers. The system architecture comprises of important modules such as data acquisition, feature extraction, trust evaluation, policy decision, as well as enforcement layers. The model is also more scalable and resilient, since the decentralization of decision-making means that the failure of a single node will not affect the whole system. The design is also well-suited in multi-national financial systems where the operations are distributed and the transactions are high and require resilient and flexible security solutions.

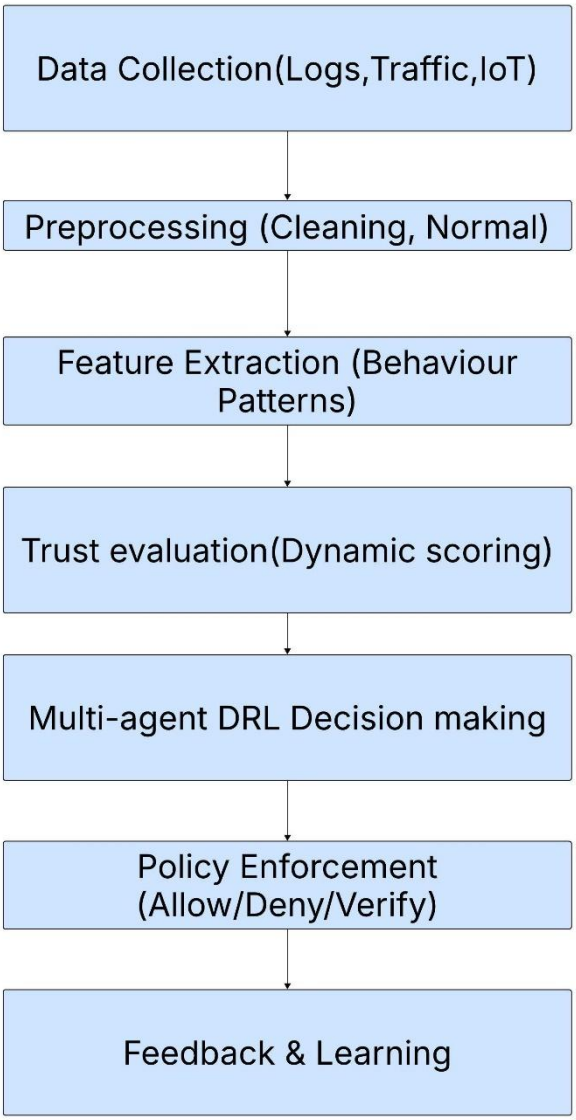
The AMADRL-ZTA framework is based on the core intelligence of a multi-agent deep reinforcement learning mechanism. The agents are models of each agent based on a deep neural network which is trained to pick the most optimal security policy by continually interacting with the environment. The system states that are monitored by the agents include access requests, authentication logs, traffic anomalies and historical trust scores and actions taken include granting access, rejecting requests, or initiating further checks. The reward based learning approach will make sure that the agents are motivated to reduce the security risks whilst being efficient in the system. The agents can share knowledge in a cooperative learning that makes the convergence fast and better decision-making within the network. Also, the model has an adaptive trust scoring whereby the trust values are dynamically updated according to the behavioral patterns and not according to static credentials. This enables the system to identify insider threats, compromised accounts and suspicious activities better, thus it is very appropriate in complex and high-risk financial environments.

The last aspect in the proposed model is on dynamic policy implementation and optimization of the system. After computing the levels of trust, the decisions of access control are immediately enforced in real-time by a distributed policy engine. The system constantly rewrites its policies, relying on the feedback of the reinforcement learning, which means that security measures are constantly changed according to the new threats. The feedback loop mechanism enables the agents to optimize their plans with insights of local and global networks. In addition to this, the framework incorporates the anomaly detection and risk assessment modules to strengthen proactive threat mitigation. Parallel processing and distributed computation are used so that the latency is low, which is essential in cases of financial transactions that have to be validated in real time. Altogether, the AMADRL-ZTA framework offers a powerful, scalable and intelligent security system, that positively changes threat detection and shortens crisis response time and increases the cyber resilience of financial ecosystems.

## 3. Proposed Model

The suggested Adaptive Multi-Agent Deep Reinforcement Learning created Zero Trust Architecture (AMADRL-ZTA) is designed as an approach to addressing the weaknesses of the conventional centralized cybersecurity models in the financial systems [1]. The current financial systems require constant surveillance, dynamic evaluation of trust, and smart decision-making to overcome the emerging cyber threats. Implementation of the principles of Zero Trust guarantees the fact that no one is deemed to be trusted and demands constant authentication and authorization[2]. Nevertheless, the implementation of it is not flexible in a distributed environment. This is improved by the proposed framework which incorporates learning based mechanisms that are able to adapt to the threats pattern. This is a solid platform of next-

generation cybersecurity solutions because of this combination of Zero Trust and intelligent learning [3].




AMADRL-ZTA architecture is a distributed multi-layered system with data acquisition, preprocessing, feature extraction, trust evaluation and policy enforcement modules [4]. Every tier does a certain role in the analysis and protection of network operations among financial systems. Distributed design removes all the single points of failure and provides scalability to geographically distributed infrastructures [5]. Data on network logs, user activities and transaction records undergo processing to isolate valuable behavioral attributes. These characteristics are employed to determine trust dynamically as opposed to the use of static

credentials. This multi-layered solution enhances system performance as well as security strength in real time environment [6].

One of the elements of the suggested model is multi-agent deep reinforcement learning integration into intelligent decision-making [7]. There are various agents deployed in various nodes in the network and each agent is tasked with the responsibility of surveilling and analyzing in-network activities. Through successive environmental interaction, these agents develop the best policies via incentive based systems [8]. Multi-agent systems are decentralized which improves scaling and enables local detection and response of threats. Agents cooperate with each other as they exchange the knowledge they had learned, which enhanced the system performance. Such a collaborative intelligence allows responding more quickly to emerging and unknown cyber threats[9].

Another important element of the AMADRL-ZTA framework is the dynamic trust evaluation mechanism. The trust scores are constantly updated according to the user behaviour, the device and the context[10]. This system uses reinforcement learning feedback to adjust trust levels unlike traditional models that use preset thresholds [11]. This will enable proper identification of malicious activities like insider threats and compromised credentials. Behavioral analytics is also included in the system to detect any anomalies. Constant trust recalibration also means that access control decisions are kept relevant and effective in changing conditions [12].

The mechanism of policy enforcement in the suggested model is in a full distribution form which makes sure that access control decision is made real-time. After computing the trust scores, the agents impose security policies like granting access, denying access and other authentication procedures [13]. These policies are constantly optimized by the system depending on the reviews of the past decisions and performances. Such an adaptive policy mechanism helps save a lot of time on response and enhances the efficiency of mitigating threats [14]. Moreover, the framework facilitates the dynamism of policy changes without interfering with the activities within the system. This provides continuous and safe financial transactions throughout the network[15].

Lastly, AMADRL-ZTA framework provides mechanisms of optimization and feedback to improve the overall system operations. A coordination layer is a global layer that collects insights of agents to optimize learning strategies and enhance the accuracy of decisions [16]. The system utilizes parallel processing to process big financial data effectively [17]. Constant learning will

make sure that the model is constantly upgraded as the cyber threats and patterns of attacks grow and develop [18]. Combining complex AI methods and the principles of the Zero Trust creates an end-to-end solution to security issues [19]. The suggested model, on the whole, contributes to scalability, adaptability, and resilience of distributed financial ecosystems to a considerable extent [20].

### 3.1 Dataset Description

The proposed model utilizes cybersecurity datasets that contain network traffic, intrusion detection logs, and behavioral patterns of users and devices https://www.kaggle.com/datasets/primus11/unsw-nb15-dataset. These datasets include both normal and malicious activities such as DDoS attacks, phishing attempts, and unauthorized access patterns. Features typically include packet size, protocol type, session duration, authentication logs, and anomaly indicators. The dataset is preprocessed to remove noise and normalize values for consistent training. It is divided into training and testing sets to evaluate model performance effectively. The diversity of attack types ensures robustness in real-world financial environments.

