Journal of Computer and Knowledge Engineering, Vol. 9, No. 1, 2025. (25-40)

Ferdowsi University of Mashhad

Journal of Computer and Knowledge Engineering

https://cke.um.ac.ir

Information and Communication Technology Association of Iran

Smart Grid Security: Proactive Prediction of Advanced Persistent Threats*

Research Article

Motahareh Dehghan [1], Erfan.Khosravian[2]

10.22067/cke.2025.91408.1141

Abstract *The increasing reliance on Internet of Things devices in smart grids has introduced significant cybersecurity challenges, particularly in the detection and prevention of Advanced Persistent Threats. These threats, characterized by their stealth and persistence, can compromise the integrity and functionality of critical grid infrastructure. This paper proposes the use of Deep Reinforcement Learning to enhance cybersecurity in smart grids by leveraging the ProAPT model, which is specifically designed to predict and mitigate Advanced Persistent Threats. The ProAPT model utilizes a Markov Decision Process to simulate and assess potential threats, dynamically adapting to the evolving security landscape. The model is trained using the CICAPT-IIoT dataset, which includes simulated attack scenarios in industrial IoT networks. The results of our experiments demonstrate the effectiveness of the ProAPT model in detecting and preventing APTs in smart grid environments. Experimental results show that the ProAPT model significantly outperforms traditional machine learning algorithms like Random Forest, Support Vector Machines, and Logistic Regression, achieving 93.8% accuracy, 93.12% precision, 95.2% recall, and 94.15% F1-Score. The feature importance analysis reveals that traffic-related features such as packet size variance and connection duration are crucial in identifying Advanced Persistent Threats. This paper demonstrates the effectiveness of Deep Reinforcement Learning in enhancing smart grid cybersecurity by proactively identifying and mitigating cyber threats, offering a promising approach to securing IoT-based critical infrastructures against sophisticated cyberattacks.*

Key Words Cyber Security, Smart Grids, Advanced Persistent Threats, Deep Reinforcement Learning, ProAPT Model, Feature Importance.

1. INTRODUCTION

The transformation from traditional power grids to smart grids has revolutionized the energy sector by integrating modern technologies such as IoT devices, sensors, and advanced communication systems. These technologies enable real-time monitoring, automated decision-making, and predictive maintenance, making energy supply more efficient, reliable, and sustainable. In particular, smart grids enable dynamic management of electricity generation, distribution, and consumption, improving energy efficiency and facilitating the integration of renewable energy sources. However, the increasing complexity of smart grids increases their vulnerability to cybersecurity threats. The emergence of IoT in smart grids has significantly increased the number of connected devices and systems, many of which are exposed to external networks or deployed in remote or insecure environments. While these IoT devices are essential to optimizing network operations, attackers can also exploit vulnerabilities in these devices to infiltrate network systems, manipulate operations, or disrupt network operations. These threats are exacerbated by the increasing sophistication and persistence of cyber-attacks targeting critical infrastructure, which can have serious consequences such as system failure, data theft, and even property damage [1].

One of the most concerning types of cyber-attacks related to smart grids is the APT. An APT is a type of advanced, stealthy cyber-attack designed to infiltrate a network and remain undetected for long periods of time. Unlike traditional cyber-attacks, which are often short-lived and detectable by traditional defense mechanisms, APTs are characterized by their multi-stage nature and long-term objectives, making them difficult to identify and contain. These threats are often launched by well-funded and organized attackers, including nation states and cybercrime organizations, who seek to maintain persistent

* Manuscript received 2024 December 29, Revised 2025 March 4, Accepted 2025 June 16.

[1] Corresponding author. Assistant Professor Department of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran. **Email:** m_dehghan@modares.ac.ir

[2] Assistant Professor, Department of Mechanical Engineering, Payame Noor University, Tehran, Iran.

access to critical systems for espionage, sabotage, or data exfiltration. The impact of APTs on the smart grid is potentially devastating [2].

If attackers successfully penetrate the smart grid, they can manipulate operational data, disrupt the flow of electricity, or compromise the security of the entire system. For example, APTs could attack the power grid's control systems, causing power outages and damaging the electrical infrastructure. Furthermore, because the smart grid is decentralized and relies heavily on IoT devices for data collection and decision-making, these attacks are increasingly difficult to detect.

Traditional defense mechanisms such as signature-based IDSs and basic anomaly detection methods are often ineffective against such complex and persistent threats.
The scale and complexity of the smart grid poses unique challenges for cybersecurity. Unlike traditional IT networks, where security measures can be deployed centrally, the smart grid is comprised of numerous interconnected devices, including smart meters, grid sensors, phases of power flow management systems, and actuators. These devices are distributed across vast geographic areas and communicate with each other in real time to ensure efficient network operation. Given this dynamic and decentralized structure, ensuring the security of the smart grid requires not only protecting individual devices, but also ensuring that all components work together securely [3].

Recent research highlights the growing importance of deep learning techniques, particularly Deep Reinforcement Learning [4], in addressing the dynamic and adaptive nature of APTs in smart grids. DRL offers a promising solution for proactive cybersecurity measures by continuously learning from interactions with the environment and adapting strategies accordingly. Studies such as [5] emphasize the role of deep learning in enhancing the resilience of smart grid networks against evolving cyber threats. In addition, Sewak et al. [6] demonstrate the effectiveness of DRL-based models in detecting complex cybersecurity threats, including APTs, by using reward-based learning frameworks. These advancements are particularly relevant for smart grid systems, where traditional cybersecurity measures are increasingly inadequate due to the rapid evolution of attack techniques and the scale of connected devices.

Moreover, recent studies such as [7,8] have proposed robust models integrating machine learning and DRL for detecting and mitigating APTs. They have designed a DRL framework for smart grid cybersecurity, highlighting its ability to adapt to the complex, dynamic nature of cyber-physical attacks. Khan et al. [7] provide an overview of the cyber threats facing modern smart grids and propose advanced machine learning models to counter these challenges. These studies reinforce the need for adaptive and proactive cybersecurity frameworks like the ProAPT model, which utilizes DRL to predict and mitigate APTs before they fully manifest, thus improving the security and reliability of smart grids. In addition, IoT devices often have limited processing power and storage capacity and may not support traditional security measures, further complicating the detection and containment of complex cyber threats. Furthermore, the growing reliance on M2M communications and cloud computing in smart grids increases the attack surface and provides attackers with numerous entry points. This is particularly problematic because attackers may exploit vulnerabilities in the software or hardware of IoT devices, as well as in communication protocols and network interfaces.

As a result, traditional cybersecurity approaches are no longer sufficient to address emerging threats to smart grids. Given the limitations of traditional techniques and the increasing sophistication of cyber-attacks, there is an urgent need for more advanced and adaptive solutions that can effectively detect, predict, and mitigate APTs in smart grids [9].

In this paper, we propose a novel solution to combat cybersecurity threats in smart grids by detecting and mitigating APTs using DRL. DRL is a branch of machine learning in which an agent learns how to make optimal decisions by interacting with the environment and receiving feedback in the form of rewards or penalties [10]. Unlike supervised learning approaches that require labeled data, DRL operates in dynamic environments and is able to continuously learn from new interactions and adapt its strategy accordingly. The proposed solution leverages the ProAPT (Prediction of Advanced Persistent Threats) model [11], which is designed to predict and mitigate APTs using deep reinforcement learning. The central idea behind the ProAPT model is to use a Markov decision process (MDP) to simulate the evolving security state of a smart grid system and determine the optimal action to address potential threats.

In the context of a smart grid, these actions might include triggering security protocols, isolating affected devices, or adjusting network configurations to prevent the attack from spreading. By continually interacting with the grid’s environment and receiving feedback, the model learns how to improve threat detection and mitigation strategies over time, enabling it to identify APTs before they fully manifest. One key innovation of this approach is its ability to proactively predict APTs. Rather than relying on reactive measures such as post-attack detection, the ProAPT model predicts possible future threats based on historical attack data and ongoing grid activity. This proactive approach significantly improves the resilience of the grid, enabling early intervention to prevent severe damage. The model is trained using the CICAPT-IIoT dataset [12], which contains simulated attack scenarios in industrial IoT networks. The ProAPT model is applied to this dataset to evaluate its effectiveness in detecting and mitigating APTs in smart grid environments.

