network

MDPI

Article

Enhancing Smart Grid Cyber Resilience Against FDI Attacks Using Multi-Agent Recurrent DDPG

Tahira Mahboob [1,*], Mingwei Li [2], Awais Aziz Shah [1] and Dimitrios Pezaros [1,*]

[1] School of Computing Science, University of Glasgow, Glasgow G12 8QQ, UK; awaisaziz.shah@glasgow.ac.uk
[2] Department of Computer Science, University of Exeter, Rennes Drive, Exeter EX4 4RN, UK; ml969@exeter.ac.uk
* Correspondence: tahira.mahboob@glasgow.ac.uk (T.M.); dimitrios.pezaros@glasgow.ac.uk (D.P.)

Abstract

Digital substations (DSs) play a critical role in modern Energy and Power Electrical Systems (EPESs), enabling intelligent control, monitoring, and automation. With increased reliance on communication and sensing technologies, DSs are vulnerable to cyberattacks such as False Data Injection (FDI). An adversary may falsify transformer temperature readings, misleading protection mechanisms and resulting in incorrect disconnection actions. These false disconnections may disrupt power delivery, cause economic losses, and reduce equipment lifespan. To address these challenges, we propose a reinforcement learning-based approach for cyber protection of smart grids against false temperature data injection attacks. Specifically, this work designs a Long Short-Term Memory Deep Deterministic Policy Gradient (LSTM-DDPG) deep reinforcement learning algorithm that learns to detect normal patterns and responds to suspicious thermal patterns by dynamically adjusting disconnection decisions. The agents process sequential state features to differentiate between legitimate overload conditions and sudden anomalies caused by FDI attacks. We implement the proposed approach on the IEEE 30-bus distribution network using the Pandapower simulator. The experimental results indicate that the LSTM-DDPG controller outperforms conventional DDPG and DQN baselines, achieving a recall of 0.897, F1 of 0.945, precision of 1.00 and accuracy of 0.981 with a confidence interval of 95%. In addition, grid stability reaches up to 0.9815, 1.0, 1.0, 0.9926 with respect to the voltage stability score, transformer stability value, disconnection stability, and stability index, respectively. The proposed method led to fewer false disconnections, providing improved robustness against sensor manipulations.

Keywords: digital substation; false data injection attack; deep reinforcement learning

Check for updates

Academic Editor: Pingyi Fan

Received: 2 March 2026
Revised: 2 April 2026
Accepted: 14 April 2026
Published: 17 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The transformation of traditional substations into digital substations is a key trend in modern power systems. By replacing fixed-function hardware with software-defined logic and intelligent electronic devices (IEDs), digital substations bring increased flexibility, observability, and control to the grid. These advancements are especially important in the context of high penetration of distributed and renewable energy sources (DRESs), introducing further dynamicity and uncertainty into the system. However, as substations become increasingly connected and automated, they are also exposed to new types of cyber threats. One significant type of cyberattack is False Data Injection (FDI) [1,2], where attackers manipulate sensor data. The adversary's aim is to mislead control decisions [3,4].

For example, in transformer protection systems, an adversary may intentionally inject high temperature readings which may trigger unnecessary disconnections or, worse, prevent critical protection actions when they are truly needed [5,6].

The growing complexity of modern power systems and DRES integration, the electrification of loads, and increasing levels of remote management and automation set the scene for a new paradigm for reliability, selectivity and adaptability of protection schemes [7–9]. In a similar way, power transformers are a critical and vulnerable resource of power networks, and are exposed to thermal and electrical stresses. Conventional transformers have relied on numerical relays that function based on electrical measurements, such as current and voltage magnitude/phase, while considering thermal estimation models such as IEC 255-8 [10,11]. Furthermore, the introduction of high-efficiency loads, electrical vehicle charging infrastructure, rooftop photovoltaic systems, and inverter-based distribution generation presents challenges to traditional power systems. Subsequently, protection mechanisms that solely rely on current measurement for overload protection may be inefficient or inadequate and warrant further investigation.

The existing literature addresses FDI detection, reinforcement learning control, and transformer protection. We present a novel resilient protection strategy using multi-agent recurrent DDPG for transformer protection under temperature sensor FDI attacks. Transformer behavior and false temperature data injection (FTDI) anomalies change over time, necessitating a technique that captures temporal relationships beyond single-step observations, which is why LSTM-DDPG is used. While DDPG manages continuous control operations more skillfully than discrete approaches, the LSTM module offers memory of previous states, allowing for more accurate detection of subtle or misleading thermal trends. When combined, LSTM-DDPG produces disconnection decisions that are more precise, reliable, and resistant to attacks than threshold-based, DQN, or conventional DDPG methods. Current approaches either lack the temporal modeling capabilities necessary for detecting gradual attack patterns or fail to coordinate protection across multiple transformers. Furthermore, most studies evaluate methods in simulation without considering the partial observability of systems under a cyberattack. To address this challenge, we propose a DRL-based transformer protection scheme that can operate under both physical and cyber uncertainties. Specifically, we design a multi-agent framework where each transformer is controlled by an autonomous agent trained using Long Short-Term Memory Deep Deterministic Policy Gradient (LSTM-DDPG) [12]. The LSTM module enables each agent to process sequences of historical state information, enabling distinction between genuine thermal overloads and sudden, deceptive anomalies caused by FDI attacks.

Our work addresses these gaps by: (1) formulating transformer protection under FTDI as a multi-agent sequential decision problem using LSTM-DDPG agents that capture temporal dependencies; (2) developing a hybrid reward structure balancing safety, power quality, and attack resilience; and (3) validating performance using realistic FDI scenarios in an IEEE 30-bus system with Pandapower. To the best of our knowledge, this represents the first application of recurrent multi-agent reinforcement learning specifically for cyber-resilient transformer protection, achieving superior recall and lower false disconnection rates compared to non-recurrent baselines.

This approach formulates the disconnection decision as a continuous control task. At each timestep, the agent outputs a disconnection probability based on temporal transformer loading, voltage, and temperature patterns. The agent is trained in a partially observable environment using feedback from the power grid simulation. This allows it to learn policies that are robust to data tampering and that strike a balance between system safety and operational continuity [13]. We validate the proposed method using a realistic power grid test case with injected FDI attacks and dynamic load variations. The

experimental results show that our LSTM-DDPG controller outperforms traditional DDPG and DQN baselines [14], achieving higher recall and lower false disconnection rates.

The rest of this paper is structured as follows: Section 2 reviews related work on transformer protection and learning-based cyberattack mitigation. Section 3 introduces the preliminaries of the LSTM-DDPG algorithm and the problem formulation. Section 4 describes the simulation setup and reward formulation. Section 5 discusses the experimental results and comparison with baseline methods. Finally, Section 6 concludes the paper and outlines future directions.

2. Related Work

False Data Injection (FDI) attacks pose significant threats to modern power systems, particularly targeting sensor measurements in digital substations. Lin et al. [15] formulated FDI detection as a Markov Decision Process using Deep Q-Networks, achieving 96–100% detection rates across IEEE test systems. Their approach demonstrates that RL agents can learn attack patterns without prior knowledge of attack signatures. Zhang and Yang [16] pioneered LSTM networks for FDI detection, achieving 94.22% accuracy with an optimal temporal sequence of few time steps, establishing benchmarks directly applicable to transformer protection. Mahi-al-rashid et al. [17] developed a CNN-LSTM autoencoder, achieving 95.58% accuracy with 40% faster detection than traditional approaches, demonstrating real-time feasibility.

In the context of digital substations, cyber vulnerabilities in IEC 61850 [18] protocols require special attention. Tobar-Rosero et al. [19] developed the first comprehensive dataset for GOOSE spoofing attacks, while Nhung-Nguyen et al. [20] achieved 97–98% accuracy with sub-4 ms detection latency using DNNs. Hussain et al. [21] proposed a dual-module sequence content resolver validated using a real-time digital simulator (RTDS), demonstrating effectiveness against replay and masquerade attacks. These works establish that combining communication protocol validation with physical system constraints provides robust defence against sophisticated FDI attacks targeting transformer control systems.

The integration of deep reinforcement learning into networks has shown significant promise, particularly when incorporating temporal modeling capabilities [22]. Lin et al. [23] presented the first successful LSTM-DDPG integration for economic dispatch, validated through hardware-in-the-loop testing with OPAL-RT, achieving a 4.26% cost reduction. Cao et al. [24] and Zhang et al. [25] provide comprehensive reviews establishing that RL approaches excel in adaptive control under uncertainty, crucial for cyber-resilient protection schemes. For multi-agent coordination, Hu et al. [26] developed EA-MAAC with attention-based communication, demonstrating 15–20% performance improvement over single-agent approaches. Zhang et al. [27] introduced physics-informed MARL, incorporating power system constraints directly into neural architectures, achieving 25% faster convergence while improving robustness against adversarial attacks. Their graph neural network architecture enables agents to understand network topology, essential for detecting coordinated attacks across multiple transformers. The centralized training with a distributed execution paradigm provides scalability for substation-wide deployment. While the centralized approach does give a global view, this may not be suitable for critical operational environments, such as power grids.