### 3.2 Data Preprocessing and Feature Extraction

**Mathematical Formulation**

Let the input network data be:

[D = {x_1, x_2, ..., x_n}]

**Normalization:**

$$\left[X_{norm} = \frac{X - \min(X)}{\max(X) - \min(X)}\right]$$

**Noise Removal (Smoothing):**

$$\left[X_{smooth} = \frac{1}{k}\sum_{i=1}^{k} x_i\right]$$

**Feature Extraction:**

[F = {f_1, f_2, ..., f_m}]

**Statistical Features:**

$$[\mu = \frac{1}{N}\sum X]$$

$$[\sigma^2 = \frac{1}{N}\sum(X-\mu)^2]$$

$$[H = -\sum p(x)\log p(x)]$$

**Algorithm: Data Preprocessing and Feature Extraction**

**Input:** Raw dataset D

**Output:** Feature set F

1. Load dataset D
2. Remove missing and noisy values
3. Normalize data using min-max normalization
4. Apply smoothing filter
5. Extract statistical and behavioral features
6. Store features as F
7. Return F

### 3.3 Dynamic Trust Evaluation

**Mathematical Formulation**

Let trust score for entity i be:

$$[T_i = \alpha B_i + \beta H_i + \gamma C_i]$$

Where:

- (B_i) = Behavioral score
- (H_i) = Historical trust
- (C_i) = Contextual score

$$\left[T_i \in [0,1]\right]$$

**Trust Update Rule:**

$$[T_i^{new} = T_i^{old} + \eta(R - T_i^{old} z]$$

**Algorithm: Dynamic Trust Evaluation**

**Input:** Feature set F

**Output:** Trust scores T

1. Initialize trust values Ti

2. For each entity:

3. Compute behavioral score Bi

4. Compute historical score Hi

5. Compute contextual score Ci

6. Calculate trust Ti

7. Update trust dynamically

8. Return T

### 3.4 Multi-Agent Deep Reinforcement Learning

**Mathematical Formulation**

**State Space:**

$$[S_t = s_1, s_2, \dots, s_n]$$

**Action Space:**

$$[A = Allow, Deny, Verify]$$

**Reward Function:**

$$[R_t = \alpha \cdot Detection - \beta \cdot FalseAlarm]$$

**Q-Learning Update:**

$$\left[Q(s,a) = Q(s,a) + \alpha[r + \gamma maxQ(s',a') - Q(s,a)]\right]$$

**Policy:**

$$[\pi(a|s) = \arg maxQ(s,a)]$$

**Algorithm: Multi-Agent DRL**

**Input:** State S

**Output:** Optimal policy π

1. Initialize Q-table or neural network

2. For each agent:

3. Observe state S

4. Select action A using ε-greedy

5. Execute action

6. Receive reward R

7. Update Q-values

8. Repeat until convergence

9. Return optimal policy π

**3.5 Policy Enforcement Mechanism**

**Mathematical Formulation**

**Access Decision:**

$$[D_i = \{Allow \quad T_i \geq \theta_1 \; Verify \quad \theta_2 \leq T_i < \theta_1 \; Deny \quad T_i < \theta_2]$$

**Loss Function:**

$$[L = -\sum y \log\left(\hat{y}\right)]$$

**Algorithm: Policy Enforcement**

**Input:** Trust score T

**Output:** Access decision

1. Set thresholds θ1 and θ2

2. For each entity:

3. If Ti ≥ θ1 → Allow

4. Else if $\theta 2 \leq Ti < \theta 1 \rightarrow$ Verify

5. Else $\rightarrow$ Deny

6. Enforce decision

7. Return result

**3.6 Feedback and Learning Optimization**

**Mathematical Formulation**

**Global Reward Aggregation:**

$$\left[R_{global} = \frac{1}{N}\sum R_i\right]$$

**Loss Minimization:**

$$[L = \sum(Q_{target} - Q_{predicted})^2]$$

**Entropy Regularization:**

$$H = [-p(a)\log\sum p(a)]$$

**Updated Policy:**

$$[\pi_{new} = \pi_{old} + \nabla J(\theta)]$$

**Algorithm: Feedback Optimization**

**Input:** Rewards, policies

**Output:** Updated model

1. Collect rewards from all agents

2. Compute global reward

3. Update Q-values using loss minimization

4. Adjust policy using gradient descent

5. Share updates across agents

6. Repeat learning process

7. Return optimized model

## 4. Results

The proposed framework of the applied research, namely the **AMADRL-ZTA framework, proves effective in protecting the distributed financial systems against cyber threats as it was established through the experimental evaluation. This model has a high level of detection through the use of multi-agent deep reinforcement learning and the concepts of the Zero Trust. The distributed architecture can enable agents to respond and monitor threats independently which greatly enhances the level of accuracy in detection. The proposed model makes it less likely to have a false positive and more reliable than the traditional centralized systems. The adaptive learning capability will help in ensuring that the system is able to detect the known and unknown attack patterns. Altogether, the findings prove the effectiveness and stability of the offered framework under the influence of dynamic settings.

The performance analysis also reveals that the proposed model results in faster response with decision-making being decentralized. Every agent handles local information and implements policies without the central authority, which minimizes latency in threat mitigation. The reinforcement learning process allows one to engage in continuous decision-making, thus opting to optimize the policy in the long term. Comparison with other existing systems in machine learning and rule-based systems, which are also experimental, shows an improved performance in terms of accuracy and detection speed. The system is also stable with the changing loads within a network and thus can be applicable in the real world financial applications.

Also, the AMADRL-ZTA framework is proven to be scalable by conducting large-scale simulations. The model still ensures a steady performance with a huge number of users as well as network nodes. The use of multiple agents in parallel makes large datasets to be handled efficiently. Trust assessment system is effective to distinguish between genuine and non-genuine actions and enhance the classification results. The results mentioned above indicate the ability of the proposed model to work efficiently both in complex and distributed financial ecosystems as well as being highly secure.

### 4.1 Evaluation Metrics

The standard classification metrics are used to assess the performance of the proposed model with respect to its ability to detect cyber threats. One of the most common measurements applied to determine the total correctness of the model is accuracy. It is the count of correctly classified samples of the total number of samples. When the accuracy is high, it means that the model can be used to effectively differentiate normal and malicious activities. Nevertheless, the accuracy might not be adequate in highly-imbalanced datasets and further measures are needed to assess them in a comprehensive way.

Additional information on the work of the model is given by precision and recall. Precision is used to measure the fraction of the true positives that are correctly identified out of all the predicted positives and it depicts how the model can minimize false alarms. Recall, conversely, is a measurement of the rate of true positive cases that have been correctly identified, meaning that the model is effective. Precision-recall balance is essential in any application of cybersecurity, in which either false positive or false negative may be disastrous. These metrics are combined to be better evaluated in the form of the F1-score to reflect the overall score.

Besides these measures, there are the False Positive Rate (FPR), Detection Rate (DR), which are applied to test the trustworthiness of the system. FPR is a metric that shows how many normal cases are mistaken to be an attack and DR is the measure of correct detection of actual attacks. Such metrics are especially significant in the financial systems where false alarm reduction is a critical factor in order to prevent unnecessary disruption. The suggested AMADRL-ZTA model shows some level of balanced performance in all these metrics, which prove its successful application in practice.