This paper makes the following key contributions to advancing smart grid cybersecurity:

- Novel Application of DRL for APT Prediction: Unlike previous works that rely on traditional machine learning approaches, this study pioneers the use of DRL to predict APTs in smart grids, enabling a more adaptive and proactive defense mechanism.
- Empirical Validation on a Real-World Industrial IoT Dataset: We rigorously evaluate our proposed ProAPT model using the CICAPT-IIoT

dataset, which includes diverse and realistic cyber-attack scenarios specific to critical infrastructure, ensuring practical relevance and generalizability.

- Feature Importance-Driven Model Optimization: Our approach integrates a feature importance analysis to systematically identify and prioritize the most critical features, enhancing model interpretability and efficiency.
- Comprehensive Performance Assessment: Unlike prior studies that focus on limited evaluation metrics, we conduct an extensive performance analysis using accuracy, precision, recall, and F1-score to provide a holistic understanding of the model's effectiveness in detecting and mitigating APTs.

This paper is structured as follows:

Section 2 provides a comprehensive overview of related research in the areas of cybersecurity in smart grids, APT detection, and the application of DRL in cybersecurity. Section 3 presents the methodology detailing the ProAPT model, its adaptation to smart grid cybersecurity, the training process and evaluation using the CICAPT-IIoT dataset. Section 4 describes the experimental setup including the results of applying the ProAPT model and compares its performance with other conventional models in terms of detection accuracy, precision and F1-score. In section 5 feature importance methods are implemented and the best features are stated. Finally in sections 6 and 7 we discuss and conclude the paper with an overview of the contributions and suggestions for future work such as improving scalability and integrating it into existing smart grid security frameworks.

2. RELATED WORK

Smart grid cybersecurity is a critical concern due to the integration of advanced technologies and data-driven systems, which, while enhancing efficiency and sustainability, also introduce vulnerabilities. These vulnerabilities manifest in various forms, such as false data injection attacks, malware, and cyber-physical attacks, posing significant risks to the integrity and reliability of smart grids. Addressing these threats requires a multifaceted approach involving detection, prevention, and mitigation strategies. Machine learning models, such as Extra Tree, Random Forest, and Extreme Gradient Boosting, have shown high accuracy (up to 98%) in detecting these attacks, providing a robust defense mechanism [13].

Cyber-Physical attacks involve manipulating power demands using IoT devices or introducing false sensor readings. A DRL framework has been proposed to counter these attacks by triggering appropriate protection sequences, verified through reachability analysis for safety [14]. The use of SCADA systems in smart grids makes them susceptible to malware, which can exploit IT-OT integration vulnerabilities. The complexity of these systems increases the risk of cyber threats, necessitating enhanced cybersecurity measures [15].

The integration of information and operations technology in smart grids introduces new vulnerabilities, requiring continuous monitoring and updating of security protocols to prevent breaches [16]. Implementing a combination of traditional and advanced security measures is crucial. This includes regular updates, intrusion detection systems, and employee training to recognize and respond to threats [7]. Ongoing research is essential to address emerging threats and develop innovative solutions, such as advanced algorithms for attack detection and mitigation [16]. While smart grids offer numerous benefits, such as improved energy efficiency and integration of renewable sources, they also present unique cybersecurity challenges. The dynamic nature of cyber threats necessitates a proactive and adaptive approach to security, ensuring the resilience and reliability of smart grid infrastructures.

Smart grids are an essential part of modern energy systems, but they are also vulnerable to various cybersecurity threats due to their increasing reliance on digital technologies and interconnected devices. Researchers have proposed several solutions to secure smart grids, which can be broadly categorized into IDS, anomaly detection techniques, and authentication protocols. One of the primary methods used to protect smart grids is the development of IDS, which monitor the network for any signs of unauthorized access or abnormal behavior. IDS in smart grids often rely on signature-based detection, which matches observed network behavior to known attack patterns. However, as smart grid environments evolve, this approach has become less effective due to the increasing sophistication of cyberattacks and the dynamic nature of smart grids. To address this limitation, anomaly detection techniques, such as statistical methods and machine learning, have been integrated into IDS to detect deviations from normal operations that could indicate a security breach [17].

These methods, though effective in detecting new types of attacks, struggle with issues such as false positives and the need for large amounts of labeled data. Detecting and responding to APTs in smart grids presents unique challenges. APTs are characterized by their stealthy, multi-stage nature and ability to remain undetected over long periods. This makes them particularly dangerous in smart grids, where attackers can potentially gain control of critical infrastructure systems without alerting security systems. Additionally, the heterogeneity of smart grid components, the presence of many IoT devices, and the distributed nature of control make it difficult to monitor and secure the entire grid effectively. These challenges require advanced, dynamic methods of detection and response that can adapt to new and evolving attack vectors. Several studies have explored using real-time monitoring and adaptive security models to mitigate these challenges [5].

APTs are one of the most critical cybersecurity concerns for modern infrastructure, including smart grids. Unlike typical cyberattacks, which tend to be short-lived and easily detectable, APTs are long-term attacks that exploit vulnerabilities in a system over an extended period. APTs often involve multiple stages, including initial infiltration, lateral movement within the network, data exfiltration, and

maintaining persistence over time. They are designed to avoid detection and maximize their impact on targeted systems [2].

APTs are usually associated with highly organized threat actors, such as nation-states or cybercriminal groups. These actors have significant resources and expertise, allowing them to plan and execute multi-phase attacks. Key characteristics of APTs include sophistication, long-term persistence, and specific targeting. APT attacks often target high-value assets, including critical infrastructure like power plants, water supplies, and transportation systems, with the goal of gaining unauthorized access, stealing sensitive data, or causing operational disruptions [18].

Some of the most infamous APT attacks targeting critical infrastructure include Stuxnet [19], which specifically targeted Iran's nuclear facilities, and BlackEnergy [20], which affected Ukraine's power grid. These attacks demonstrate the high stakes involved in cybersecurity for critical infrastructure and the potential consequences of a successful APT. Stuxnet, for example, manipulated control systems within the targeted facility, leading to significant physical damage. Traditional methods for detecting APTs include signature-based approaches, which compare network traffic to predefined attack patterns, and statistical methods, which look for anomalies in system behavior that may indicate an attack. However, these approaches often struggle to detect sophisticated, low-and-slow APTs. Recent research has focused on leveraging machine learning techniques to improve APT detection. Models such as random forests, support vector machines (SVM), and deep learning have shown promise in identifying previously unknown attack patterns. Despite this progress, a major challenge remains the lack of labeled data for training models, as APTs are rare and difficult to simulate in a controlled environment [21].

DRL has emerged as a powerful tool for addressing complex decision-making problems in dynamic environments, including cybersecurity. DRL involves training an agent to take actions in an environment to maximize cumulative rewards, making it an ideal approach for security tasks that require continuous adaptation and learning. DRL has shown great potential in the field of cybersecurity due to its ability to adapt to evolving threats and optimize long-term security strategies. DRL-based models have been used for tasks such as intrusion detection, vulnerability scanning, attack detection, and incident response. By continuously learning from the environment and adjusting its actions based on feedback, DRL can provide an adaptive, proactive defense mechanism against cyberattacks, including APTs. For example, DRL has been used to model intrusion detection in IoT networks, where it learns to distinguish between benign and malicious activities based on observed behaviors [22].