In addition to examining load frequency control (LFC) in smart grids under False Data Injection (FDI) attacks, ref. [28] suggests a unified data-driven self-healing system that combines control, data repair, and attack detection. The method uses a multi-agent reinforcement learning asynchronous advantage actor–critic (MARL-A3C) scheme in conjunction with an attention-enhanced autoencoder within a GAN architecture to provide precise detection, effective data recovery, and stable decentralized frequency regulation

with insufficient labeled data, achieving an F1-score of 0.99. Due to latency, bandwidth limitations, and cybersecurity concerns, in general, the current multi-agent reinforcement learning techniques for smart grid control rely on centralized coordination or inter-agent communication, assumptions that are frequently infeasible. For reliable and efficient transformer disconnection decision-making, completely decentralized, communication-free frameworks—especially those that use temporal models like LSTM to manage partial observability—remain understudied.

Advanced machine learning techniques have revolutionized transformer diagnostics and protection. Rokani et al. [29] achieved 90.7% accuracy using hyperparameter-optimized neural networks with dissolved gas analysis, providing optimization guidance for deep RL networks. Notably, Attallah et al. [30] achieved 100% classification accuracy for transformer fault diagnosis using thermal imaging with ResNet-18, demonstrating that thermal signatures provide independent validation of temperature sensors—crucial for detecting FDI attacks. The 2.85 s training time enables real-time deployment.

For temporal anomaly detection, Duan [31] developed a transformer–GAN hybrid, achieving 95.18% accuracy through self-attention mechanisms and capturing long-term dependencies in power system operational data. This architecture's sliding window processing and adaptive learning specifically address sophisticated FDI patterns in transformer temperature sensors. These advances demonstrate that combining multiple sensing modalities with temporal pattern recognition provides robust protection against sensor manipulation [32,33].

FDI attacks have been extensively studied in power systems [3], where adversaries craft malicious inputs to compromise sensor integrity. One dangerous subtype is FTDI [34], which targets the thermal protection logic of transformers by injecting falsified overtemperature readings. Transformers are critical components in substations, responsible for voltage regulation and power flow between voltage levels. To avoid overheating and thermal breakdown, substations deploy protection schemes that disconnect a transformer when its sensed temperature exceeds a predefined threshold [5]. However, if attackers successfully manipulate the reported temperature, the protection mechanism can trigger false disconnections (causing service interruption and asset wear) or fail to respond to genuine thermal threats.

Traditional protection relies on static thresholds, which are vulnerable to deceptive but statistically plausible sensor readings. In this work, we propose an RL-based approach to mitigate FTDI attacks, where each transformer is controlled by an autonomous agent trained with LSTM-DDPG. Using the temporal dynamics of the state and temperature trends of the system, the agent learns to detect inconsistencies that arise from falsified measurements. This allows for a more intelligent disconnection policy that balances grid safety and service availability, even under adversarial conditions.

3. Methodology

3.1. Problem Formulation

In this work, the transformer protection problem under FTDI is formulated as an MARL task [35]. Each transformer in the distribution network is controlled by an autonomous agent that learns to decide whether to disconnect the transformer in response to both physical overloads and cyber-induced risks. The environment is built upon the Pandapower simulation framework, incorporating time-varying loads, thermal dynamics, and adversarial data injection mechanisms.

The power grid is defined as a tuple $\mathcal{N} = (\mathcal{B}, \mathcal{L}, \mathcal{T})$, where $\mathcal{B}$ denotes the set of buses, $\mathcal{L}$ the set of lines, and $\mathcal{T}$ the set of transformers. Each transformer $i \in \mathcal{T}$ is managed by a local RL agent that interacts with the system at discrete timesteps $t = 0, 1, \ldots, T$.

At each timestep, the agent receives a local observation $s_i^t \in \mathbb{R}^d$ comprising features such as transformer loading, reported temperature (which may be falsified), temperature trend, voltage magnitude, and power flow in nearby lines. Historical state sequences are retained and encoded using recurrent structures, i.e., LSTM [36], to enhance temporal awareness.

The agent outputs a continuous action $a_i^t \in [0,1]$, representing the probability of disconnecting transformer i at time t. A disconnection is triggered if a_i^t exceeds a predefined threshold θ. Through interactions with the environment, each agent learns a policy $\pi_i(a_i^t \mid h_i^t)$, where h_i^t denotes the internal hidden state or historical embedding. The objective is to maximize the expected discounted cumulative reward:

$$\pi_i^* = \arg\max_{\pi_i} \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t R_i^t\right], \tag{1}$$

where $\gamma \in [0,1]$ is the discount factor and R^t is the scalar reward signal assigned at time t based on system safety, power quality, and trend detection outcomes (detailed in Section 3.3).

This formulation poses several challenges. Due to the nature of FTDI attacks, the reported sensor data may not reflect the true transformer state, making the environment partially observable. The agent must learn to reason beyond instantaneous sensor readings and instead rely on context, such as historical behavior and physical consistency, to detect anomalies and make resilient control decisions. This stands in contrast to traditional threshold-based protection, which may be easily misled by data tampering.

An overview of the LSTM–DDPG algorithm is presented in Figure 1. Each transformer is assigned an autonomous RL agent operating in the Pandapower environment. At every timestep, the agent observes local features such as load, voltage, and temperature trends, which may include adversarially manipulated measurements. The agent then outputs a disconnection probability based on its policy. A disconnection is triggered if this probability exceeds a predefined threshold. Through continuous interaction with the environment and feedback from reward signals, the agents learn to distinguish genuine overload conditions from falsified sensor data, thereby ensuring resilient transformer protection against FTDI attacks.

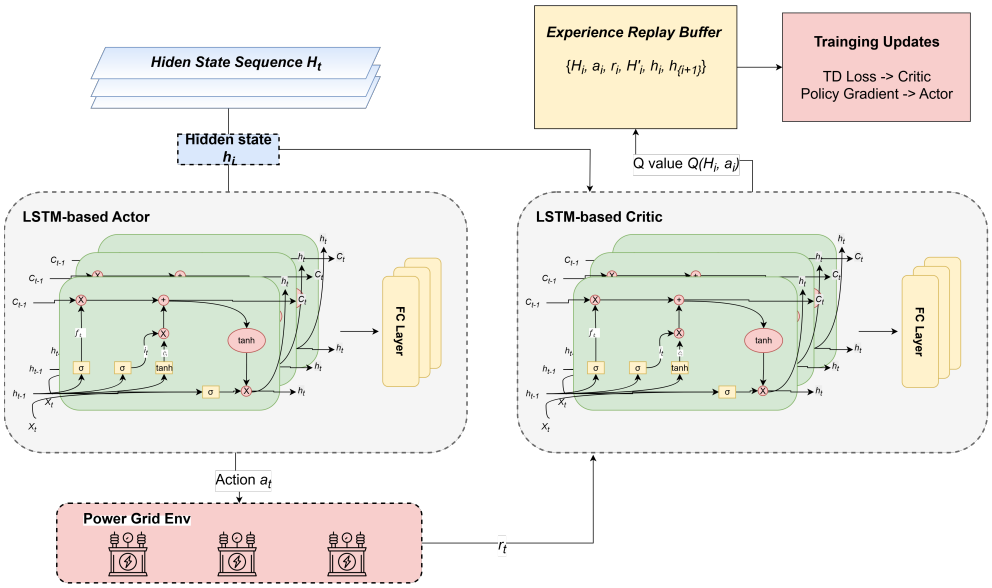


Figure 1. The LSTM-DDPG algorithm architecture.

3.2. Multi-Agent LSTM-DDPG Algorithm Architecture

DDPG is an off-policy, actor–critic reinforcement learning algorithm designed for continuous action spaces. It combines deterministic policy gradient updates with deep neural networks, including separate actor and critic models and experience replay, allowing stable learning in high-dimensional control tasks. In the context of using time series data as input, by concatenating historical observations, temporal information is implicitly recorded in a conventional DDPG or DQN, essentially treating the sequence as a static feature vector [37]. However, learning via implicit temporal information in input data prevents the model from determining which historical data is more pertinent and restricts its temporal comprehension to a predetermined window size. Furthermore, these basic models lack an adaptive memory mechanism to dynamically remember, forget, or prioritize past information—a crucial learning step in settings with long-term dependencies or delayed effects—because they do not maintain an internal state. Therefore, to take full advantage of the adaptive learning via recurrent architecture, we implement LSTM-DDPG, so the internal states and long-term dependencies are captured by the algorithm.

To address the transformer disconnection problem under FTDI attacks, we propose a multi-agent recurrent DDPG framework. Each transformer is controlled by an agent that observes a sequence of past states and outputs a disconnection probability ranging from 0 to 1. By integrating LSTM layers in both the actor and critic networks, the agents can capture temporal dependencies and identify anomalous patterns over time. Each agent maintains a historical observation buffer $H_i = [s_{t-\tau}^{(i)}, \ldots, s_t^{(i)}]$, hidden state vectors h_i^t for recurrent reasoning, and a local actor network A_i and critic network Q_i. The algorithm proceeds as described in Algorithm 1. The suggested approach manages transformer disconnection choices in a dynamic power grid by leveraging the LSTM-DDPG framework. Every transformer is represented as an independent agent that interacts with its surroundings over time. By capturing temporal dependencies such as changes in voltage and load, LSTM agents make optimal decisions under dynamic environments.