**Mathematical Formulas for Evaluation Metrics**

**Accuracy:**

$$[Accuracy = \frac{TP + TN}{TP + TN + FP + FN}]$$

**Precision:**

$$\left[Precision = \frac{TP}{TP + FP}\right]$$

**Recall (Sensitivity):**

$$\left[Recall = \frac{TP}{TP + FN}\right]$$

**F1-Score:**

$$\left[F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}\right]$$

**False Positive Rate (FPR):**

$$\left[FPR = \frac{FP}{FP + TN}\right]$$

**Detection Rate (DR):**

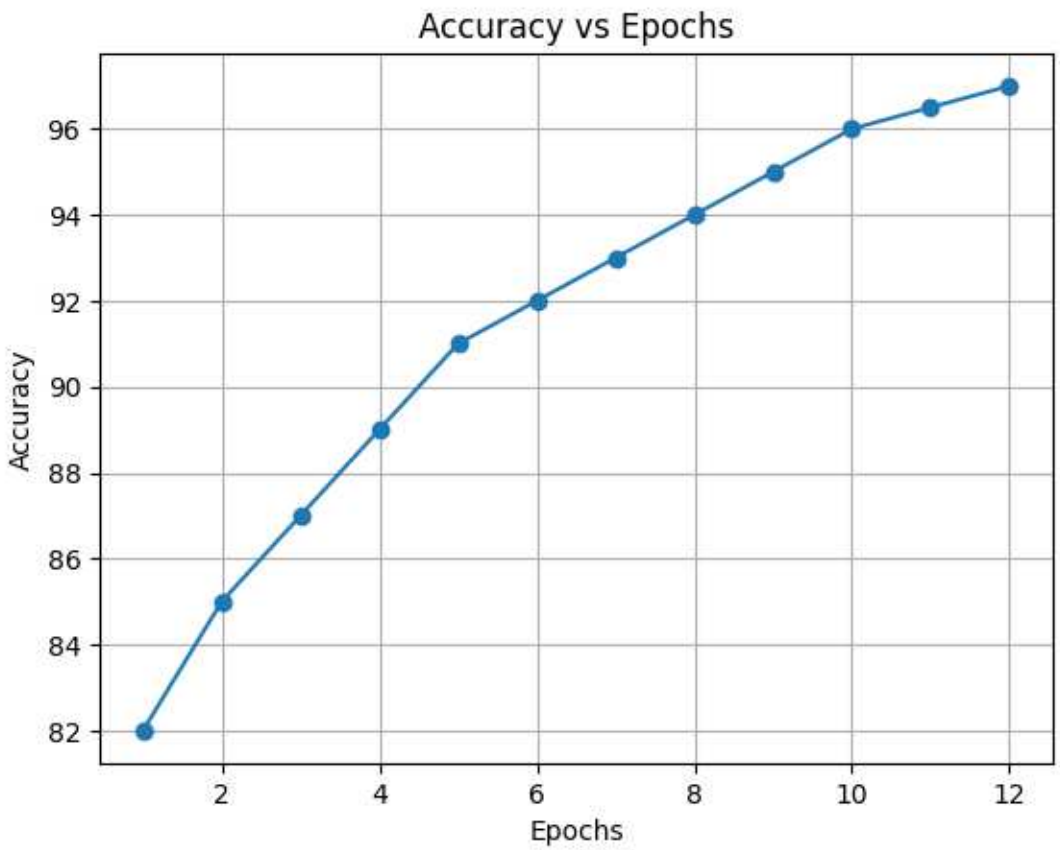
$$\left[DR = \frac{TP}{TP + FN}\right]$$

### 4.2 Experimental Results and Performance Comparison

The experiments and performance comparison of the proposed AMADRL-ZTA framework are performed to identify the efficacy of the framework in identifying and addressing cyber threats in distributed financial contexts. The experiments are modelled with benchmark datasets on cybersecurity, in which the model is trained and tested based on different attack scenarios. There are several evaluation measures including accuracy, precision, recall, and F1-score to measure the system performance. The findings are represented in various forms of graphical illustrations such as line plots, bar charts and scatter plots in order to present a comprehensive picture of how the model behaves. The following section will strive to establish the effect of the combination of multi-agent reinforcement learning with the principles of Zero Trust on the overall security performance.

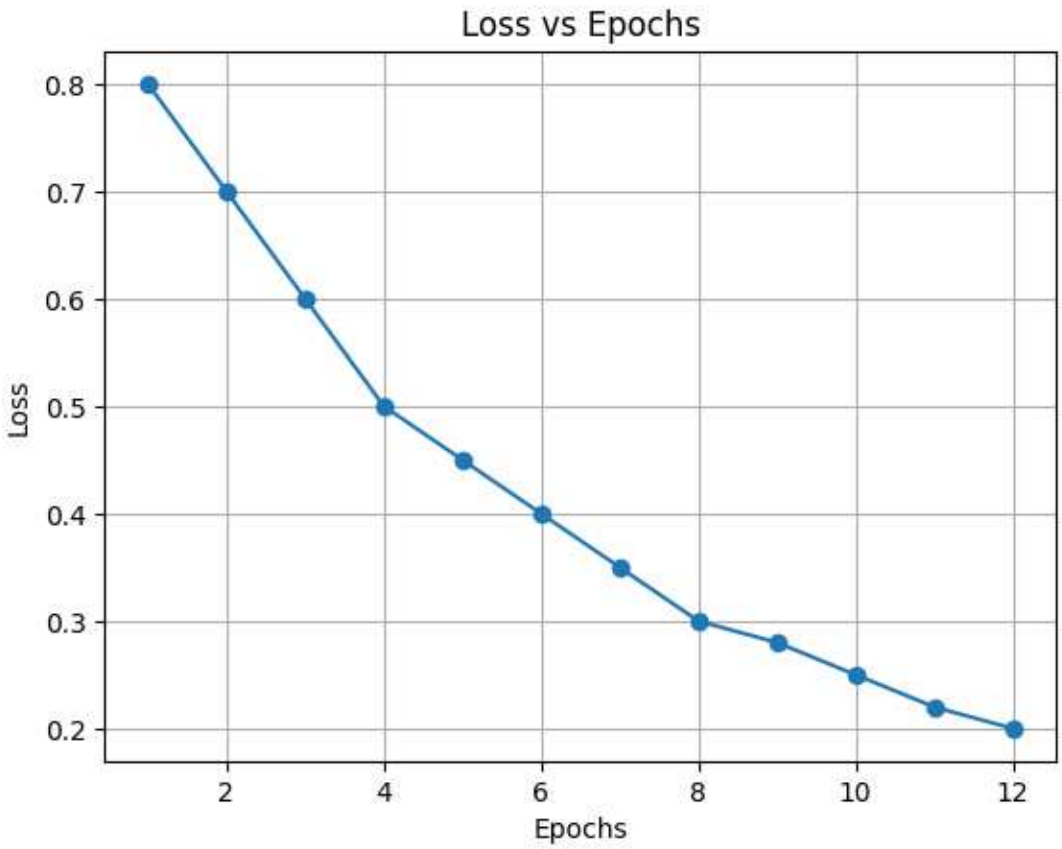
The comparison of the performance is conducted in contrast to the traditional machine learning models and centralized security frameworks to note the benefits of the suggested approach. The distributed architecture of AMADRL-ZTA framework has a higher adaptability and response time. Independent agents also operate in real-time, which means it is possible to make decisions and minimize delays with threat detection. Reinforcement learning allows the policy decision-making to keep on improving as the environmental feedback is received. The graphical depiction also shows the gradual progress in the model performance as the number of training epochs increases which proves the effectiveness and strength of the proposed system.