One of the main advantages of DRL is its ability to learn optimal decision policies from raw data without relying on hand-crafted rules or predefined attack signatures. This capability is particularly useful in environments like smart grids, where attack patterns are constantly evolving. Moreover, DRL-based models can handle complex, multi-step security tasks that require dynamic adjustments based on the state of the system. For instance, DRL can optimize actions to prevent attacks while minimizing the impact on system performance and resource consumption. Despite its potential, applying DRL to cybersecurity poses several challenges. One of the main challenges is the sample inefficiency of deep reinforcement learning algorithms, where a large number of interactions with the environment are often needed to converge on an optimal policy. Additionally, reward shaping can be difficult, as determining the appropriate rewards for specific security actions in dynamic environments like smart grids is not straightforward. Finally, training DRL models in real-world cybersecurity scenarios often requires access to large amounts of labeled data, which is typically not available for rare events such as APTs [6].

DRL, a promising technique for cybersecurity, enables models to learn optimal responses by interacting with the environment and adapting over time. It has shown significant potential in various fields, including robotics, gaming, and cybersecurity. One notable application is ProAPT [11], which uses DRL to predict the next stages of APTs. The model learns from historical attack data and environmental conditions to anticipate the next steps in an ongoing attack, enabling proactive defense mechanisms.

Recent advances in DRL have led to a surge of research focused on enhancing cybersecurity in smart grids and critical infrastructures. Abdi et al. [5] provided a comprehensive survey on the application of deep learning, particularly DRL, to proactively secure smart grid environments. They emphasized how DRL frameworks can adaptively counter zero-day attacks and sophisticated APTs. Veith et al. [23] explored how DRL agents trained on misuse cases can learn novel attack vectors, representing a significant leap in proactive APT detection. Sinha et al. [24] extended this work by proposing a cyber-resilient demand response system, which not only optimizes grid operations but also integrates DRL for enhanced security against APTs and false data injection. Furthermore, Li et al. [25] introduced a state-adversarial DRL-based scheduler for integrated energy systems that mitigates the effect of data manipulation attacks on demand-response coordination. To support secure communication in grid CPS, Sun et al. [26] proposed a DRL-based multi-agent scheme for secure resource allocation under adversarial conditions.

These contributions collectively reinforce the relevance and applicability of DRL—especially DQN variants—in detecting and mitigating APTs across multiple smart grid environments [27]. While the previous research demonstrate important progress in applying machine learning and deep learning methods to smart grid cybersecurity, several critical gaps remain that hinder their real-world applicability. Most of the existing deep learning models—such as LSTM, CNN, and GRU—operate in a supervised learning setting and rely heavily on large volumes of labeled data. This is a significant limitation in the context of APTs, which are rare, highly complex, and difficult to label accurately due to their stealthy and evolving nature. Moreover, many previous solutions are static in their behavior and lack the ability to adapt over time. As cyber threats in smart grid environments grow more dynamic, fixed models trained on historical data may

struggle to detect novel attack strategies. Another notable limitation is the frequent separation between different data modalities. Prior studies often focus on either network traffic or system behavior independently, rather than combining both for richer context-aware detection. The proposed ProAPT model addresses these limitations through its integration of deep reinforcement learning with LSTM-based temporal modeling, allowing it to dynamically learn and predict sequential attack stages. Unlike static models, ProAPT can adapt to new patterns without requiring manual retraining.

3. METHODOLOGY

The ProAPT model [11] is a novel approach designed for predicting and mitigating APTs using DRL**.** The model leverages DRL's ability to continuously learn and adapt to dynamic environments, making it ideal for addressing the evolving nature of cyber threats in complex systems like smart grids. Smart grids present unique challenges due to their complexity, scale, and reliance on interconnected IoT devices.

The ProAPT model is based on Q-learning and LSTM to project the following step of APTs. As some relations exist between the attack steps, LSTM is used for value function approximation. LSTM is a modified version of RNN and facilitates the recall of past data and solves the problems of RNN. LSTM is employed to keep the previous states over long periods. The APT projection problem can be considered as a Markov Decision Process. Detection of normal or abnormal behavior at the current time step will alter the environment. The changing environment will also influence the next decision. Hence, it is natural to adapt this problem to the framework of Reinforcement Learning. We describe the Deep Reinforcement Learning System for the APT projection problem as follows: We demonstrate each state by features such as the source IP address, destination IP address, source port number, destination port number, timestamp, attack type, header length, flow duration. The agent receives the current state and selects the best action based on the ϵ -greedy policy. Indeed, the agent receives the correlated alerts and selects the following attack step. The reward is 1 or 0 for a correct/incorrect attack prediction. We use a Q-learning algorithm to learn the agent. To approximate the Q function, we employ LSTM, as some relations exist between attack steps. A Q function provides the maximum expected reward at a specific state and action. We employ APT datasets instead of interacting with the environment to reduce the time spent learning, testing, and evaluating. Although employing datasets increases the speed of learning and testing, interacting with the environment is suitable for predicting unknown APTs.

As mentioned, we give data from an APT dataset as input to DRLS. Based on the input data, the agent learns how to predict the following step of attacks.

Based on Fig. 1, we randomly divide the input dataset into sections and select the index. Then, from the selected index, we consider N number of data as training data. Each Training data, as input for LSTM, include the features of the alerts such as source IP address, destination IP address, source port number, destination port number, timestamp, attack type, header length, and flow duration. The second part is the data label in step t+1. This part shows the attack label in step t+1 such as automated collection, screen capture, exfiltration over C2 channel, ingress tool transfer.

For example, S_0 represents the attack step at (t_0), and a_1^* expresses the attack label at time (t_1) and for the state S_1. Since we want to recognize the following step of the attack in the DRLS, we consider the following step label in each state and use it to determine its reward. Fig. 2 demonstrates a DRLS to predict the following attack step. As mentioned, we give data from an APT dataset as input to DRLS. Based on the input data, the agent learns how to predict the following step of attacks. Input data consists of three parts. The first part expresses the state at time (t). This part includes the features of the correlated alerts at time (t). The second part is the data label in step (t+1). This part shows the attack label in step (t+1). The third part describes the state at time (t+1). That is a feature of correlated alerts at time (t+1). The first part of the input is entered into the LSTM neural network to approximate the value function of different actions for the state at time (t). In this context, LSTM approximates the value function for the following step of the ongoing attack. We display the approximated value with ($\hat{a_{t+1}}$), which has the value ($\hat{q_{t+1}}$). Then, based on the ϵ -greedy policy, the action with the highest value function is selected by a probability of ϵ . Finally, the approximated value ($\hat{a_{t+1}}$) is compared with the main label of the following step of the attack, which is the second part of the input data (a_{t+1}^*). If the comparison result is equal, the reward (+1) is given to the agent; otherwise, the reward (0) is given. The third part of the input is used to calculate the error function and update the LSTM. So that the state-expressing features at time (t+1) are entered in the second LSTM for approximation of value functions for different actions. At this point, the policy is the selection of action with the most significant value. In our problem, actions are the following step of attacks. Then, the obtained value is used to calculate the Mean Square Error Loss between the Q-value approximated by the LSTM for the state at time t and a reference value (q_{ref}). The reference value ($q_{ref} = r_t + \lambda \times \hat{q_{t+1}}$.) is obtained by adding the reward at time t (r_t) to the Q-value for the state at time t+1 multiplied by a discount factor (λ). The pseudocode for the smart grids APT prediction is depicted in Algorithm 1.

The output space (actions) corresponds to predicting the next attack step in an APT sequence is stated in Table 1.

Algorithm 1. Smart Grids APT prediction

Initialize Replay Buffer B
Initialize Q-network with LSTM: Q(s, a; θ)
Initialize Target Q-network with parameters θ− ← θ
Set discount factor γ and exploration rate ε

for episode = 1 to MaxEpisodes:
  Initialize state s_0 using features of alert at time t_0
  for each step t in episode:
    # Step 1: Choose action using epsilon-greedy policy
    if random() < ε:
      Select random action a_t (i.e., randomly select next predicted attack step)
    else:
      Select action a_t = argmax Q(s_t,a; θ)

    # Step 2: Perform action (predict next step)
    Observe reward r_t (1 if correct prediction, 0 otherwise)
    Observe new state s_{t+1} from alert features at time t_{+1}
    Store transition (s_t, a_t, r_t, s_{t+1}) in B

    # Step 3: Sample mini-batch of transitions from B
    for each sampled (s_i, a_i, r_i, s_{i+1}):
      Predict target value:
      Q_target = r_i + γ * max_a' Q_target(s_{i+1}, a'; θ−)
      Update Q-network using MSE loss:
        Loss = (Q(s_i, a; θ) - Q_target)²

    Every C steps:
      θ− ← θ # Update target network

    $s_t \leftarrow s_{t+1}$

  Decay ε (exploration rate)

- State (s_t): Vector of features at time t such as source IP, destination IP, source port, destination port, timestamp, attack type, header length, flow duration
- Action (a_t): The predicted next attack step from a predefined set of APT steps derived from the MITRE ATT&CK framework.