The system state is ascertained at each time step by a power flow analysis, and recent states are recorded in a history buffer. The actor network receives this sequence and the LSTM hidden state as input and produces a disconnection probability. After that, a binary action is sampled to determine whether the transformer should be disconnected. After implementing these actions, the system state is updated and reevaluated. Rewards for agents are determined by the efficiency and stability of the system. Stable and adaptive control policies can be learned by storing experiences and using them to train the actor and critic networks.

This architecture enables each agent to implicitly learn temporal correlations between loading conditions and temperature evolution, thereby improving the ability to detect and mitigate FTDI attacks. The use of LSTM networks allows the agents to distinguish between naturally rising temperatures and those suspiciously altered, while the multi-agent setup ensures localized, distributed protection across the entire grid.

Algorithm 1 Proposed Recurrent DDPG with LSTM for Transformer Disconnection

Initialize power grid network $\mathcal{N}$ with transformer set $\mathcal{T}$
Initialize recurrent actor and critic networks with LSTM layers for each agent A_i
Initialize replay buffer $\mathcal{D}$, temperature history, and disconnect probabilities p_i
Initialize hidden states $h_i^{(0)}$ for each transformer agent
Offline: Determine optimal threshold θ^* using validation dataset:
    Evaluate multiple thresholds $\theta \in [0, 1]$
    Select $\theta^* = \arg\max_\theta$ F1-score(θ)
for each episode $e = 1$ to E **do**
    Reset network and hidden states $h_i^{(0)}$
    for each time step $t = 1$ to T **do**
        Run power flow on $\mathcal{N}$
        for each transformer $i \in \mathcal{T}$ **do**
            Update historical state buffer $H_i = [s_{t-\tau}^{(i)}, \ldots, s_t^{(i)}]$
            Compute action probability a_i, hidden state $h_i^{(t+1)} \leftarrow A_i(H_i, h_i^{(t)})$
            if Training phase then
                Sample binary action $b_i \sim$ Bernoulli(a_i)
            else ▷ Execution phase
                $b_i \leftarrow \mathbb{I}(a_i \geq \theta^*)$ ▷ Deterministic thresholding
            end if
            if $b_i = 1$ **then**
                Disconnect transformer i
            else
                Keep transformer i connected
            end if
            Save $p_i \leftarrow a_i$
        end for
        Run power flow again on updated network
        for each transformer $i \in \mathcal{T}$ **do**
            Observe next state $s_{t+1}^{(i)}$
            Update $H_i' = [s_{t-\tau+1}^{(i)}, \ldots, s_{t+1}^{(i)}]$
            Compute reward r_i
            Store transition $(H_i, a_i, r_i, H_i', h_i^{(t)}, h_i^{(t+1)})$ into $\mathcal{D}$
        end for
        if buffer $\mathcal{D}$ has enough samples **then**
            Sample mini-batch of sequences
            Update critic using TD-loss over sequences
            Update actor using policy gradient over sequences
        end if
    end for
end for

At each timestep t, agent i observes a local state $s_i^t \in \mathbb{R}^{10}$, which is represented as

$$s_i^t = \begin{pmatrix} P_{\ell,i}^t,\ Q_{\ell,i}^t,\ \Delta V_i^{\mathrm{LV}}(t),\ \Delta V_i^{\mathrm{HV}}(t),\ I_i^{\max}(t), \\ \phi_i^t,\ \dot{L}_i^t,\ \Delta V_i(t),\ \delta T_i^t,\ \zeta_i^t \end{pmatrix}, \tag{2}$$

where $P_{\ell,i}^t$ and $Q_{\ell,i}^t$ denote real and reactive power losses, $\Delta V_i^{\mathrm{LV}}(t)$ and $\Delta V_i^{\mathrm{HV}}(t)$ are bus voltage deviations, $I_i^{\max}(t)$ is the maximum current, ϕ_i^t is the power factor, $\dot{L}_i^t$ is the loading change rate, $\Delta V_i(t)$ is the inter-bus voltage difference, δT_i^t is the anomaly indicator derived from raw sensor readings, and $\zeta_i^t \in [0, 1]$ is the sensor trust score. Formally, the input to the recurrent policy is the state sequence represented by

$$H_i^t = \left[s_i^{t-\tau+1}, \ldots, s_i^t\right] \in \mathbb{R}^{\tau \times 10}. \tag{3}$$

The actor network π_i maps this sequence into a continuous action

$$a_i^t = \pi_i(H_i^t) = \sigma\big(f_\theta(H_i^t)\big) \in (0,1), \tag{4}$$

where $\sigma(\cdot)$ is the sigmoid and f_θ is the LSTM-based model. At the training phase, a binary disconnection decision is obtained by

$$b_i^t = \mathbb{I}[\, a_i^t > \theta \,] \quad \text{or} \quad b_i^t \sim \text{Bernoulli}(a_i^t), \tag{5}$$

and the transformer is disconnected if $b_i^t = 1$. The model learns a continuous mapping from states to probabilities at different risk levels. Values near 1 indicate high risk, and values near 0 indicate safe conditions. This creates a smooth decision boundary between safe and unsafe regions. An optimal threshold is estimated with the validation dataset, resulting in optimal performance for metrics such as the F1-score. The learned threshold is then evaluated at execution time. For example, $a > 0.5$ extracts the learned decision boundary. It converts continuous outputs into binary actions without changing the learned policy. To summarize, the algorithm promotes stochastic exploration; transformer disconnection choices are sampled from a Bernoulli distribution depending on the actor's output probabilities during training. A deterministic optimal threshold is then used during execution to translate probabilities into deterministic actions for consistent and dependable control. This threshold is chosen through validation trials that maximize the F1-score.

3.3. The Reward Function

The reward function is carefully designed to encourage safe transformer disconnection, ensure power system stability, and enable agents to detect suspicious patterns potentially caused by FTDI attacks. The main structure is presented in Table 1. The reinforcement-learning agent is guided toward safe and dependable transformer operation by the categories assigned by the reward table. Disconnecting when the transformer is actually overloaded is the most desirable behavior, and it is reinforced by Large Positive (LP) rewards. The most hazardous situations, like failing to disconnect during overload or breaking grid restrictions like low voltage or heavy line loading, are penalized by Large Negative (LN) incentives. Correct noncritical conduct, such as stable normal operation or gently handling FTDI-like anomalies, is recognized with Small Positive (SP) incentives. Mildly suboptimal behaviors, such as needless disconnections or disregarding rising temperature trends, are discouraged by Small Negative (SN) rewards. A compact, understandable incentive structure is maintained while safety, stability, and anomaly-awareness objectives are clearly prioritized when magnitude-based categories are used instead of set values.

Table 1. Reward for various operations.

Condition	Type	Rationale
Overloaded and disconnected	LP	Correct protection
Overloaded and not disconnected	LN	Unsafe; severe risk
Disconnected when not overloaded	SN	Undesirable event
FTDI-like anomaly and disconnected	SP	Encourages caution
Normal state and online	SP	Stable operation
Low voltage/high line loading	LN	Grid violation
Temp rising fast and no trip	SN	Missed anomaly
Overloaded but temp decreasing	SP	Distrust misleading data

Note: LP = Large Positive reward; LN = Large Negative reward; SP = Small Positive reward; SN = Small Negative reward. These categories reflect relative importance rather than absolute numeric values.

At each timestep t, the reward received by the agent controlling transformer i is computed as a weighted sum of three components:

$$r_i^t = \lambda_1 \cdot r_{\text{safety}} + \lambda_2 \cdot r_{\text{power}} + \lambda_3 \cdot r_{\text{trend}}, \tag{6}$$

where $\lambda_1 = \beta_1$, $\lambda_2 = \beta_2$, and $\lambda_3 = \beta_2$ are weighting factors empirically chosen to prioritize safety. For instance, it is appropriate to assign a higher value to $\beta_1 = 10$ to ensure accurate disconnections and operational safety, and to assign $\beta_2 = 1$ and $\beta_3 = 1$. The first term, r_{safety}, evaluates whether the agent has made an appropriate disconnection decision given the actual transformer loading. For instance, if the transformer is overloaded (i.e., $loading > 1.5$) and is disconnected, a reward of +3 is assigned. If it is overloaded but remains connected, the agent is penalized with -5. If the transformer is not overloaded yet and gets disconnected, a penalty of -1 is given. However, in cases where the disconnection coincides with features indicative of FTDI attacks, such as a sharp positive temperature jump, increasing temperature, and decreasing load, a small compensation of +0.5 is added. Lastly, if the transformer is neither overloaded nor disconnected, the agent receives a Small Positive reward of +0.5 for maintaining stable operation.

The second term, r_{power}, penalizes control actions that negatively affect the electrical stability of the grid. If the voltage magnitude at the transformer's low-voltage bus falls below 0.95 per unit (i.e., $vm_pu < -0.05$ after centering), a penalty is applied in proportion to the deviation. Additionally, if the transformer remains online and the maximum loading of adjacent lines exceeds 120%, the agent is penalized based on the excess loading.