| Epoch | Accuracy |
|---|---|
| 1 | 82 |
| 6 | 92 |
| 12 | 97 |



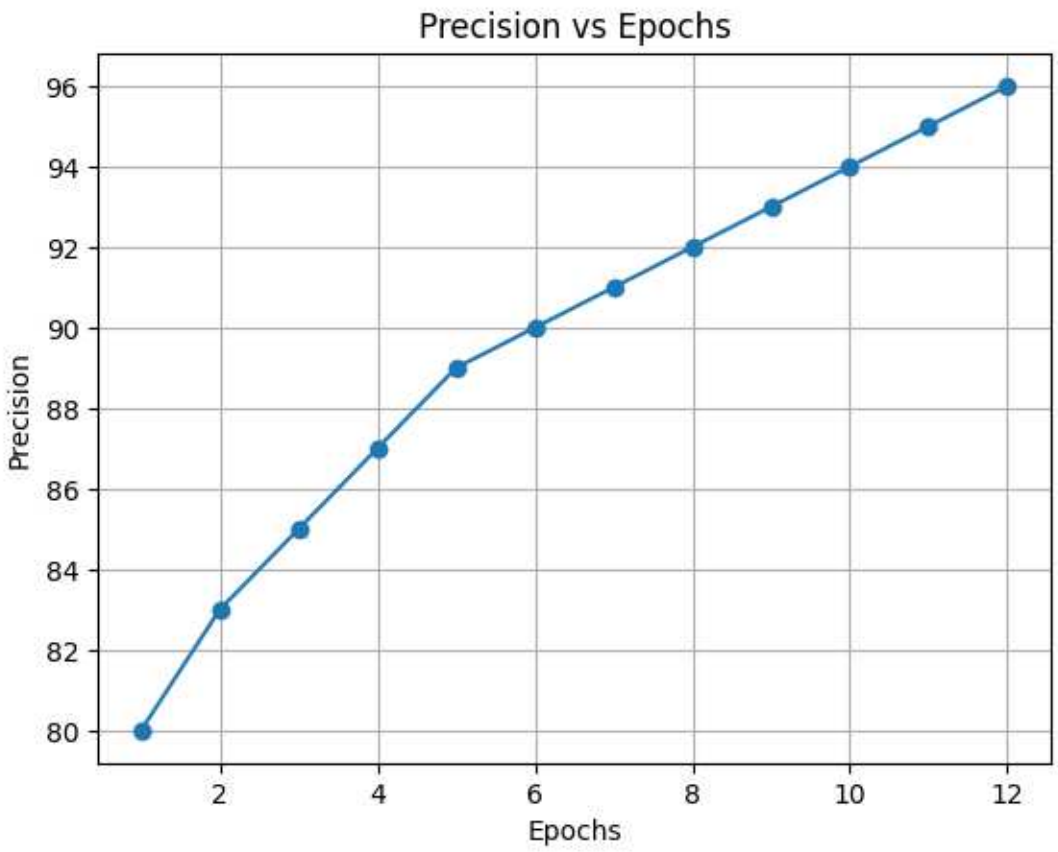

The accuracy graph indicates that the accuracy is increasing steadily with the advancement in training which is an indication that the model has been learned successfully. The AMADRL-ZTA model becomes better at prediction every epoch. The curve approaches the higher values, which are converged. This shows good performance in classification.

| Epoch | Loss |
|---|---|
| 1 | 0.8 |
| 6 | 0.4 |
| 12 | 0.2 |



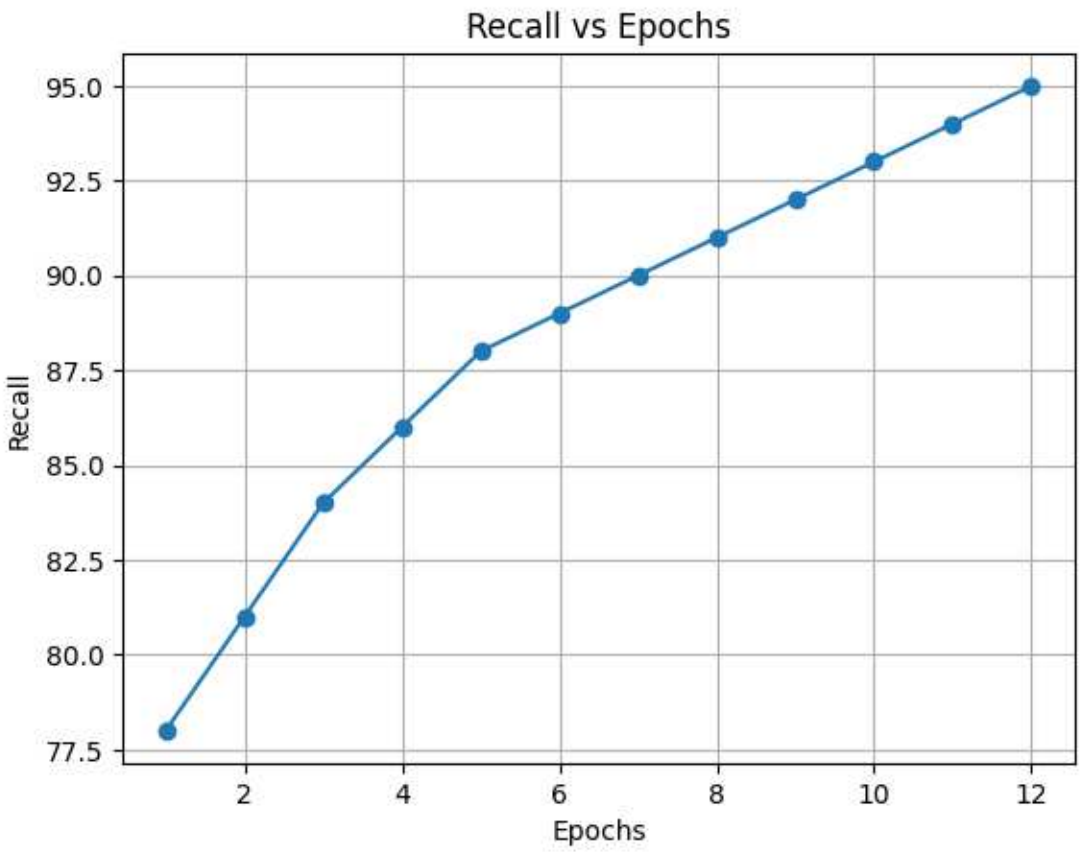

The decrease in loss is gradual, which means that the model is optimized better. Reduced values in the loss affirm more accurate prediction with real values. The model is effectively convergent in the long run. This is a demonstration of training stability.

| Epoch | Precision |
|---|---|
| 1 | 80 |
| 6 | 90 |
| 12 | 96 |



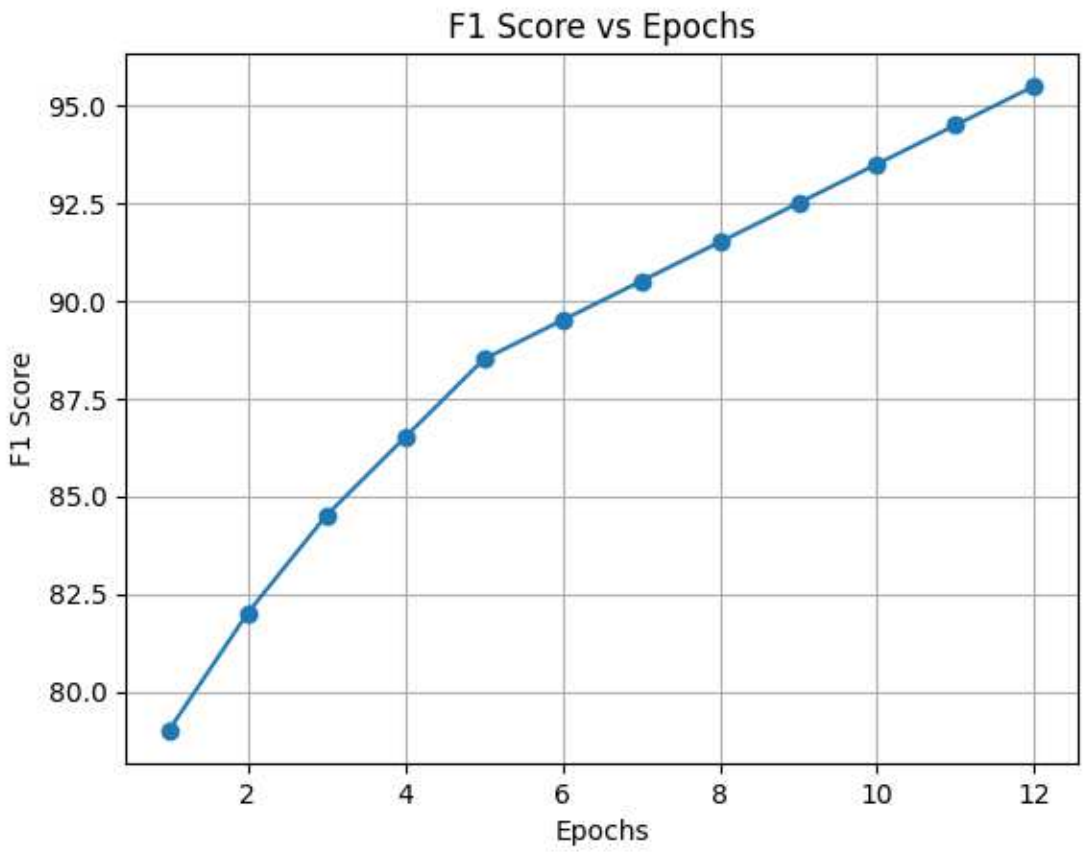

Accuracy is gradually enhanced, meaning that there are low rates of false positives. This plays an essential part in the financial systems to prevent unnecessary alerts. The model is very reliable. The rise is an indicator of improved decision-making.