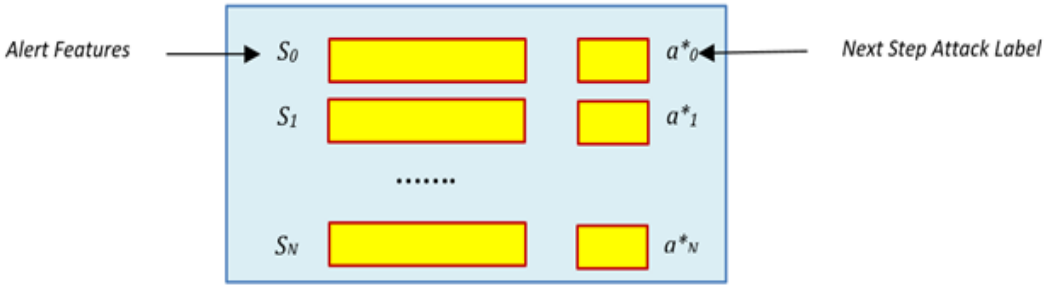

Fig. 1. Data Preparation in ProAPT model (Dehghan et al., 2022)

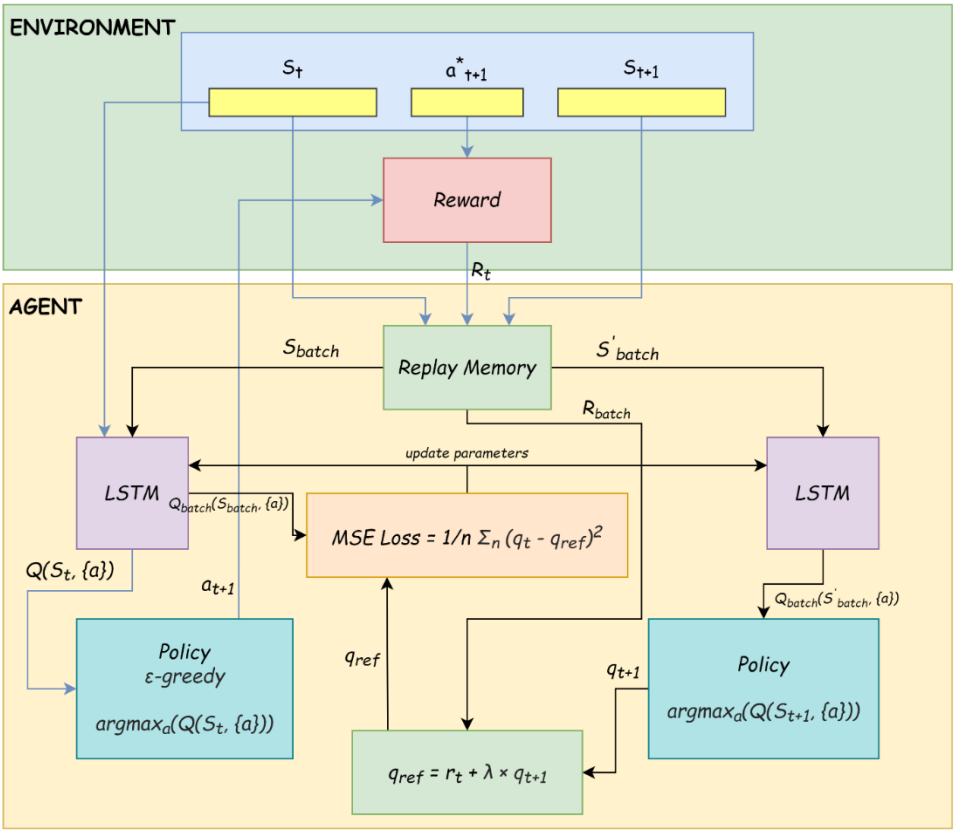

Fig. 2. The architecture of the ProAPT Model (Dehghan et al., 2022)

TABLE 1
The output space corresponds to APT prediction

Action ID	Predicted Attack Step	Description
0	Automated Collection	Data staged for exfiltration
1	Screen Capture	Attacker takes screenshots
2	Exfiltration over C2 Channel	Sensitive data exfiltrated using covert communication
3	Ingress Tool Transfer	Uploading malicious tools for further exploitation
4	Credential Dumping	Extraction of credentials from memory or files
5	Remote SSH	Remote access for lateral movement
6	Masquerading	Use of deceptive filenames/paths to evade detection
7	Data Destruction	Deletion of logs or sabotage
...	(Additional tactics as needed)	Aligned with MITRE categories from the CICAPT-IIoT dataset

Fig. 3. Pre-processing Steps

The steps of our methodology is as follows:

Data Preprocessing: Initially, the data is preprocessed to standardize features and address any missing values, as shown in Fig. 3.

Hyperparameters Tuning: Key hyperparameters like learning rate, discount factor, and exploration rate are carefully tuned to maximize the model's performance. Grid search is used to find the optimal combination of these hyperparameters, ensuring the model performs at its best.

Feature Selection: Performed using Random Forest-based feature importance to select the most informative attributes.

Data Splitting: The dataset is split into 70% for training, 30% for testing

Model Training: The model undergoes training using the reinforcement learning framework. As it trains, the model updates its Q-values based on the feedback it receives from the reward function, gradually refining its predictions over time.

Test and Evaluation: Once trained, the model is evaluated using standard classification metrics like accuracy, precision, recall, F1-score. These metrics gauge how well the model predicts the next attack in the sequence, considering both correct predictions and penalties for mistakes. By evaluating the model with these metrics, we can assess its effectiveness in predicting the next step in an attack sequence and its overall value in enhancing smart grid cybersecurity with proactive defense strategies.

4. EXPERIMENTS AND RESULTS

The CICAPT-IIoT dataset [12] is employed to evaluate the proposed prediction model. This dataset is designed for cybersecurity research, specifically for detecting APTs in industrial Internet of Things environments. The dataset simulates a sophisticated APT campaign based on the APT29 attack group, capturing both provenance logs and network traffic data from a hybrid testbed that integrates real and simulated IIoT components. The CICAPT-IIoT dataset was generated using a controlled IIoT testbed built on the Brown-IIoTbed framework, featuring a combination of physical and virtual components. It consists of two main data types: provenance logs, and network traffic logs.

The provenance logs capture system-level interactions and process relationships through a provenance graph. It includes 32 unique features, tracking process execution, file access, and network connections. The network traffic logs include an attack information file, detailing attack timestamps, process IDs of malicious actions, and attack categories, enabling researchers to correlate network activity with specific APT tactics. This dataset realistically replicates multi-stage APT campaigns relevant to smart grid cybersecurity. The attack framework follows MITRE ATT&CK tactics, encompassing over 20 distinct attack techniques across eight major categories as stated in Table 2. [12].

The dataset's attack scenarios closely mimic real-world threats to smart grids, where attackers exploit vulnerabilities in IIoT devices, industrial control systems, and network infrastructure. By incorporating provenance-based monitoring and network traffic analysis, this dataset provides a robust foundation for machine learning-based APT detection in critical infrastructure security.

As stated above, the dataset used for this research is the CICAPT-IIoT dataset, which provides a rich set of features related to the operation of Industrial Internet of Things (IIoT) devices and the detection of network-based threats, including. This dataset includes real-time network traffic data, device status, and attack patterns, which serve as the input for our DQN and LSTM models. To train the ProAPT model, we preprocess the dataset following the steps outlined in Fig. 3 Next, we select the best hyperparameters. Table 3. demonstrates the best selected one.