The third term, r_{trend}, aims to capture suspicious patterns in thermal behavior that may result from FTDI attacks. If the agent does not disconnect the transformer while the temperature exhibits a sharp increasing trend (i.e., $jump > 0.3$, $\Delta temp > 0$) and the transformer is not yet overloaded, a penalty of -1 is assigned. Conversely, if the transformer is truly overloaded but the temperature is unexpectedly decreasing ($\Delta temp < 0$), a bonus of $+0.5$ is added to encourage distrust of misleading data.

To stabilize training and avoid gradient explosion, the final reward is clipped to a fixed range:

$$r_i^t \in [-\Gamma, \Gamma] \tag{7}$$

For instance, a value of $[-20, 20]$ can be configured. This hybrid reward structure enables the agent to learn nuanced and risk-aware control strategies that balance correct transformer protection with robustness to stealthy FDI attacks.

4. Simulation Setup

4.1. Power Grid Structure

In this work, a simulated substation is implemented in advance to carry out the electric power flow simulation. Both false temperature injection simulation and DRL-based mitigation algorithms are implemented within this system. Pandapower is a python-based library used for power systems analysis. The performance is evaluated on the IEEE 30-bus test system, implemented via the pandapower library [38]. The IEEE 30-bus system was originally designed for transmission-level studies but has been adapted here to represent a heterogeneous distribution grid with multiple voltage layers and control components. The detailed electrical parameter changes, for instance, impedance and fault levels, have not been explicitly modeled.

We enhance the standard IEEE 30-bus system by adding backup connections or generators to improve reliability and mimic real-world power grid resilience strategies. The base topology is imported from the MATPOWER format using the case_ieee30() interface. To better simulate realistic distribution networks and enhance fault resilience, we introduce

structural redundancy to the system. Specifically, for each transformer in the network, we identify its low-voltage bus and connect it to an alternative bus via an artificially created tie-line or, when not feasible, equip it with a backup generator. These modifications emulate typical utility strategies for ensuring supply continuity during equipment disconnection or overload mitigation.

The final network contains 30 buses and multiple two-winding transformers that serve as the bridge between high-voltage transmission buses and local distribution feeders. Load points are attached primarily to the low-voltage side of the transformers. Figure 2 illustrates the topology of the modified IEEE 30-bus grid, where transformers, high-voltage backbones, tie-lines, and multi-voltage feeders are explicitly visualized. This environment is the testing ground for our reinforcement learning-based transformer protection algorithm under FTDI attacks.

To estimate transformer thermal conditions, we adopt a simplified yet widely accepted power-based temperature model [39]. The internal temperature T of a transformer is estimated as:

$$T = T_{\text{ambient}} + \Delta T_{\text{rated}} \times (\text{loading percent})^n, \tag{8}$$

where T_{ambient} is the ambient temperature, ΔT_{rated} is the rated temperature rise under nominal load (typically around 65 °C), loading percent is the current transformer load relative to nominal capacity, and n is the thermal exponent, often taken as 1.6 under air-cooled conditions. We assume air-cooled transformers and set $n = 1.6$ accordingly. This equation provides a physical basis for calculating the actual internal temperature of each transformer, which is compared against threshold values to determine overheat risk and reward function assignment.

To simulate realistic operating stress in distribution grids, we manually inject synthetic overload events into the system. These overloads serve as the physical basis for legitimate transformer disconnection and act as ground-truth for evaluating agent decisions. For each transformer, we randomly generate multiple overload events per episode. Each event consists of a randomly selected start time t_{start} and duration Δt; a load multiplier $\alpha \in [2.0,\ 3.0]$, applied to the transformer's base load; and a sinusoidal or stepped injection pattern. These events are applied by modifying the time-series power demand on the transformer's low-voltage bus using ConstControl. If a transformer has no attached load, a synthetic load is added to enable controlled overload. The injected profiles increase local heating and transformer temperature via thermal modeling and affect both operational safety and power flow. These overloads are essential for generating true positive cases for disconnection and for forcing learning agents to understand the relationship between load and thermal state.

4.2. FTDI Modeling

In addition to physical overloads, we simulate stealthy cyberattacks in the form of FTDI. These attacks aim to mislead controller decisions by forging sensor readings while leaving physical parameters untouched. Each FTDI attack is defined by a tuple (t, T^{false}), where t is the timestep and $T^{\text{false}} \in [110\ °\text{C},\ 160\ °\text{C}]$ is an anomalously forged temperature. We randomly generate a fixed number, e.g., 50, of such attacks and assign them to random transformers during each episode. A falsified temperature measurement is injected directly into the measured value of temperature_measured at time t, bypassing the normal thermal evolution model. The actual temperature, derived from power flow and loading, remains unchanged. The attack logic is encapsulated within the TransformerDisconnect controller class. These attacks create scenarios where the transformer appears overheated (high T^{false}) while actually operating normally. The agent must distinguish real overheating from deceptive signals using auxiliary features like loading, temperature trend, and ΔT.

This model tests the robustness of the LSTM-DDPG controller against false positive and false negative decisions in a partially observable environment. To visualize the impact of FTDI attacks, Figure 3 presents the actual and measured temperature curves for Transformer 1. While the actual temperature evolves smoothly based on load and thermal dynamics, the measured temperature exhibits sharp artificial spikes during several timesteps, corresponding to injected false data. These deviations can potentially mislead traditional protection schemes. The injected attacks raise the sensor reading to unrealistic values (e.g., >160 °C) while the actual temperature remains below the threshold. Such cases emphasize the necessity of intelligent policies that analyze trends, correlations, and load consistency instead of relying solely on raw sensor inputs.

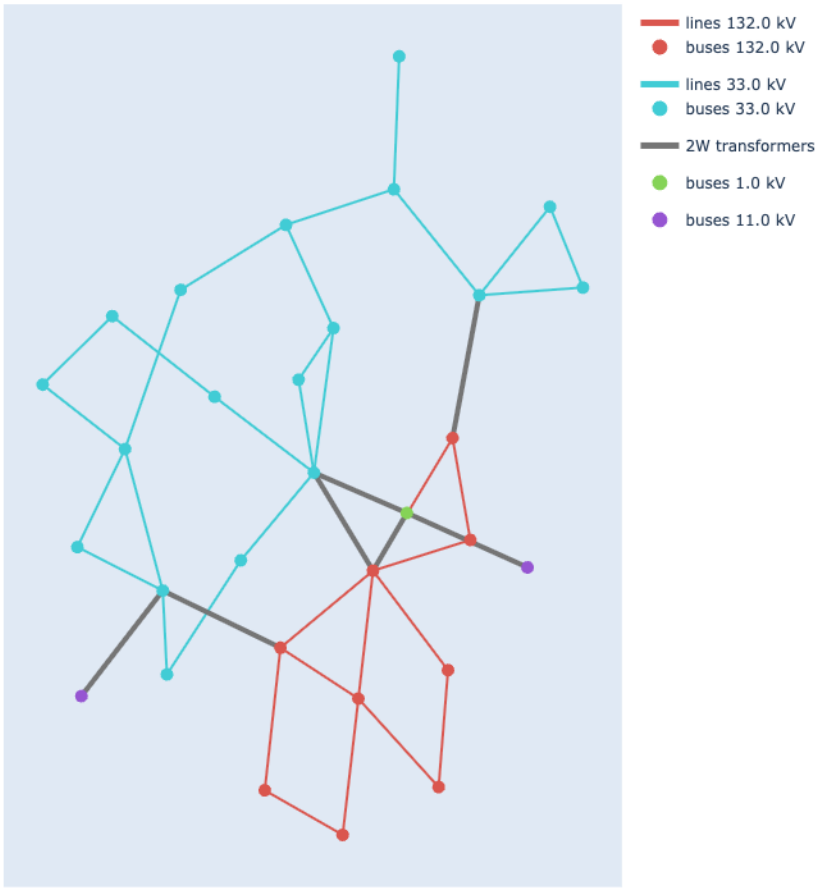


Figure 2. Power grid structure.

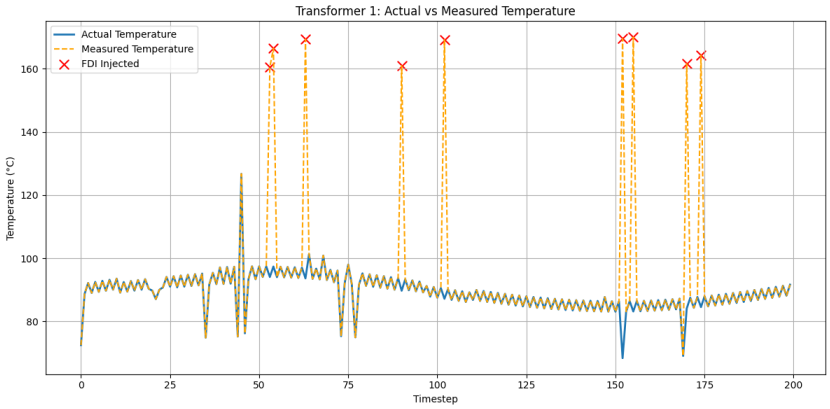


Figure 3. Comparison between actual and falsified transformer temperatures (Transformer 1). Red markers indicate injected false data (FTDI).