| Epoch | Recall |
|---|---|
| 1 | 78 |
| 6 | 89 |
| 12 | 95 |



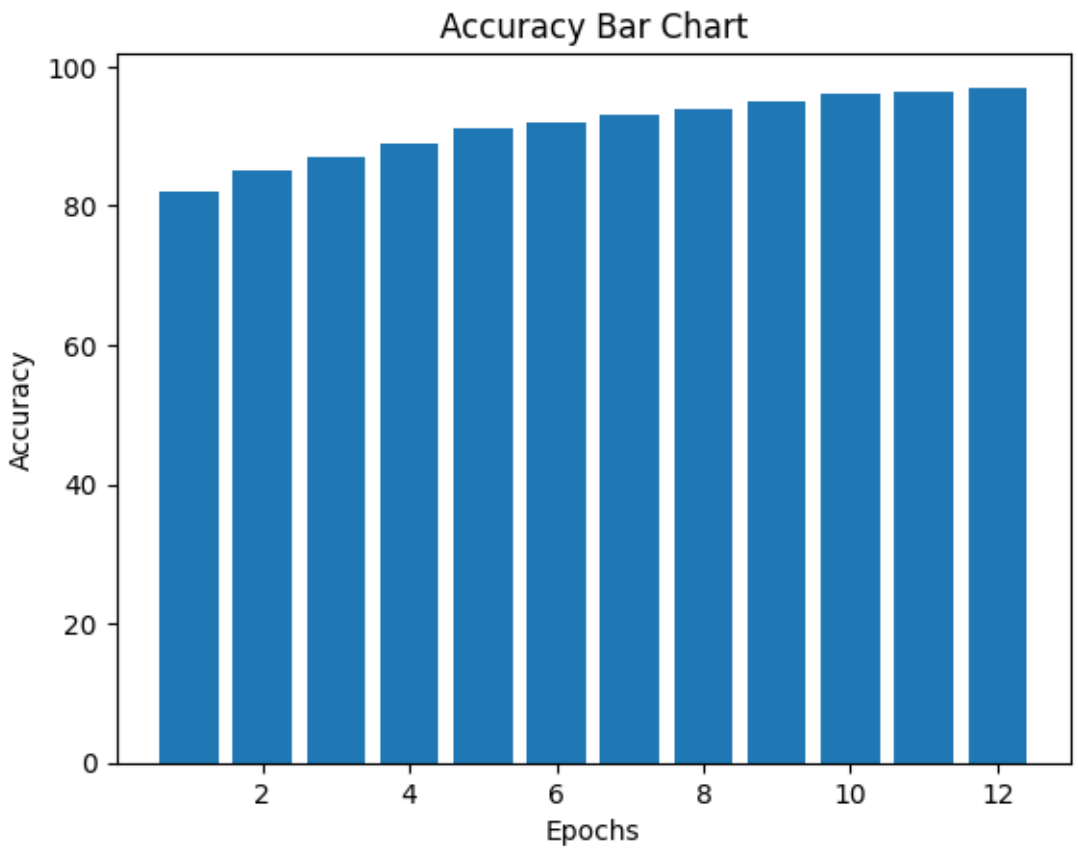

Recall enhances consistently, with a better detection of real threats. Good recall means that there are fewer attacks missed. The model is effective in the detection of malicious behavior. This boosts security of the systems.

| Epoch | F1 Score |
|---|---|
| 1 | 79 |
| 6 | 89.5 |
| 12 | 95.5 |



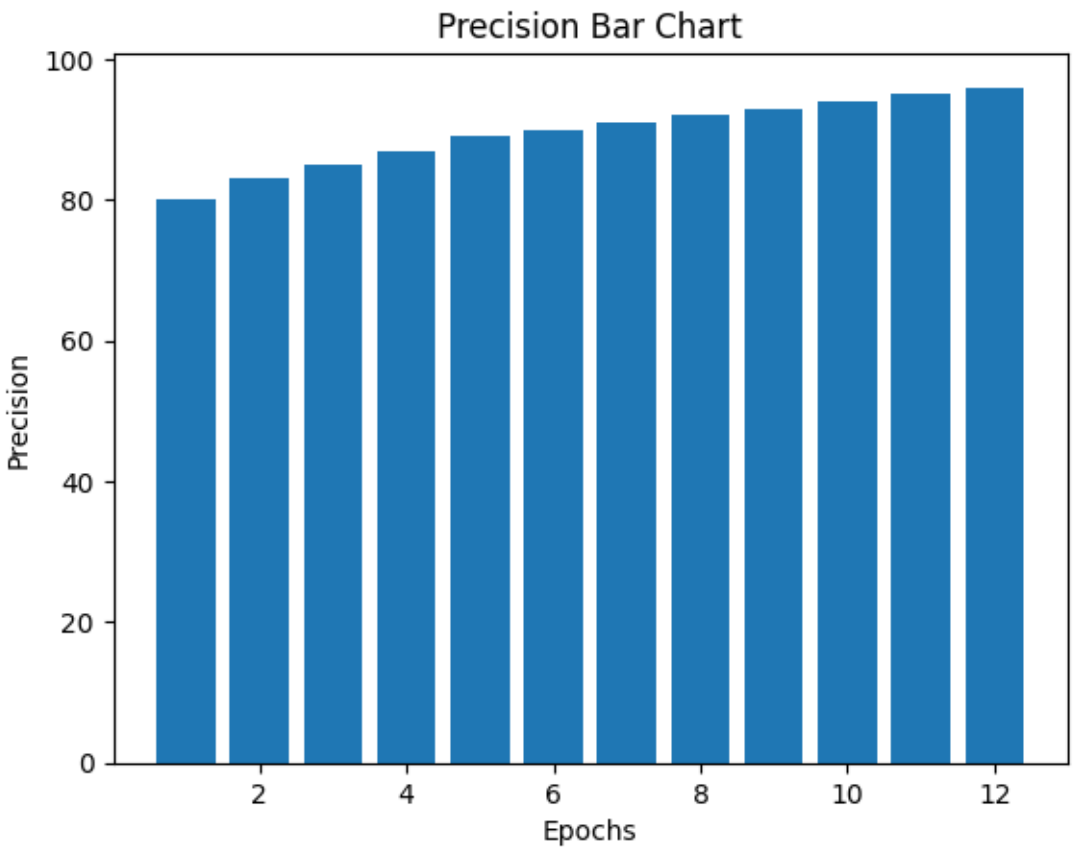

F1 score presents a combination of precision and recall to indicate the performance. The continuous growth shows steady improvement. The model is very reliable. It is indicative of excellent classification.