By setting a low learning rate, we ensure that the updates to the model remain stable. Additionally, a high discount factor emphasizes long-term rewards, helping the model prioritize future outcomes. A low exploration rate encourages the model to exploit the policies it has already learned, while a larger batch size and higher update frequency help stabilize the training process.

TABLE 2
Attack Techniques Used in CICAPT-IIoT Dataset [12]

Tactic	**Example Techniques**	**Relevant APT Groups**
Collection	Data Staging, Screen Capture	APT28, APT29, APT39
Exfiltration	Exfiltration over C2 Channels	Lazarus, APT3, APT32
Command & Control	Ingress Tool Transfer	APT29, APT3
Persistence	Event-Triggered Execution	APT28, APT29, APT3
Discovery	System & Network Discovery	Chimera, Dragonfly, APT29
Credential Access	Unsecured Credentials, Password Extraction	APT3, APT39, HEXANE
Lateral Movement	Remote SSH Access	APT29, Lazarus
Defense Evasion	Masquerading, Data Destruction	APT28, APT29, Dragonfly

TABLE 3
The best hyperparameters

Model	Hyperparameter	Value
DQN with LSTM	Action Space	Security measures (e.g., block traffic, adjust security policies)
	State Space	Network traffic features, device status, attack signatures
	Neural Network Architecture	Fully connected feedforward network
	Learning Rate	0.001
	Replay Buffer Size	10,000
	Batch Size	64
	Epsilon (for exploration)	1 (decaying to 0.1)
	Target Network Update Frequency	Every 100 steps
	Input	Sequences of time-series data (traffic, device status)
	Number of LSTM Units	100
	Learning Rate	0.001
	Epochs	50
	Batch Size	64
	Activation Function	ReLU (hidden layers), Softmax (output)

To evaluate the performance of our prediction model, we use several key metrics, as outlined by Carvalho et al. [28]:

Accuracy: This measures the proportion of correct predictions out of all predictions. It provides an overall indication of how well the model is performing in predicting the next attack step in the sequence.

Precision: Precision assesses how many of the predicted attacks are actually correct. This is particularly important in cybersecurity, as false positives can have significant consequences. A high precision ensures that the model isn't falsely predicting attack steps.

Recall: Recall measures how many of the actual attacks were correctly predicted. In cybersecurity, this metric is crucial because we want to make sure the model doesn't miss any attacks, even if it leads to a few false positives.

F1-Score: The F1-score is the harmonic mean of precision and recall, providing a balanced measure of both. It is especially valuable when dealing with imbalanced datasets, such as when attacks are less frequent than normal behavior.

Time Consumption (ms): The amount of time each model takes to process the data and make predictions. More complex models like DQN typically take longer to process due to their deeper architectures and the need for more computations.

Bandwidth Usage (KB/s): The amount of bandwidth consumed during data transfer between the model and the system. Models that require processing more complex data often use more bandwidth due to the need for transmitting larger volumes of information.

Throughput (ops/s): The number of operations the model can perform per second. Models with optimized architectures and faster computation capabilities generally have higher throughput, meaning they can handle more operations in a shorter amount of time

These metrics are essential for assessing how well the model can predict the next steps in a multi-step attack sequence. In particular, precision and recall are crucial in cybersecurity to minimize false positives and ensure that attacks are detected in a timely manner [29]. We compared the proposed ProAPT model with additional deep learning (non-reinforcement) baselines beyond traditional ML models. Specifically, we included models widely used in temporal classification tasks such as GRU, Bi-LSTM, CNN- LSTM, and Transformer architectures, as depicted in Table 4. These models were trained on the same CICAPT-IIoT dataset and evaluated using the same metrics as ProAPT to ensure a fair comparison.

As shown in Table 4, ProAPT achieved the best accuracy and F1-score but required slightly more processing time and bandwidth compared to simpler models like GRU and LSTM. However, its ability to handle complex multi-stage attack sequences and maintain high throughput demonstrates its suitability for real-time cybersecurity in smart grid systems.

To select the most suitable deep reinforcement learning algorithm for the proposed model and the dataset, we evaluated various algorithms (DQN, Double DQN, PPO, A3C), among which DQN delivered the best results. A comparison of these algorithms is presented in Table 5.

These algorithms are widely used in complex reinforcement learning environments due to their stability and robustness in continuous and asynchronous settings. PPO employs a clipped objective function to maintain policy updates within a trust region, improving learning stability. A3C, on the other hand, leverages multiple asynchronous agents to stabilize training and efficiently explore large state spaces [30].

We implemented PPO and A3C using the same environment setup, state space, and reward functions used for DQN and Double DQN to ensure consistency. Our

results, summarized in Table 5, show that while both PPO and A3C performed competitively, the proposed DQN-based ProAPT model outperformed them in terms of accuracy, precision, and recall. Specifically, PPO. These results reinforce the suitability of DQN for discrete action spaces typical of smart grid security environments, where decisions like blocking IPs or raising alarms are categorical in nature. Moreover, we considered a hybrid model combining feature-engineered inputs with a lightweight anomaly detection layer before feeding into DRL. Although this hybrid approach improved interpretability slightly, it did not outperform the standalone DRL models in overall metrics. These additional comparisons support our choice of DQN as a highly effective and practical baseline for APT detection in smart grid environments, while also highlighting avenues for future exploration in combining DRL with hybrid or ensemble methods [31].

TABLE 4
Performance Comparison between ProAPT and Deep Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Time (ms)	Bandwidth (KB/s)	Throughput (ops/s)
ProAPT (DQN + LSTM)	92.5	91.8	93.2	92.5	150	120	5000
LSTM	89.8	88.4	91.0	89.7	95	90	5800
GRU	89.3	88.1	90.4	89.2	87	85	5900
Bi-LSTM	90.2	89.6	91.3	90.4	110	100	5600
CNN-LSTM	90.7	89.8	92.0	90.9	125	105	5400
Transformer	91.0	90.5	92.2	91.3	140	115	5100

TABLE 5
Performance Comparison between DQN, and Other DRL Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Time Consumption (ms)	Bandwidth Usage (KB/s)	Throughput (ops/s)
DQN	92.5	91.8	93.2	92.5	150	120	5000
Double DQN	90.3	89.5	91.4	90.4	160	115	4800
PPO	91.0	90.2	92.1	91.1	180	110	4700
A3C	88.2	86.9	89.4	88.1	200	100	4500
Hybrid	90.1	89.0	90.2	89.6	250	130	4700

5. FEATURE IMPORTANCE

Feature importance indicates how much each feature contributes to the predictions made by a machine learning model. In the case of the Random Forest Classifier, the importance of each feature is determined by its ability to reduce uncertainty or enhance decision-making at each split in the trees [32]. Decision trees within the Random Forest algorithm aim to find patterns in the data that best separate the different classes, such as benign behavior and various types of attacks (DoS, etc.). Features that result in the most impactful splits—those that effectively distinguish between these classes—are considered more important. Fig. 4 provides an overview of the feature importance results.

For the Network Traffic dataset, features related to traffic patterns, such as packet size variance, connection duration, and protocol usage, dominated the importance scores. Features indicating irregularities in network flow (e.g., unusually large data packets or abrupt connection terminations) were highly predictive of threats. Certain features, like general connection metadata, showed low importance and could potentially be excluded to streamline model training. The leading features include packet size variance, connection duration, and frequency of specific protocols. This highlights that deviations in normal traffic patterns and protocol behaviors are indicative of advanced persistent threats. After implementing feature importance, we train and test the model, and summarize the results as demonstrated in Table 6.

- The confusion matrix for multi-stage attacks before feature selection is presented in Table 7. and Fig. 5. This matrix shows the actual vs predicted values for each of the 7 attack classes. The TP (True Positives), FP (False Positives), FN (False Negatives), and TN (True Negatives) for Class 0 (as an example) are calculated as follows:
- True Positives (TP): 10000 (correctly classified instances of Class 0).
- False Positives (FP): 1250 (the number of instances from other classes that were misclassified as Class 0).
- False Negatives (FN): 860 (the number of instances of Class 0 that were incorrectly classified into other classes).
- True Negatives (TN): 2084650 (all other instances not related to Class 0).