5. Experimental Results

We evaluate the proposed LSTM-DDPG-based transformer protection strategy under scenarios involving physical transformer overloads and FTDI attacks. The DL-model configuration is presented in Table 2. The experiments are conducted on a modified IEEE 30-bus distribution system simulated using the Pandapower framework, illustrated in Figure 4. We compare our method against two baseline reinforcement learning approaches: DDPG and DQN. These baselines follow the same training setup but do not include recurrent structures or historical state modeling.

Table 2. LSTM-DDPG network architecture and state representation.

Network	Layer	Configuration	Output Size	Activation
Actor	Input	Sequence $(T \times D)$	(5×10)	–
	LSTM	Hidden Units = 128	(128)	Tanh
	FC	$128 \rightarrow 1$	(1)	Sigmoid
Critic	Input	Sequence $(T \times D)$	(5×10)	–
	LSTM	Hidden Units = 128	(128)	Tanh
	Concatenate	+Action (1)	(129)	–
	FC1	$129 \rightarrow 128$	(128)	ReLU
	FC2	$128 \rightarrow 1$	(1)	Linear
State Representation				
Sequence Length (T)		5		
Feature Dimensions (D)		10 $[P^t_{\ell,i}, Q^t_{\ell,i}, \Delta V^{\mathrm{LV}}_i, \Delta V^{\mathrm{HV}}_i, I^{\max}_i, \phi^t_i, \dot{L}^t_i, \Delta V_i, \delta T^t_i, \zeta^t_i]$		
Training Parameters				
Batch Size		64		
Replay Buffer Size		10,000		
Discount Factor (γ)		0.99		
Soft Update (τ)		0.005		
Actor Learning Rate		1×10^{-4}		
Critic Learning Rate		1×10^{-3}		

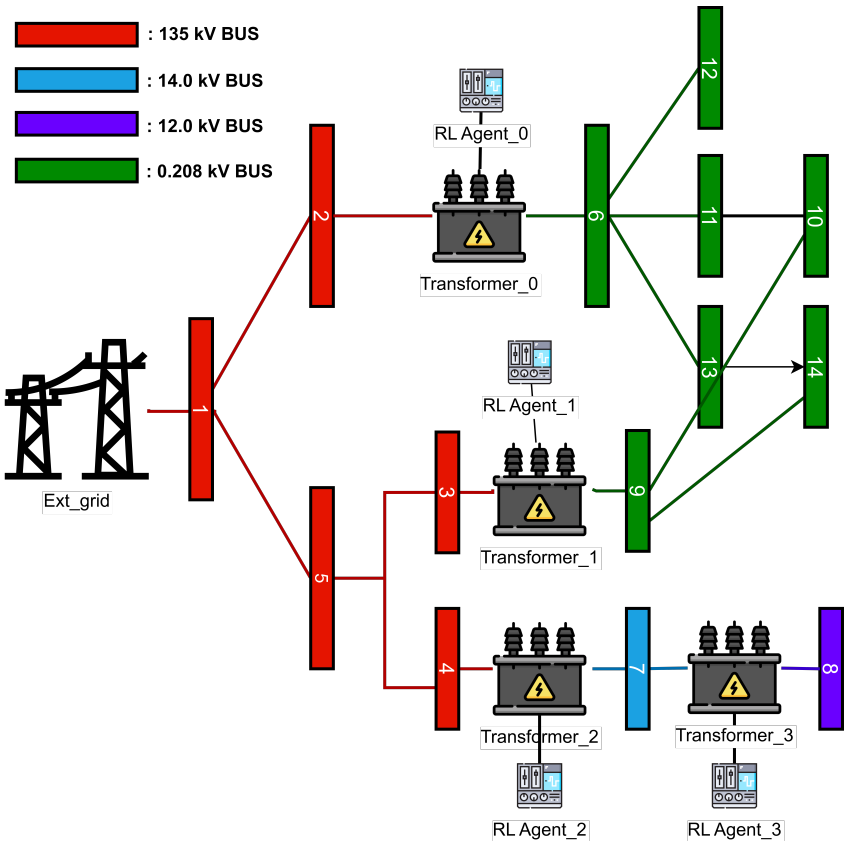


Figure 4. Power system architecture.

The evaluation focuses on the agent's ability to make correct disconnection decisions for each transformer at every timestep. A correct decision means disconnecting the transformer when it is overloaded, and keeping it connected otherwise. In contrast, false disconnections (false positives) may unnecessarily interrupt service, while missed disconnections (false negatives) can result in safety risks. To evaluate these decisions, we treat the task as a binary classification problem. The positive class corresponds to the transformer requiring disconnection, and the negative class corresponds to no disconnection needed.

5.1. Performance Metrics

The performance is evaluated in terms of deep learning metrics and system stability metrics. To quantitatively assess the protection and control performance of the DQN, DDPG, and LSTM–DDPG agents, four standard classification metrics are employed: accuracy, precision, recall, and the F1-score. These metrics are computed from the confusion matrix terms, namely true positive (TP), false positive (FP), false negative (FN), and true negative (TN).

- Accuracy evaluates the overall correctness of the agent's decisions:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}.$$

- Precision quantifies the proportion of correct disconnection actions among all disconnection decisions:

$$\text{Precision} = \frac{TP}{TP + FP}.$$

- Recall, or true positive rate, measures how effectively the agent disconnects transformers when an over-temperature condition truly occurs:

$$\text{Recall} = \frac{TP}{TP + FN}.$$

- F1-score is the harmonic mean of Precision and Recall, providing a balanced performance evaluation:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$

5.2. System Stability Metrics

In addition to classification-based assessment, four stability-oriented indices are defined to evaluate the resilience of the grid under FDI attacks and overload conditions [40,41]. Each metric lies within $[0, 1]$, where values closer to 1 indicate higher stability.

- Voltage stability score (VSS) measures the average deviation of bus voltages from the nominal value:

$$\text{VSS} = 1 - \frac{1}{N} \sum_{i=1}^{N} |V_i - 1.0|.$$

- Transformer stability value (TSV) quantifies the proportion of time during which all transformers remain below the thermal limit:

$$\text{TSV} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}(T_{\text{meas}}(t) < T_{\max}).$$

https://doi.org/10.3390/network6020025

- Disconnection stability (DS) represents the fraction of correct switching actions:

$$\mathbb{I}(x) = \begin{cases} 1, & \text{if } x \text{ is true} \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

$$\text{DS} = 1 - \frac{1}{T}\sum_{t=1}^{T} \mathbb{I}(T_t \leq T_{\max} \wedge \text{in_service}_t = 0). \tag{10}$$

- Stability index (SI) is a composite score combining voltage, thermal, and decision stability:

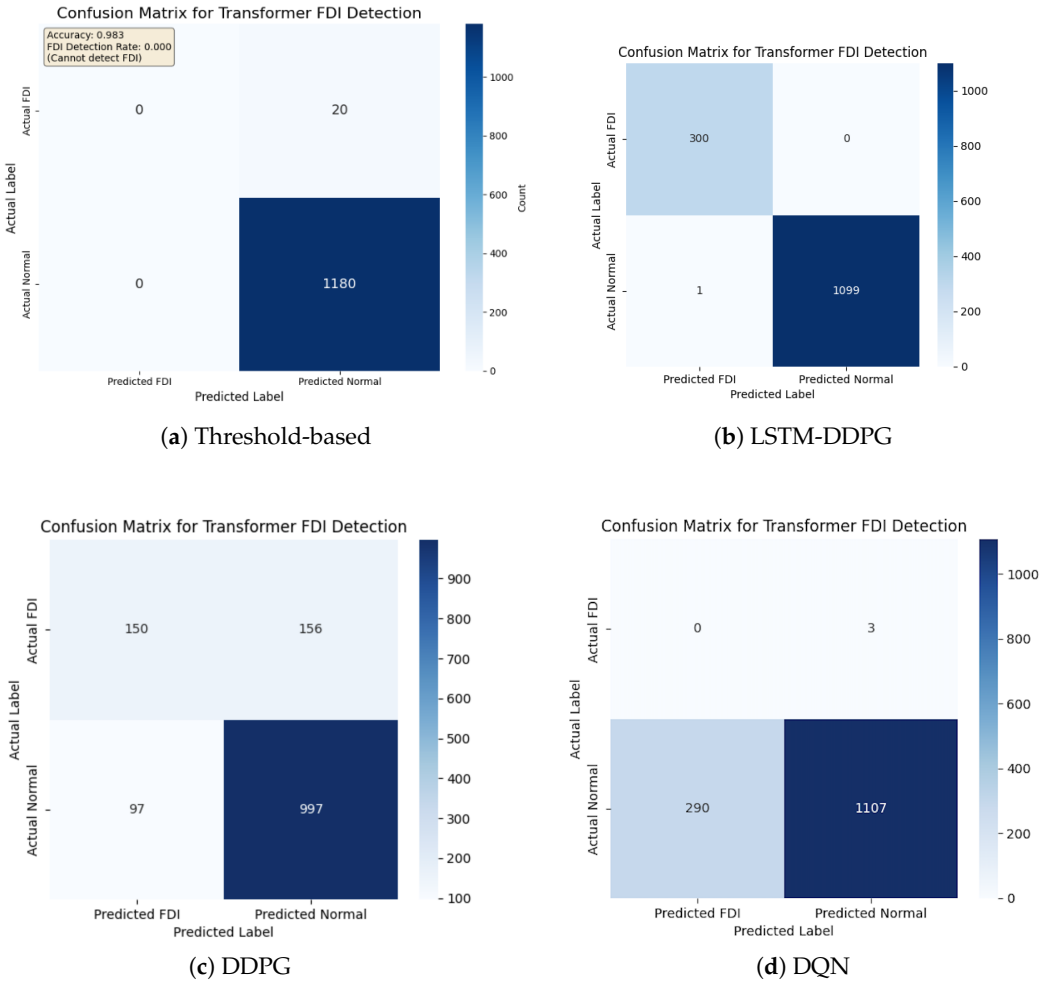
$$\text{SI} = \frac{1}{3}(\text{VSS} + \text{TSV} + \text{DS}).$$