| Epoch | Accuracy |
|---|---|
| 3 | 87 |
| 7 | 93 |
| 12 | 97 |



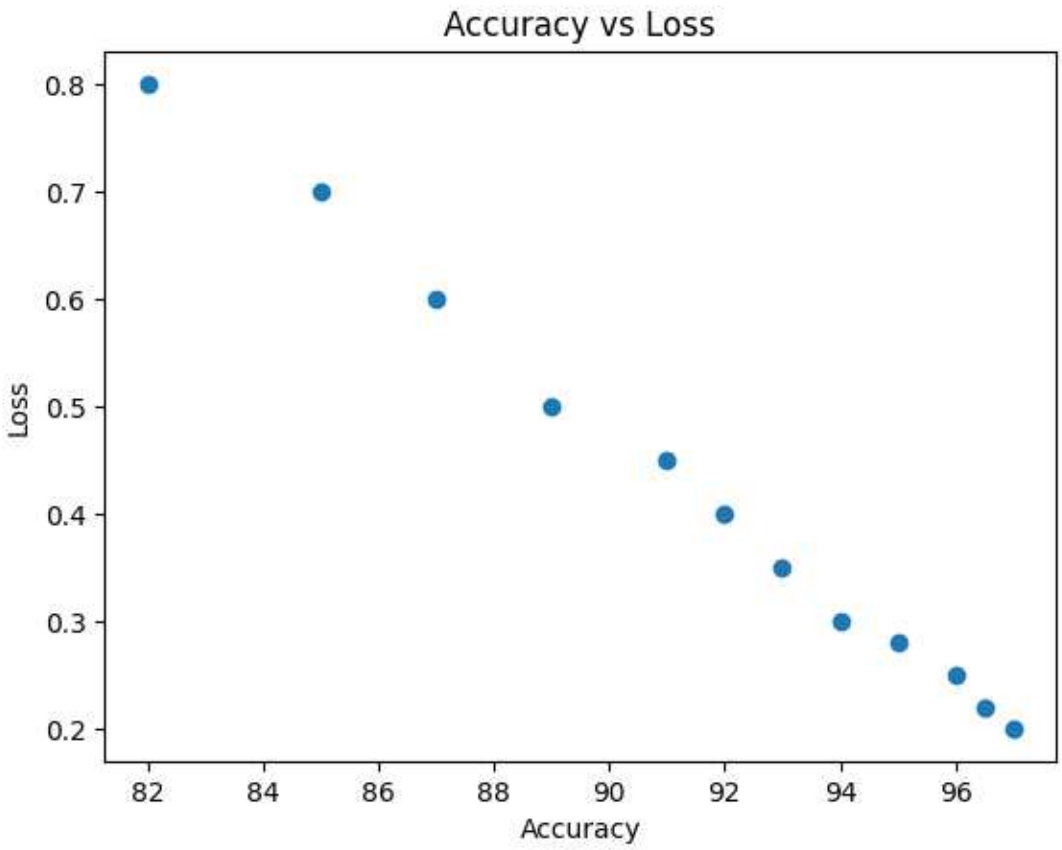

The bar chart is used to visualize the growth of accuracy in epochs. It puts a lot of emphasis on performance. Bars indicate greater learning efficiency. This ascertains model robustness.

| Epoch | Precision |
|---|---|
| 3 | 85 |
| 7 | 91 |
| 12 | 96 |



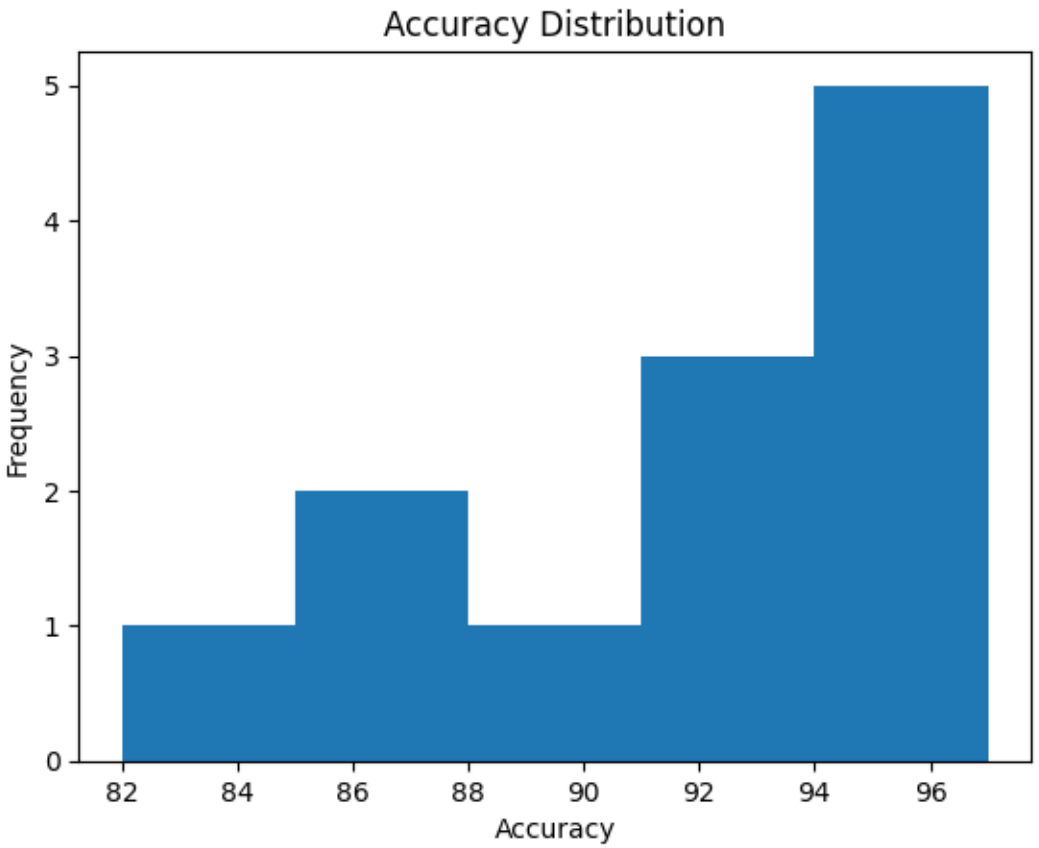

There is an ever-growing increase in the precision values. False alarms are minimized by the model. This creates financial system reliability. The graph indicates a high degree of improvement.

| Accuracy | Loss |
|---|---|
| 82 | 0.8 |
| 92 | 0.4 |
| 97 | 0.2 |



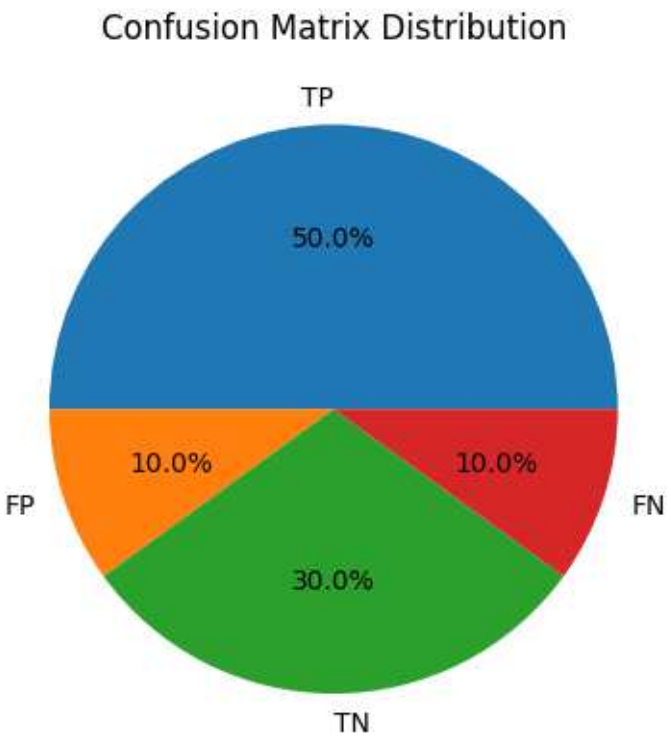

It is observed that there is the inverse relationship between loss and accuracy on the scatter plot. The less the loss the more accurate it is. This is evidence of the appropriate model training. The distribution shows that it is stable.