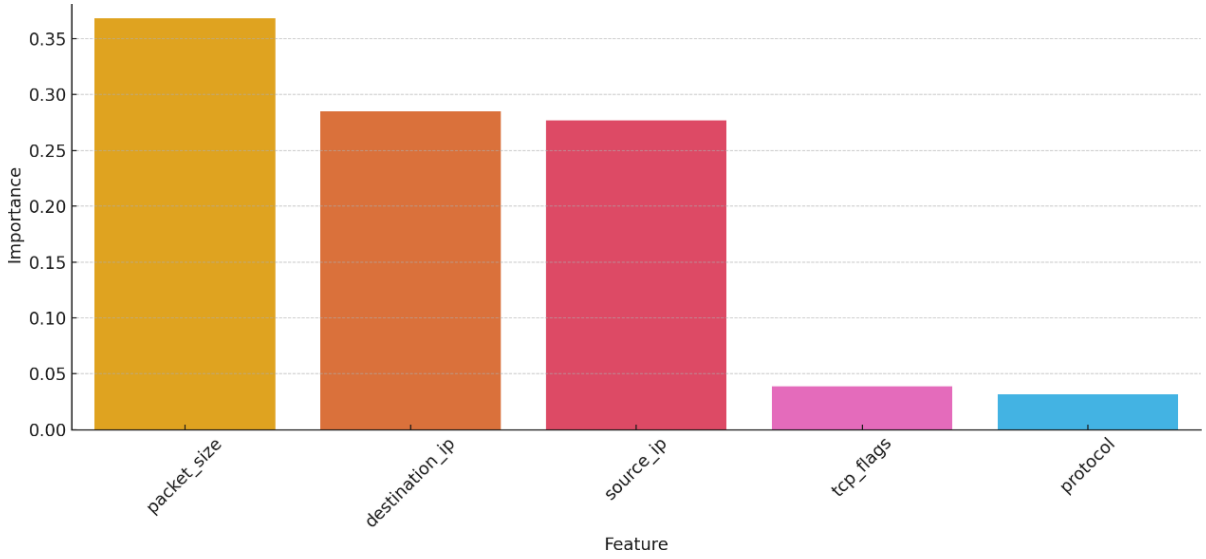

Fig. 4. Feature Importance Result for Network Traffic Dataset using Random Forest

TABLE 6
The results of prediction after feature importance implementation

Metric	Value
Accuracy (%)	93.8
Precision (%)	93.12
Recall (%)	95.2
F1-Score (%)	94.15
Time Consumption (ms)	150
Bandwidth Usage (kb/s)	120
Throughput (ops/s)	5000

TABLE 7
Confusion matrix before feature selection

	Predicted Class 0	**Predicted Class 1**	**Predicted Class 2**	**Predicted Class 3**	**Predicted Class 4**	**Predicted Class 5**	**Predicted Class 6**
Actual Class 0	10000	500	300	200	100	50	100
Actual Class 1	400	9500	400	300	200	100	150
Actual Class 2	200	300	9600	500	300	200	150
Actual Class 3	100	150	300	9700	400	300	200
Actual Class 4	50	100	200	400	9600	500	300
Actual Class 5	30	60	100	200	350	9800	500
Actual Class 6	80	120	150	300	400	450	9500

Moreover, the confusion matrix for multi-stage attacks after feature selection is presented in Table 8. and Fig. 5. After feature selection, the TP, FP, FN, and TN for Class 0 (as an example) are recalculated:

1) True Positives (TP): 10500 (correctly classified instances of Class 0).
2) False Positives (FP): 1100 (the number of instances from other classes that were misclassified as Class 0).
3) False Negatives (FN): 800 (the number of instances of Class 0 that were incorrectly classified into other classes).
4) True Negatives (TN): 2086250 (all other instances not related to Class 0).

Table 9. presents a comparison of the proposed ProAPT model with several recent works in the field of APT detection in smart grids. This comparison includes evaluation metrics such as accuracy, precision, recall, and F1-score, as well as important factors like the method used, dataset, attack types, and the year of publication. The selected works focus on applying deep learning techniques and machine learning methods to address cybersecurity threats in smart grids and IoT environments.

TABLE 8
Confusion Matrix after Feature Selection

	Predicted Class 0	**Predicted Class 1**	**Predicted Class 2**	**Predicted Class 3**	**Predicted Class 4**	**Predicted Class 5**	**Predicted Class 6**
Actual Class 0	10500	400	250	150	50	30	50
Actual Class 1	350	9800	350	250	150	80	100
Actual Class 2	150	250	9800	400	250	150	100
Actual Class 3	50	100	250	9800	300	250	150
Actual Class 4	30	60	150	300	9700	400	250
Actual Class 5	20	50	80	150	300	9700	400
Actual Class 6	60	100	120	250	350	400	9700

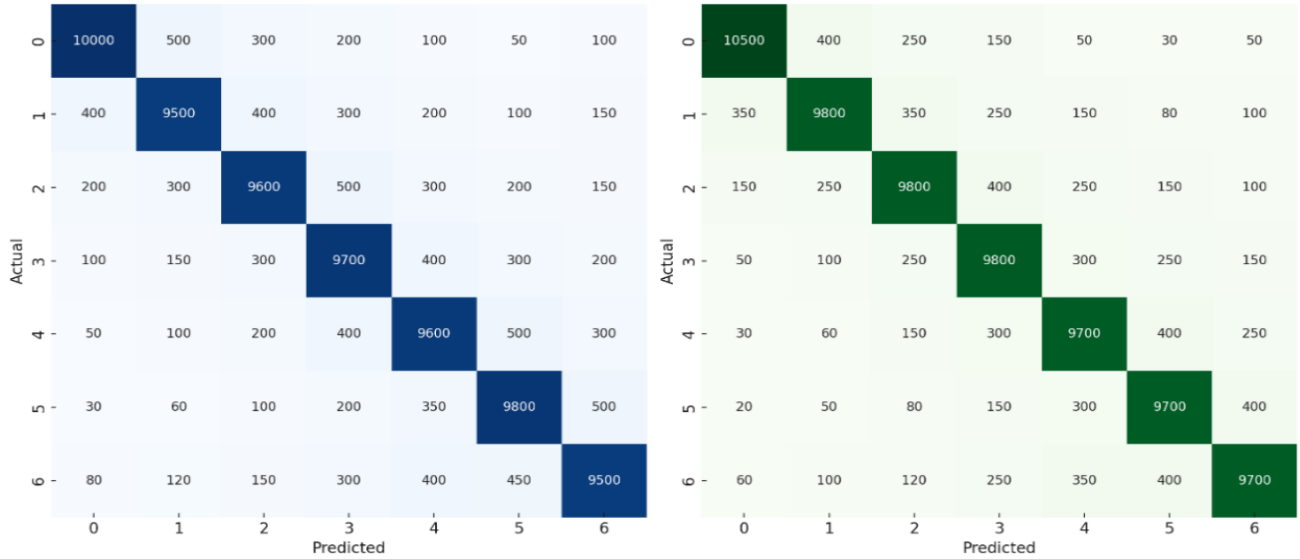

Fig. 5. Confusion Matrix before Feature Selection (Blue Diagram) and after Feature Selection (Green Diagram)

TABLE 9
Comparison of the Proposed Method with Recent Works

Study/Model	**Method**	**Dataset**	**Attack Types**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**
ProAPT (DQN) [11]	Deep Reinforcement Learning (DQN)	CICAPT-IIoT (2024)	APTs	93.8	93.12	95.2
Abdi et al. (2024) [5]	Deep Learning	Smart Grid Dataset	Malware, DoS, DDoS	90.0	89.5	91.0
Maiti & Dey (2024) [8]	Deep Reinforcement Learning	Simulated Smart Grid Data	Cyber-physical attacks	91.5	92.0	93.5
Khan et al. (2024) [7]	Machine Learning (Random Forest)	Smart Grid Cyber Attack Dataset	False Data Injection, APT	87.8	86.7	89.2
Sewak et al. (2023) [6]	Deep Reinforcement Learning (PPO)	IoT Network Traffic Dataset	APT, DoS, Ransomware	92.1	91.5	92.8

6. DISCUSSION

The proposed ProAPT model, powered by DQN, offers a compelling approach for enhancing the cybersecurity of smart grids by enabling proactive and adaptive responses to Advanced Persistent Threats (APTs). The model's strong performance—achieving over 92% accuracy, precision, and recall—demonstrates its effectiveness in detecting complex attack patterns, particularly in highly dynamic IIoT environments. One of the key strengths of the ProAPT model lies in its ability to continuously learn and adapt to new threats using reinforcement signals from the environment. Unlike traditional machine learning models that rely on static rules or labeled datasets, the DRL-based approach can dynamically adjust its policies based on feedback, making it especially suitable for environments where attack vectors evolve rapidly.