We further examine the behavior of the agent under false temperature data injection (FTDI) attack, visualizing the actual and measured temperatures of a transformer across time. As shown in Figure 3, the measured temperature is manipulated at several time steps, creating spikes that do not match the true physical state. A rule-based system or non-sequence-aware agent may treat these spikes as valid, resulting in unnecessary disconnections. However, the LSTM-DDPG agent learns to identify these anomalies by comparing temporal patterns and recognizing that the reported measurements are inconsistent with recent trends or loading behavior.

Figure 5 shows the confusion matrices for the three methods. LSTM-DDPG achieves the best balance among true positives and true negatives, with very low false positives and false negatives. This suggests that the agent is able to correctly identify critical overloads without overreacting to safe conditions. In contrast, the DDPG baseline produces more false positives, meaning that it tends to disconnect transformers even when they are not under threat. The DQN agent performs worst in recall, with several missed disconnections, indicating that it struggles to react to overloads, especially under FDI influence.

We present accuracy, precision, recall, and F1-score in Figure 6 with FDI attacks in the $[150\ ^\circ\text{C},\ 160\ ^\circ\text{C}]$ range, and seed, $s = 42$. Among the four metrics, recall is the most critical in this application, as it reflects the system's ability to avoid safety failures by disconnecting transformers when necessary. LSTM-DDPG achieves the highest recall and F1-score, confirming that its use of historical state sequences helps detect gradual overload trends and filter out FDI-induced spikes. DDPG performs better than DQN overall, but it lacks the sequence modeling capability needed to handle deceptive signals effectively. Furthermore, we evaluate the performance of the proposed framework to validate its efficacy on multiple runs with various seed initialization. We report the performance of the framework in terms on accuracy, precision, recall, and F1-score on various runs under varying seeds, $s = \{42, 84, 168, 356, 400\}$, and temperatures, $[110\ ^\circ\text{C},\ 160\ ^\circ\text{C}]$, of FTDI attacks, with a confidence interval of 95%. The results show average performance at the execution phase, achieving an average of 0.981, 1, 0.897, and 0.945 for accuracy, precision, recall, and F1-score, respectively, indicating encouraging results, as indicated in Figure 7.

Furthermore, as presented in Figure 8, the LSTM–DDPG agent demonstrates resilience against FTDI attacks and dynamic thermal changes in the range of $[110\ ^\circ\text{C},\ 160\ ^\circ\text{C}]$, as evidenced by high scores across VSS, TSV, and DS. The LSTM-DDPG framework maintains voltage regulation, avoids thermal violations, and ensures accurate control decisions under adversarial conditions by consistently achieving high values across VSS, TSV, and DS, leading to a strong overall stability index (SI). As shown, the attacks are handled feasibly by the LSTM–DDPG algorithm, demonstrating the agent's temporal reasoning capability and resilience to falsified measurements.



(**a**) Threshold-based (**b**) LSTM-DDPG

(**c**) DDPG (**d**) DQN

Figure 5. Confusion matrices of the four methods for transformer FDI disconnection.

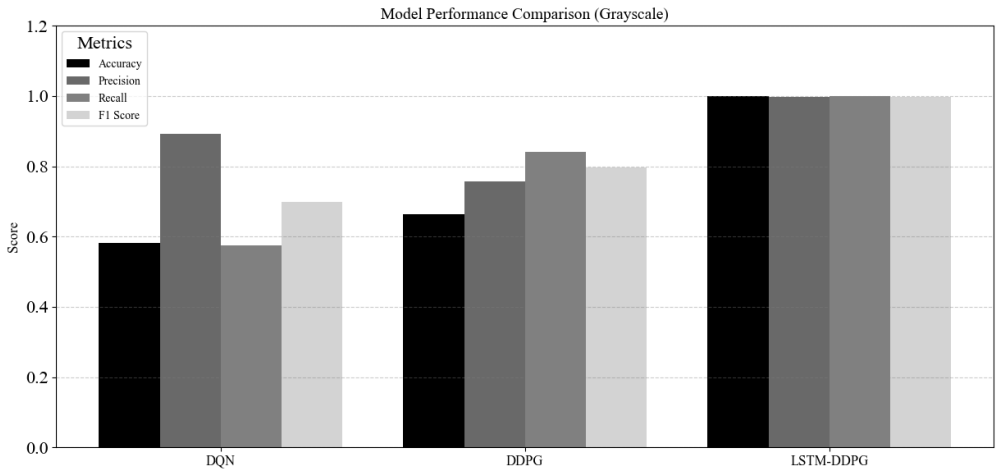


Figure 6. FTDI mitigation performance comparison for DQN, DDPG, and LSTM-DDPG.

We analyze the training stability of the LSTM-DDPG agent by tracking the loss curves of the actor and critic networks. Figure 9 shows that the loss decreases steadily over the training episodes, suggesting that the learning process is stable. This is particularly important in multi-agent settings where learning dynamics can be noisy or unstable. The results indicate that the proposed agent is not only effective in performance but also easy to train and deploy.

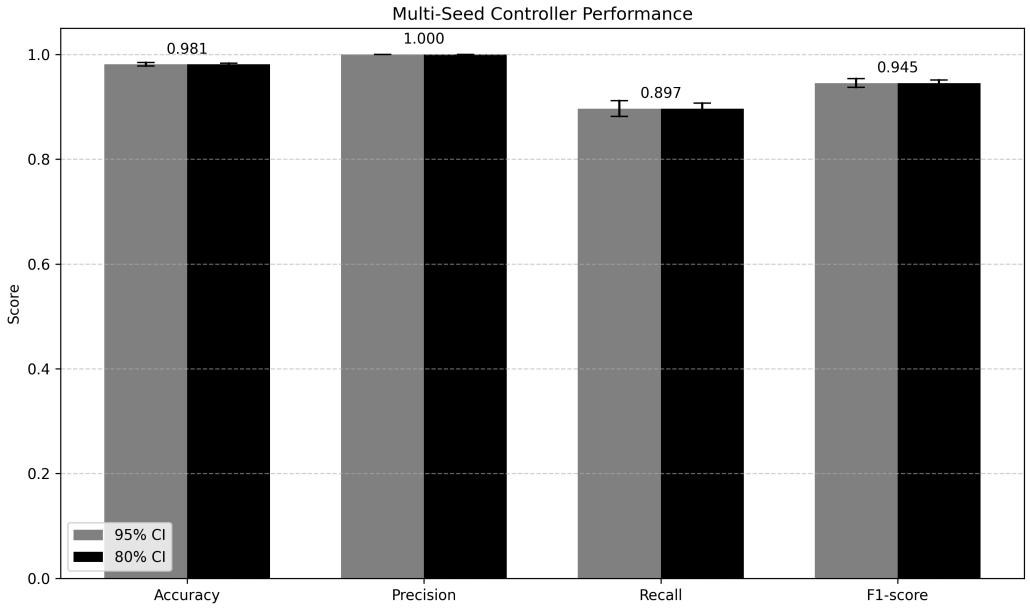


Figure 7. LSTM-DDPG detection performance under various runs with a confidence interval of 95%.