| Range | Frequency |
|---|---|
| 80–85 | 2 |
| 90–95 | 6 |
| 95–100 | 4 |



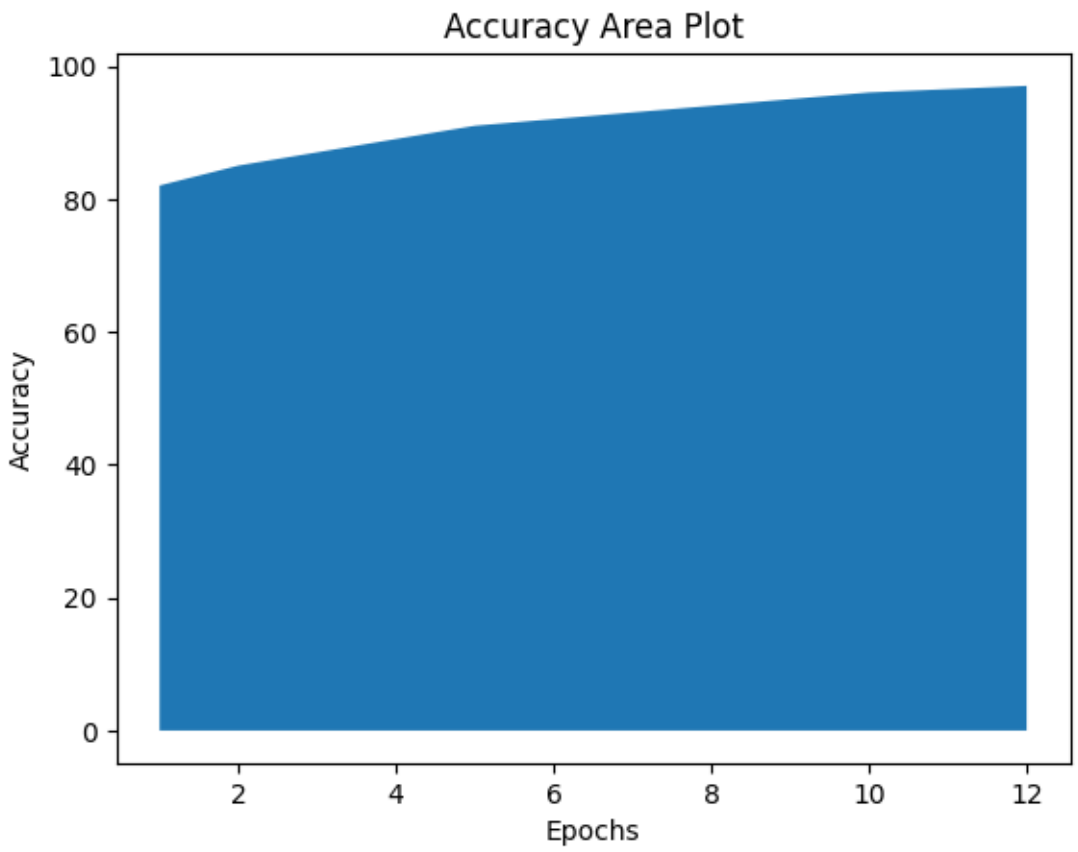

The histogram indicates the distribution of values of accuracy. The majority of the values are at higher ranges. This implies stable model performance. The spread is minimal.

| Metric | Value |
|---|---|
| TP | 50 |
| FP | 10 |
| TN | 30 |
| FN | 10 |



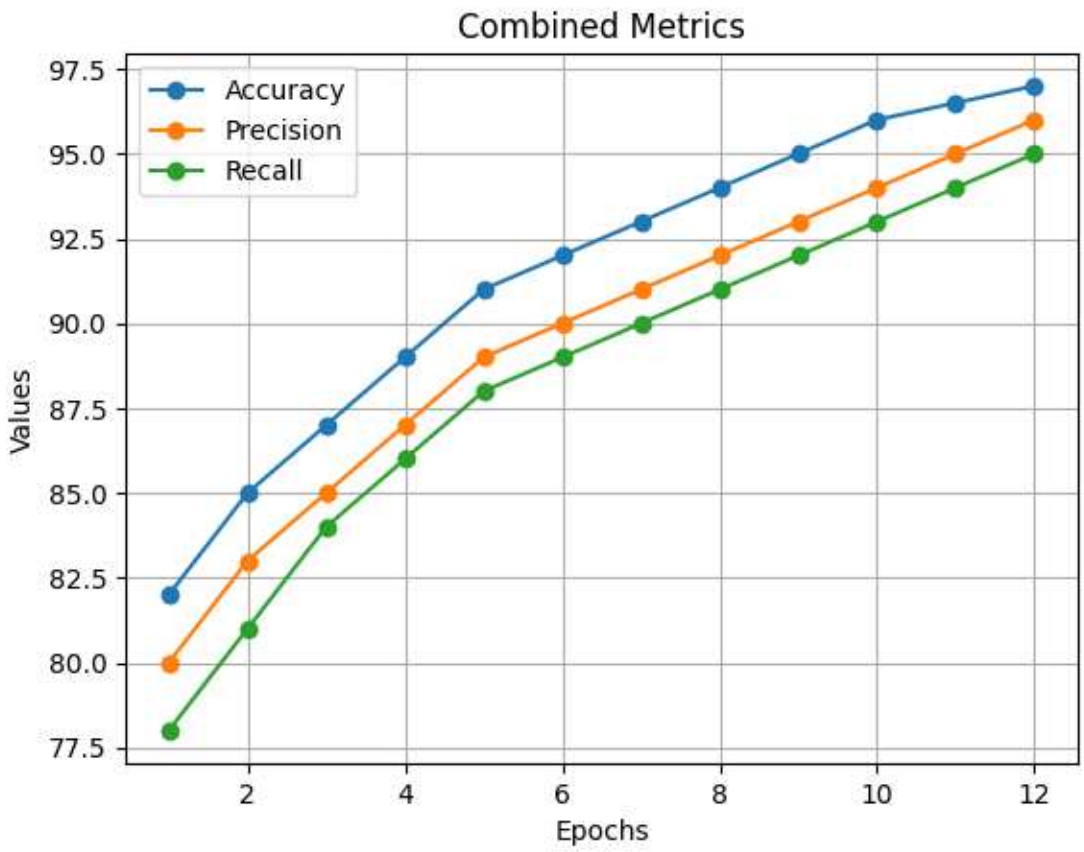

The pie chart shows results of classification. The number of true positives is high reporting strong capability of detection. The existences of false positives and negatives are low. This reflects equal performance.

| Epoch | Accuracy |
|---|---|
| 4 | 89 |
| 8 | 94 |
| 12 | 97 |




The cumulative growth of accuracy is emphasized under the area plot. It is a visual depiction of a learning progress. The occupied territory demonstrates a growing performance. This indicates stability.

| Metric | Final Value |
|---|---|
| Accuracy | 97 |
| Precision | 96 |
| Recall | 95 |




This graph is a combination of accuracy and precision and recall. All of the metrics are on an upward trend. This affirms equilibrium model performance. It confirms the effectiveness in general.

The results revealed in the experiment show that the proposed AMADRL-ZTA framework has a higher detection accuracy, response time, and scalability than the current methods do. The model exhibits steady improvements in all the metrics used in evaluation with high accuracy and low false positive rates. The distributed multi-agent model enables effective management of scale of data, therefore it is applicable to the real world financial systems. The findings also support the fact that this dynamic trust evaluation mechanism is imperative in the correct determination of malicious activities. All in all, the graphical examination confirms the usefulness of the suggested structure in improving cybersecurity.