In this paper, feature selection was guided both by domain knowledge and empirical importance measures derived from training the DRL model. Specifically, features such as packet size variance, connection duration, number of failed login attempts, and inbound/outbound byte ratios were selected due to their proven relevance in

identifying abnormal behaviors associated with APTs. These features reflect the temporal and statistical properties of network flows that are often manipulated during different stages of an attack, such as reconnaissance, lateral movement, or data exfiltration. To further validate their influence on DRL decision-making, we conducted a permutation-based feature importance analysis, revealing that traffic-related features had the highest impact on the agent's Q-value updates. For instance, packet size variance was frequently associated with stealthy data transfers, while connection duration helped differentiate between persistent sessions initiated by malicious actors and short-lived benign activity. By incorporating these features into the state representation, the DRL agent learned to prioritize observations that carry strong signals of attack behavior, thereby enhancing its ability to make accurate, context-aware decisions in real-time. Incorporating feature selection and emphasizing its impact on DRL decision-making helps provide a deeper understanding of how the model works and why certain features are critical for success in detecting and mitigating APTs in IIoT environments.

However, translating this success to real-world deployment scenarios presents several challenges that merit further discussion. Scalability is one such concern. While the ProAPT model performs well in controlled simulations, deploying it across large-scale, heterogeneous smart grid infrastructures may require distributed training frameworks or federated learning approaches to handle high-volume data streams without overwhelming central systems.

Another important consideration is computational efficiency and real-time responsiveness. Although DQN provides a solid balance between performance and complexity, models like Double DQN introduce architectural overhead that may hinder real-time inference in latency-sensitive applications. In this study, we observed that Double DQN, despite its theoretical advantage in mitigating Q-value overestimation, slightly underperformed compared to standard DQN. This was likely due to slower convergence in environments with strong temporal dependencies, as found in the CICAPT-IIoT dataset. Nevertheless, this does not diminish the potential of Double DQN; rather, it emphasizes the importance of careful hyperparameter tuning and task-specific architecture selection. For example, techniques such as prioritized experience replay, reward shaping, or even incorporating temporal abstraction (e.g., options frameworks or recurrent networks) may enhance the model's ability to capture long-term attack strategies while preserving inference speed. To address real-time decision-making constraints, future implementations could leverage lightweight model compression techniques (e.g., pruning, quantization) or offload computations to edge-cloud collaborative architectures. Such hybrid setups allow for scalable deployment without compromising responsiveness. Furthermore, the integration of explainability mechanisms—such as attention layers, saliency maps, or SHAP values—can significantly improve the trustworthiness of DRL decisions in operational contexts. This aligns with ongoing efforts in critical infrastructure security, where human operators require transparent and justifiable decision-making processes to support real-time incident response.

In summary, while the proposed ProAPT model demonstrates excellent potential as a next-generation defense mechanism for smart grids, addressing its implementation challenges through targeted enhancements can further solidify its applicability. The insights gained from this study also underscore the importance of balancing model sophistication with practicality, suggesting promising directions for future research in explainable, scalable, and robust DRL-based cybersecurity systems.

7. CONCLUSION

The ProAPT model showcases the promise of DRL in enhancing smart grid cybersecurity by predicting and mitigating APTs. With high performance metrics—accuracy of 92.5%, precision of 91.8%, and recall of 93.2%—the model proves its ability to detect complex attack sequences in real-time. One of the model's strengths lies in the engineering of its state space and the careful selection of relevant features, such as packet size variance, connection duration, and protocol usage. These features provide critical insights into network behavior, making the model more efficient and effective in detecting attacks. By focusing on the most important features, the model reduces computational complexity, improves accuracy, and enhances the interpretability of its decisions.

However, there are still significant challenges to overcome in deploying this model in real-world smart grid environments. The scalability of the model must be improved to accommodate larger systems with vast amounts of data, and real-time adaptability must be enhanced to respond to new attack patterns. Furthermore, the interpretability of DRL models must be addressed to ensure that cybersecurity professionals can trust and understand the model's decisions in critical infrastructure contexts.

Future work should focus on addressing these challenges by improving scalability, integrating additional data sources for enhanced predictive accuracy, and enhancing the model's interpretability. Additionally, reducing false positives will be crucial for ensuring that the system can operate without causing unnecessary disruptions. Exploring hybrid models that combine DRL with other machine learning techniques could further enhance the robustness of the ProAPT model, enabling it to better handle new and emerging threats. Finally, incorporating explainability into DRL models, especially for applications in high-stakes environments like smart grids, will be essential to ensure that automated systems can work effectively alongside human experts.

In conclusion, while the ProAPT model demonstrates great potential, ongoing research and development are necessary to refine its scalability, adaptability, and transparency, ensuring that it can provide reliable and effective protection against the evolving landscape of smart grid cybersecurity threats.

8. NOMENCLATURE & UNITS

IoT	Internet of Things
APT	Advanced Persistent Threats
IDS	Intrusion Detection Systems
DRL	Deep Reinforcement Learning
M2M	Machine-to-Machine
LSTM	Long Short Term Memory
MDP	Markov Decision Process
DQN	Deep Q-Networks
IIoT	Industrial Internet of Things

9. REFERENCES

[1] W. Wang and Z. Lu. (2013). Cyber Security in the Smart Grid: Survey and Challenges. *Computer Networks*. [Online]. *57(5)*, pp. 1344–1371. Available: https://doi.org/10.1016/j.comnet.2012.12.017

[2] M. Dehghan and E. Khosravian. (2023). Private Federated Learning for APT Detection in Internet of Drones. *Quarterly Scientific Journal of National University of Skills*. [Online]. *20(3)*, pp. 465–484. Available: https://karafan.tvu.ac.ir/article_179732.html?lang=en

[3] Gunduz, M. Z., & Das, R. (2020). Cyber-security on smart grid: Threats and potential solutions. *Computer networks*. [Online]. *169*, p. 107094. Available: https://doi.org/10.1016/j.comnet.2019.107094

[4] Z. Ding, Y. Huang, H. Yuan, and H. Dong. (2020). Introduction to Reinforcement Learning. *Deep Reinforcement Learning: Fundamentals, Research and Applications, Singapore: Springer Singapore*. [Online]. pp. 47–123. Available: https://doi.org/10.1007/978-981-15-4095-0_2

[5] N. Abdi, A. Albaseer, and M. Abdallah. (2024). The Role of Deep Learning in Advancing Proactive Cybersecurity Measures for Smart Grid Networks: A Survey. *IEEE Internet of Things Journal*. [Online]. *11(9)*, pp. 16398–16421. Available: https://doi.org/10.1109/JIOT.2024.3354045

[6] M. Sewak, S. K. Sahay, and H. Rathore. (2023). Deep Reinforcement Learning in the Advanced Cybersecurity Threat Detection and Protection. *Information Systems Frontiers*. [Online]. *25(2)*, pp. 589–611. Available: https://doi.org/10.1007/s10796-022-10333-x