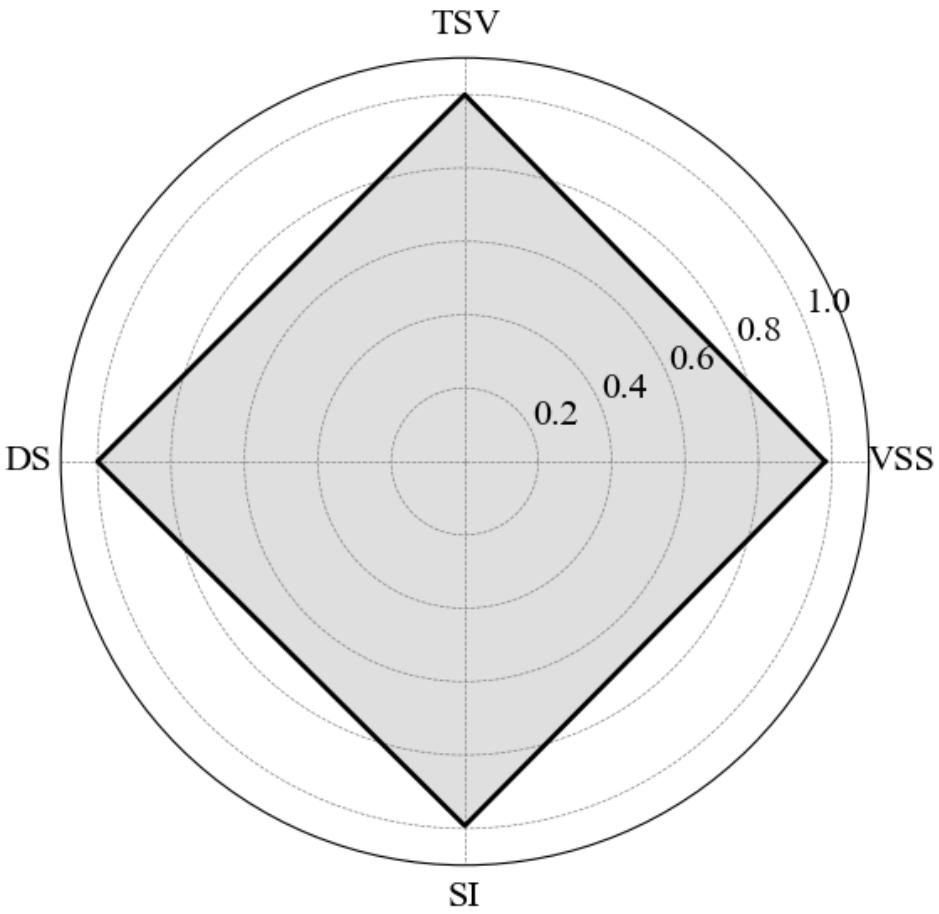


Figure 8. System stability for the proposed LSTM-DDPG.

The computational overhead of the proposed LSTM-DDPG approach was evaluated through real-time decision measurements across 1400 control steps. The experimental results reveal that the LSTM-DDPG controller requires an average of 1.875 ms per decision with median of 1.800 ms, 95th percentile with 2.314 ms, achieving a throughput of 533.2 decisions per second. Note that the latencies are reported for the inference times

for the LSTM-DDPG algorithm within the controller module during the execution phase. This represents an approximately 188× higher computational cost compared to traditional threshold-based methods requiring 0.01 ms per decision. However, this increased overhead is justified by the substantial improvement in FDI resilience. The LSTM-DDPG approach successfully detects FDI attacks, while threshold-based methods completely fail to detect any FDI attempts. The sub-2 ms decision latency remains well within acceptable bounds for transformer protection applications, where thermal time constants typically range from minutes to hours.

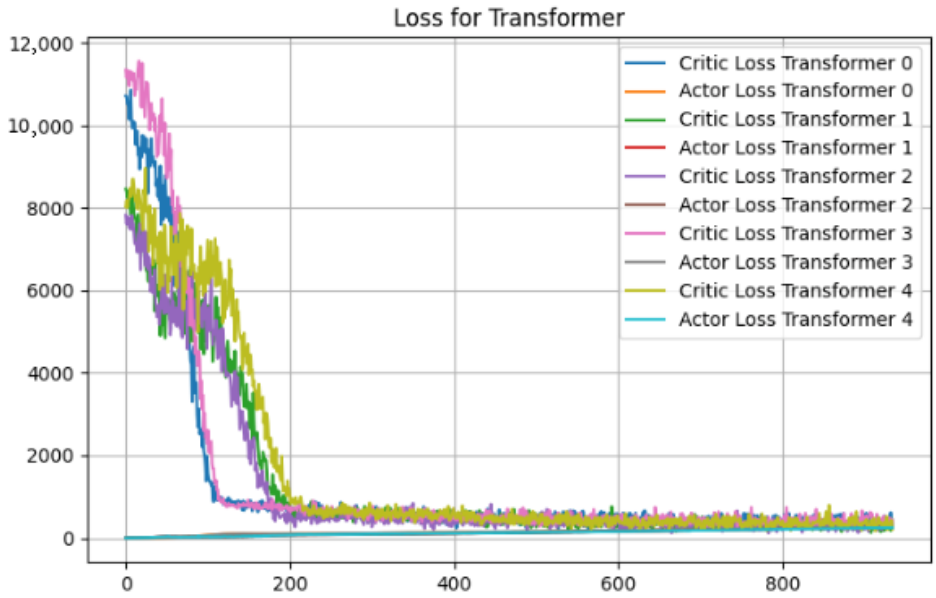


Figure 9. Training loss curve of LSTM-DDPG.

Furthermore, the consistent performance across all transformers with standard deviation of 0.859 ms and the bounded worst-case latency with a maximum of 25.625 ms ensure reliable real-time operation. This demonstrates that the LSTM-DDPG framework successfully balances computational efficiency with cyber resilience, making it suitable for deployment in modern digital substations where both performance and security are critical requirements.

6. Conclusions

This work presented a novel reinforcement learning-based approach for transformer protection under FTDI attacks. By formulating the transformer disconnection task as a multi-agent sequential decision-making problem under partial observability, we enabled each transformer to be autonomously controlled by an LSTM-DDPG agent capable of learning temporal patterns and reacting to both physical overloads and cyber-induced threats. The simulation environment, built on a modified IEEE 30-bus system using Pandapower, incorporates realistic overload conditions, temperature dynamics, and synthetic FDI attacks. The experimental results demonstrate that the proposed LSTM-DDPG method outperforms traditional DDPG and DQN baselines across multiple detection metrics and stability metrics. In particular, the use of LSTM layers significantly enhances the agent's ability to distinguish between genuine overheating and manipulated sensor data, leading to safer and more robust disconnection decisions. The findings confirm the effectiveness of deep reinforcement learning for cyber-resilient grid operation. Future work will explore more scalable multi-agent architectures, various topologies and deployment in real-world digital substation environments. Additionally, incorporating uncertainty quantification

and online adaptation may further enhance the robustness of the proposed method in dynamic and adversarial environments.

Author Contributions: Conceptualization, T.M., M.L., A.A.S. and D.P.; methodology, T.M., M.L., A.A.S. and D.P.; software, T.M. and M.L.; validation, T.M., A.A.S. and D.P.; formal analysis, T.M. and A.A.S.; investigation, T.M. and M.L.; resources, A.A.S. and D.P.; data curation, T.M. and M.L.; writing—original draft preparation, T.M. and M.L.; writing—review and editing, T.M., A.A.S. and D.P.; visualization, T.M. and M.L.; supervision, A.A.S. and D.P.; project administration, A.A.S. and D.P.; funding acquisition, A.A.S. and D.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Royal Academy of Engineering under grant number RCSRF2223-1645, Dstl under order number DSTL0000014002, EDF under order number 4840659490, UK Engineering and Physical Research Sciences Research Council (EPSRC) under reference EP/Z533221/1, and UK Research and Innovation (UKRI) under reference 10091225.

Data Availability Statement: The original data presented in the study are openly available in UofG-netlab GitHub repository at https://github.com/UofG-netlab/DDPG_FDI_SG (accessed on 13 April 2026).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Jafarigiv, D.; Sheshyekani, K.; Kassouf, M.; Seyedi, Y.; Karimi, H.; Mahseredjian, J. Countering FDI attacks on DERs coordinated control system using FMI-compatible cosimulation. *IEEE Trans. Smart Grid* **2020**, *12*, 1640–1650. [CrossRef]
2. Mahboob, T.; Holik, F.; Shah, A.A.; Pezaros, D. Adaptive Learning Feature Quantization for In-network FDI Detection in IEC 61850 Digital Substations. In *Proceedings of the 2025 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*; IEEE: North York, ON, Canada, 2025; pp. 1–6.
3. Ahmed, M.; Pathan, A.S.K. False data injection attack (FDIA): An overview and new metrics for fair evaluation of its countermeasure. *Complex Adapt. Syst. Model.* **2020**, *8*, 4. [CrossRef]
4. Mahboob, T.; Holik, F.; Shah, A.A.; Pezaros, D. Securing Smart Grids Against PSSE FDI Attacks. In Proceedings of the IEEE International Conference on Communications (ICC) (Accepted), Glasgow, UK, 24–28 May 2026; pp. 1–6. Available online https://eprints.gla.ac.uk/378639/ (accessed on 12 April 2026).
5. Swift, G.; Zocholl, E.; Bajpai, M.; Burger, J.; Castro, C.; Chano, S.; Cobelo, F.; de Sa, P.; Fennell, E.; Gilbert, J.; et al. Adaptive transformer thermal overload protection. *IEEE Trans. Power Deliv.* **2001**, *16*, 516–521. [CrossRef]
6. Cook, M.; Marnerides, A.; Johnson, C.; Pezaros, D. A survey on industrial control system digital forensics: Challenges, advances and future directions. *IEEE Commun. Surv. Tutor.* **2023**, *25*, 1705–1747. [CrossRef]
7. Ahmadi, S.A.; Sanaye-Pasand, M.; Abedini, M. Information Fusion Approach for Transformer Protection: Concept, Application, and Field Data Verification. *IEEE Trans. Ind. Electron.* **2023**, *70*, 12816–12828. [CrossRef]
8. Wu, B.; Hu, Y.F. Analysis of Substation Joint Safety Control System and Model Based on Multi-Source Heterogeneous Data Fusion. *IEEE Access* **2023**, *11*, 35281–35297. [CrossRef]
9. Rao, W.; Zhu, L.P.; Pan, S.; Yang, P.; Qiao, J.F. Bayesian Network and Association Rules-Based Transformer Oil Temperature Prediction. *Proc. J. Phys. Conf. Ser.* **2019**, *1314*, 012066. [CrossRef]
10. Chen, C.R.; Lee, C.H.; Chang, C.J. Optimal Overcurrent Relay Coordination in Power Distribution System Using a New Approach. *Int. J. Electr. Power Energy Syst.* **2013**, *45*, 217–222. [CrossRef]
11. Najy, W.K.; Zeineldin, H.H.; Woon, W.L. Optimal Protection Coordination for Microgrids with Grid-Connected and Islanded Capability. *IEEE Trans. Ind. Electron.* **2012**, *60*, 1668–1677. [CrossRef]
12. Zhang, Q.; Dehghanpour, K.; Wang, Z.; Qiu, F.; Zhao, D. Multi-agent safe policy learning for power management of networked microgrids. *IEEE Trans. Smart Grid* **2020**, *12*, 1048–1062. [CrossRef]
13. Liu, X.; Xu, H.; Liao, W.; Yu, W. Reinforcement learning for cyber-physical systems. In *Proceedings of the 2019 IEEE International Conference on Industrial Internet (ICII)*; IEEE: Orlando, FL, USA, 2019; pp. 318–327.
14. Hester, T.; Vecerik, M.; Pietquin, O.; Lanctot, M.; Schaul, T.; Piot, B.; Horgan, D.; Quan, J.; Sendonaris, A.; Osband, I.; et al. Deep q-learning from demonstrations. In *Proceedings of the AAAI Conference on Artificial Intelligence*; IEEE: New Orleans, LA, USA, 2018; Volume 32.