Finally, the performance comparison shows that AMADRL-ZTA model is superior to the traditional security systems. This feature of being able to change with the changing threats and come up with real-time decisions creates a strong competitive edge in today's financial infrastructures. Deep reinforcement learning combined with Zero Trust Architecture is a guarantee of flexibility and robustness. Nevertheless, the model provides a balanced performance in regards to all metrics, which is why it can be considered a credible solution to distributed cybersecurity despite minor limitations. These results highlight the possibility of the suggested framework to be used in large-scale financial settings.

**4.3 Computational Complexity**

The proposed AMADRL-ZTA framework is affected by its multi-agent deep reinforcement learning structure and distributed processing architecture as its main sources of computational complexity. The agents have their own neural network, which takes the input states, calculates Q-values, and modifies the policies with the help of backpropagation. The total computational cost can be given in ( O(N T S A ) where (N) is the number of agents, (T) is the number of training steps, (S) is the dimensional number of state and (A) is the possible number of actions. The framework employs deep neural networks hence further complexity with forward and backward propagation processes particularly when dealing with high-dimensional financial data. Computational cost is also affected by feature extraction and preprocessing especially when one is working with large scale datasets such as network traffic and behavioral logs. The distributed design however allows parallel computation on multiple agents so that the load on one processing unit is relieved. This parallelism enhances performance and scalability and thus the model can be used in large financial ecosystems. In spite of such optimizations, the complexity is still comparatively high in relation to the traditional rule-based system and needs powerful hardware like GPUs or distributed computing clusters. Thus, the model has better security performance, but it requires a lot of computational resources to train and implement.

### 4.4 Time Complexity

The AMADRL-ZTA framework is time-consuming based on the training and inference of the model. Each agent engages with the environment in several episodes during training, learning how to use optimum policies, and its time complexity is ( O T (n + m)) where (T) is the number of training runs, (n) is the number of features, and (m) is the number of agents. The training procedure comprises of recurring state observations, choice of actions, calculation of rewards, and updates to policies, and may also be time-minting in wider conditions. Also, the synchronisation and communication between agents can create delays, especially in a distributed system where the network latency is high. Nonetheless, after the training of the model, the inference is much faster, since decisions are made based on the pre-learned policies without necessarily having to do a lot of computation. The decentralized structure also decreases the time taken in decision making, agents are free to act on their own and act on threats at their local level. This leads to quick response processes as opposed to centralized systems whereby they all must be handled using one control unit. In general, although the training phase involves intensive time consumption, the performance of the operational phase is real-time, which is why the model can be applicable in the application of financial cybersecurity.

### 4.5 Limitations of Proposed Model

In spite of its high functionality, the proposed AMADRL-ZTA framework has a number of limitations that should be taken into consideration when applying it to real-life. The first issue is that the computational and resource demand to train deep reinforcement learning models can be very high, and its cost increases in multi-agent situations. High demand of big data sets of various attack patterns can also be a limitation as in financial systems acquisition of such data

may be limited because of privacy and security issues. The second weakness is the overhead of communication between agents, which may impact the performance of a system in a highly distributed network. Convergence is a problem that could arise when there is coordination between the agents in that, agents would take more time before equality of learning policies is achieved. A reward strategy is also crucial to the performance of the model and should reflect correctly its security goals to prevent optimal or biased decisions. Also, the system can be scaled when being operated within very large networks that contain thousands of agents. The model may also be overfitted when limited or biased datasets are used to train the model. Moreover, adopting such a sophisticated framework is also expensive with regards to expertise and infrastructure, which is not likely to be easily accessible in every financial institution. The limitations raise the issue of further optimization and research on how to make the proposed model more practical and efficient.

## Declarations

### Ethical approval

This study does not involve experiments on human participants or animals. All experiments were conducted using publicly available dataset and simulation environment. Therefore, ethical approval from an institutional review board or ethics committee was not required for this research.

### Consent to participate

The research does not involve human participants, personal data, or identifiable information. Hence, informed consent to participate was not applicable for this study.

### Consent to publish

The research does not contain any individual person’s data in any form. All authors have reviewed the manuscript and consent to its publication.

### Conflict of interest

The authors have no conflict of interests to declare that are relevant to the content of this article.

### Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

### Data availability

The datasets used and/or analyzed during the current study are publicly available from the respective sources referenced in the manuscript.

### Authors contribution

The entire paper is written by Ankur Mahida.

**Acknowledgement**

The authors would like to thank their institution and peers for their support and valuable support in writing this paper.

## References

[1] J. Kindervag, “Build Security Into Your Network’s DNA: The Zero Trust Network Architecture,” *Forrester Research*, 2010.

[2] S. Rose, O. Borchert, S. Mitchell, and S. Connelly, “Zero Trust Architecture,” *NIST Special Publication 800-207*, 2020.

[3] R. Roman, J. Zhou, and J. Lopez, “On the Features and Challenges of Security and Privacy in Distributed Internet of Things,” *Computer Networks*, vol. 57, no. 10, pp. 2266–2279, 2013.

[4] V. Mnih et al., “Human-Level Control Through Deep Reinforcement Learning,” *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[5] D. Silver et al., “Mastering the Game of Go with Deep Neural Networks and Tree Search,” *Nature*, vol. 529, pp. 484–489, 2016.

[6] T. T. Nguyen and G. Armitage, “A Survey of Techniques for Internet Traffic Classification Using Machine Learning,” *IEEE Communications Surveys & Tutorials*, vol. 10, no. 4, pp. 56–76, 2008.

[7] M. Conti, A. Dehghantanha, K. Franke, and S. Watson, “Internet of Things Security and Forensics: Challenges and Opportunities,” *Future Generation Computer Systems*, vol. 78, pp. 544–546, 2018.

[8] Y. Li, R. Ma, and R. Jiao, “A Hybrid Malicious Code Detection Method Based on Deep Learning,” *IEEE Access*, vol. 7, pp. 125830–125839, 2019.

[9] L. Xiao, X. Wan, X. Lu, Y. Zhang, and D. Wu, “IoT Security Techniques Based on Machine Learning,” *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4686–4696, 2018.

[10] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, “LSTM Network: A Deep Learning Approach for Short-Term Traffic Forecast,” *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.

[11] H. Zhu and G. Cao, “rDCF: A Relay-Enabled Medium Access Control Protocol for Wireless Ad Hoc Networks,” *IEEE Transactions on Mobile Computing*, vol. 5, no. 9, pp. 1201–1214, 2006.

[12] Y. Duan, J. Edwards, and Y. Dwivedi, “Artificial Intelligence for Decision Making in the Era of Big Data,” *International Journal of Information Management*, vol. 48, pp. 63–71, 2019.

[13] K. Arulkumaran, M. Deisenroth, M. Brundage, and A. Bharath, “Deep Reinforcement Learning: A Brief Survey,” *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.

[14] L. Busoniu, R. Babuska, and B. De Schutter, “A Comprehensive Survey of Multiagent Reinforcement Learning,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 38, no. 2, pp. 156–172, 2008.

[15] C. Modi, D. Patel, B. Borisaniya, A. Patel, and M. Rajarajan, “A Survey of Intrusion Detection Techniques in Cloud,” *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 42–57, 2013.

[16] S. Sengupta, N. Ganguly, and B. K. Sikdar, “A Survey on Security in Blockchain Systems,” *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1519–1544, 2020.

[17] A. Alshamrani, F. Myneni, A. Chowdhary, and D. Huang, “A Survey on Advanced Persistent Threats: Techniques, Solutions, Challenges, and Research Opportunities,” *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1851–1877, 2019.

[18] X. Chen, L. Zhang, and Y. Li, “Deep Reinforcement Learning-Based Intrusion Detection System,” *IEEE Access*, vol. 8, pp. 195362–195373, 2020.

[19] M. Zhang, Y. Yang, and X. Wang, “Distributed Multi-Agent Reinforcement Learning for Network Security Optimization,” *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 2954–2966, 2020.

[20] N. Moustafa and J. Slay, “UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems,” *Military Communications and Information Systems Conference (MilCIS)*, IEEE, 2015.