[7] Khan, M. A., Saleh, A. M., Waseem, M., & István, V. (2024, Sep.). Smart Grid Cyber Attacks: Overview, Threats, and Countermeasures. In 2024 22nd International Conference on Intelligent Systems Applications to Power Systems (ISAP). [Online]. pp. 1-5. Available: https://doi.org/10.1109/ISAP63260.2024.10744349

[8] S. Maiti, S. Adhikary, S. Dey, and A. R. Hota. (2024). Learning-Enabled Adaptive Voltage Protection Against Load Alteration Attacks On Smart Grids. *arXiv preprint*. [Online]. Available: https://doi.org/10.48550/arXiv.2411.15229

[9] Y. Yang, T. Littler, S. Sezer, K. McLaughlin, and H. F. Wang. (2011, Dec.). Impact of Cyber-Security Issues on Smart Grid. In International Conference and Exhibition on Innovative Smart Grid Technologies, pp. 1–7. [Online]. Available: https://doi.org/10.1109/ISGTEurope.2011.6162722

[10] Khosravian, E. and Dehghan, M. (2025). Cyber Risk Prediction for UAVs in Space-Related Missions Using Deep Reinforcement Learning. *Journal of Space Science and Technology*. [Online]. *18*, pp. 1-15. Available: https://doi.org/10.22034/jsst.2025.1527

[11] M. Dehghan , B. Sadeghiyan , E. Khosravian , A. Sedighi Moghadam and F. Nooshi. (2025). ProAPT: Projection of APTs with Deep Reinforcement Learning. The ISC International Journal of Information Security. 17 (1). pp. 25-41, doi: 10.22042/isecure.2024.428569.1052

[12] Ghiasvand, E., Ray, S., Iqbal, S., Dadkhah, S., & Ghorbani, A. A. (2024). CICAPT-IIOT: A provenance-based APT attack dataset for IIoT environment. *arXiv preprint*. [Online]. Available: https://doi.org/10.48550/arXiv.2407.11278

[13] Shees, A., Tariq, M., & Sarwat, A. I. (2024). Cybersecurity in Smart Grids: Detecting False Data Injection Attacks Utilizing Supervised Machine Learning Techniques. *Energies*. [Online]. 17(23), p. 5870. Available: https://doi.org/10.3390/en17235870

[14] S. Maiti and S. Dey. (2024). Smart Grid Security: A Verified Deep Reinforcement Learning Framework to Counter Cyber-Physical Attacks. *arXiv preprint*. [Online]. Available: https://doi.org/10.48550/arXiv.2409.15757

[15] H. Biswas. (2024, Sep.). Malware Trend in Smart Grid Cyber Security. *IEEE Region 10 Symposium (TENSYMP)*. [Online]. pp. 1–5. Available: https://doi.org/10.1109/TENSYMP61132.2024.10752141

[16] Paul, B., Sarker, A., Abhi, S. H., Das, S. K., Ali, M. F., Islam, M. M., ... & Saqib, N. (2024). Potential smart grid vulnerabilities to cyber attacks: Current threats and existing mitigation strategies. *Heliyon*. [Online]. *10*(19). Available: https://doi.org/10.1016/j.heliyon.2024.e37980

[17] N. Sahani, R. Zhu, J. H. Cho, and C. C. Liu. (2023). Machine Learning-Based Intrusion Detection for Smart Grid Computing: A Survey. *ACM Transactions on Cyber-Physical Systems*. [Online]. *7(2)*, pp. 1–31. Available: https://doi.org/10.1145/3578366

[18] H. Shadabfar, M. Dehghan, and B. Sadeghiyan. (2024). DSRL-APT-2023: A New Synthetic Dataset for Advanced Persistent Threats. In *21st International ISC Conference on Information Security and Cryptology (ISCISC 2024)*. [Online]. Available: https://doi.org/10.22042/isecure.2025.214212

[19] D. Kushner. (2013). The Real Story of Stuxnet. *IEEE Spectrum*. [Online]. *50(3)*, pp. 48–53.

[20] Khan, R., Maynard, P., McLaughlin, K., Laverty, D., & Sezer, S. (2016, August). Threat analysis of blackenergy malware for synchrophasor based real-time control and monitoring in smart grid. In 4th International Symposium for ICS & SCADA Cyber Security Research. [Online]. pp. 53-63.

[21] A. S. AL-Aamri, R. Abdulghafor, S. Turaev, I. Al-Shaikhli, A. Zeki, and S. Talib. (2023). Machine Learning for APT Detection. *Sustainability*. [Online].

15(18), p. 13820. Available: https://doi.org/10.3390/su151813820

[22] Nguyen, T. T., & Reddi, V. J. (2021). Deep reinforcement learning for cyber security. *IEEE Transactions on Neural Networks and Learning Systems*. [Online]. *34*(8), 3779-3795. Available: https://doi.org/10.1109/TNNLS.2021.3121870

[23] E. M. S. P. Veith, A. Wellßow, and M. Uslar. (2023). Learning new attack vectors from misuse cases with deep reinforcement learning. *Frontiers in Energy Research*. [Online]. *11*, p. 1138446. Available: https://doi.org/10.3389/fenrg.2023.1138446

[24] A. Sinha, R. Vyas, F. Alasali, W. Holderbaum, and O. P. Vyas. (2025). A deep reinforcement learning-based approach for cyber resilient demand response optimization. *Frontiers in Energy Research*. [Online]. *12*, 1494164. Available: https://doi.org/10.3389/fenrg.2024.1494164

[25] Y. Li, W. Ma, Y. Li, S. Li, Z. Chen, and M. Shahidehpour. (2025). Enhancing Cyber-Resilience in Integrated Energy System Scheduling with Demand Response Using Deep Reinforcement Learning. *Applied Energy*. [Online]. *379*, p. 124831. Available: https://doi.org/10.1016/j.apenergy.2024.124831

[26] Q. Sun, G. Lian, Z. Cao, X. Zeng, Z. Lv, L. Liu, ... and T. X. Zheng. (2023, Sep.). Deep Reinforcement Learning Based Secure Communication and Computing Resource Allocation for Grid Cyber-Physical System. In Proceedings of the 2nd International Conference on Internet of Things, Communication and Intelligent Technology, pp. 274–283. Singapore: Springer Nature Singapore. [Online]. Available: https://doi.org/10.1007/978-981-97-2757-5_29

[27] M. Dehghan and B. Sadeghiyan. (2018, May). An Efficient Secure Generalized Comparison Protocol. In *Electrical Engineering (ICEE), Iranian Conference on*, pp. 1487–1492. [Online]. Available: https://doi.org/10.1109/ICEE.2018.8472437

[28] Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*. [Online]. *8(8)*, p. 832. Available: https://doi.org/10.3390/electronics8080832

[29] M. Dehghan and B. Sadeghiyan. (2020, Oct.). Secure Multi-Party Sorting Protocol Based on Distributed Oblivious Transfer. In 10th International Conference on Computer and Knowledge Engineering (ICCKE), pp. 011–017. [Online]. Available: https://doi.org/10.1109/ICCKE50421.2020.9303630

[30] Dehghan, M. , Mahdi Zadeh, A. and Sadeghian, B. (2024). *A* Model to Measure Effectiveness in Cyber Security Situational Awareness. *Computer and Knowledge Engineering*. [Online]. 7(1), pp. 17-26. Available: https://doi.org/10.22067/cke.2024.83723.1101

[31] M. Dehghan and E. Khosravian. (2024). A Review of Cognitive UAVs: AI-Driven Situation Awareness for Enhanced Operations. *AI and Tech in Behavioral and Social Sciences*. [Online]. *2(4)*, pp. 54–65. [Online]. Available: https://doi.org/10.61838/kman.aitech.2.4.6

[32] M. Saarela and S. Jauhiainen. (2021). Comparison of Feature Importance Measures as Explanations for Classification Models. SN Applied Sciences. [Online]. 3(2), p. 272. Available: https://doi.org/10.1007/s42452-021-04148-9