15. Lin, X.; An, D.; Cui, F.; Zhang, F. False data injection attack in smart grid: Attack model and reinforcement learning-based detection method. *Front. Energy Res.* **2023**, *10*, 1104989. [CrossRef]
16. Zhang, F.; Yang, Q. False data injection attack detection in dynamic power grid: A recurrent neural network-based method. *Front. Energy Res.* **2022**, *10*, 1005660. [CrossRef]
17. Mahi-al rashid, A.; Hossain, F.; Anwar, A.; Azam, S. False Data Injection Attack Detection in Smart Grid Using Energy Consumption Forecasting. *Energies* **2022**, *15*, 4877. [CrossRef]
18. *IEC TR 61850-1-1:2026*; Communication Networks and Systems for Power Utility Automation-Part 1-1: Introduction and Overview. Corrected Version 2026-02; International Electrotechnical Commission (IEC): Geneva, Switzerland, 2026.
19. Tobar-Rosero, O.A.; Roa-Romero, O.A.; Rueda-Carvajal, G.D.; Leal-Piedrahita, A.; Botero-Vega, J.F.; Gutierrez-Betancur, S.A.; Branch-Bedoya, J.W.; Zapata-Madrigal, G.D. GOOSE Secure: A Comprehensive Dataset for In-Depth Analysis of GOOSE Spoofing Attacks in Digital Substations. *Energies* **2024**, *17*, 6098. [CrossRef]
20. Nhung-Nguyen, H.; Girdhar, M.; Kim, Y.; Hong, J. Machine-Learning-Based Anomaly Detection for GOOSE in Digital Substations. *Energies* **2024**, *17*, 3745. [CrossRef]
21. Hussain, S.; Iqbal, A.; Hussain, S.M.S.; Zanero, S.; Shikfa, A.; Ragaini, E.; Khan, I.; Alammari, R. A novel hybrid methodology to secure GOOSE messages against cyberattacks in smart grids. *Sci. Rep.* **2023**, *13*, 1857. [CrossRef]
22. Simpson, K.A.; Rogers, S.; Pezaros, D.P. Per-host DDoS mitigation by direct-control reinforcement learning. *IEEE Trans. Netw. Serv. Manag.* **2019**, *17*, 103–117. [CrossRef]
23. Lin, F.; Chang, C.; Huang, Y.; Su, T. A Deep Reinforcement Learning Method for Economic Power Dispatch of Microgrid in OPAL-RT Environment. *Technologies* **2023**, *11*, 96. [CrossRef]
24. Cao, D.; Hu, W.; Zhao, J.; Huang, Q.; Chen, Z.; Blaabjerg, F. Reinforcement learning and its applications in modern power and energy systems: A review. *J. Mod. Power Syst. Clean Energy* **2020**, *8*, 1029–1042. [CrossRef]
25. Zhang, D.; Han, X.; Deng, C. Reinforcement Learning for Selective Key Applications in Power Systems: Recent Advances and Future Challenges. *IEEE Trans. Smart Grid* **2022**, *13*, 2935–2958. [CrossRef]
26. Hu, D.; Ye, Z.; Gao, Y.; Ye, Z.; Peng, Y.; Yu, N. Multi-agent deep reinforcement learning for voltage control with coordinated active and reactive power optimization. *IEEE Trans. Smart Grid* **2022**, *13*, 4873–4886. [CrossRef]
27. Zhang, B.; Cao, D.; Hu, W.; Ghias, A.; Chen, Z. Physics-Informed Multi-Agent deep reinforcement learning enabled distributed voltage control for active distribution network using PV inverters. *Int. J. Electr. Power Energy Syst.* **2024**, *155*, 109641. [CrossRef]
28. Cui, Y.; Wu, T.; Zhang, Y. Load Frequency Control of Smart Grid Based on Data-Driven FDI Attacks Detection and Data Repair. *IEEE Trans. Smart Grid* **2025**, *16*, 5404–5415. [CrossRef]
29. Rokani, V.; Kaminaris, S.; Karaisas, P.; Kaminaris, D. Power Transformer Fault Diagnosis Using Neural Network Optimization Techniques. *Mathematics* **2023**, *11*, 4693. [CrossRef]
30. Attallah, O.; Ibrahim, R.A.; Zakzouk, N.E. A lightweight deep learning framework for transformer fault diagnosis in smart grids using multiple scale CNN features. *Sci. Rep.* **2025**, *15*, 14505. [CrossRef]
31. Duan, J. Deep learning anomaly detection in AI-powered intelligent power distribution systems. *Front. Energy Res.* **2024**, *12*, 1364456. [CrossRef]
32. Ni, T.; Lan, G.; Wang, J.; Zhao, Q.; Xu, W. Eavesdropping mobile app activity via Radio-Frequency energy harvesting. In *Proceedings of the 32nd USENIX Security Symposium (USENIX Security 23)*; USENIX Association: Anaheim, CA, USA, 2023; pp. 3511–3528.
33. Li, J.; Zhou, H.; Wu, S.; Luo, X.; Wang, T.; Zhan, X.; Ma, X. FOAP: Fine-Grained Open-World android app fingerprinting. In *Proceedings of the 31st USENIX Security Symposium (USENIX Security 22)*; USENIX Association: Boston, MA, USA, 2022; pp. 1579–1596.
34. Roomi, M.M.; Biswas, P.P.; Mashima, D.; Fan, Y.; Chang, E.C. False data injection cyber range of modernized substation system. In *Proceedings of the 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*; IEEE: Tempe, AZ, USA, 2020; pp. 1–7.
35. Buşoniu, L.; Babuška, R.; De Schutter, B. Multi-agent reinforcement learning: An overview. In *Innovations in Multi-Agent Systems and Applications-1*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 183–221.
36. Kawakami, K. Supervised Sequence Labelling with Recurrent Neural Networks. Ph.D. Thesis, School of Computation, Information and Technology (CIT), Technical University of Munich, Munich, Germany, 2008.
37. Sumiea, E.H.; Abdulkadir, S.J.; Alhussian, H.S.; Al-Selwi, S.M.; Alqushaibi, A.; Ragab, M.G.; Fati, S.M. Deep deterministic policy gradient algorithm: A systematic review. *Heliyon* **2024**, *10*, e30697. [CrossRef]
38. Thurner, L.; Scheidler, A.; Schafer, F.; Menke, J.H.; Dollichon, J.; Meier, F.; Meinecke, S.; Braun, M. pandapower—An open-source Python tool for convenient modeling, analysis, and optimization of electric power systems. *IEEE Trans. Power Syst.* **2018**, *33*, 6510–6521. [CrossRef]
39. Correa, W.; García, D.; García, B. Power transformer temperature–moisture dynamics modeling: An experimental validation. *Electr. Eng.* **2023**, *105*, 761–773. [CrossRef]

40. Pasculescu, D.; Riurean, S.; Ilieva-Obretenova, M.; Lazar, T.; Tatar, A.M.; Fita, N.D. Intelligent modeling of PV–BESS microgrids for enhanced stability, cyber–physical resilience and blackout prevention. *Energies* **2025**, *19*, 148. [CrossRef]
41. Wang, C.; Wu, J.; Wang, J.; Zhao, W. Reliability analysis and overload capability assessment of oil-immersed power transformers. *Energies* **2016**, *9*, 43. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.