Attack-Resistant, Energy-Adaptive Monitoring for Smart Farms: Uncertainty-Aware Deep Reinforcement Learning Approach

Qisheng Zhang, *Student Member, IEEE,* Dian Chen, Yash Mahajan, Ing-Ray Chen, *Member, IEEE,* Dong Sam Ha, *Fellow, IEEE,* and Jin-Hee Cho, *Senior Member, IEEE*

Index Terms—Smart farm, Internet-of-Things, deep reinforcement learning, adversarial attacks, uncertainty.

Abstract—This work proposes an energy-adaptive monitoring system for a wireless sensor network-based smart farm using solar sensors attached to cows. The proposed system aims to achieve a high monitoring quality in a smart farm under fluctuating energy and cyberattacks disrupting the operations of collecting sensed data from solar sensors, such as protocol non-compliance, false data injection, denial-of-service, and state manipulation. We adopt *Subjective Logic*, a belief model to consider multiple dimensions of sensed data uncertainties. We employ Deep Reinforcement Learning (DRL) for agents on gateways to collect high-quality sensed data from the solar sensors. The DRL agents aim to collect high-quality sensed data with low uncertainty and high freshness under fluctuating energy levels in solar sensors. We analyze the performance of the proposed energy-adaptive smart farm system in terms of accumulated reward, monitoring error rate, and system overload. We conduct a comparative performance analysis of two proposed uncertainty-aware DRL algorithms (deep Q-learning, DQN, and multi-agent proximal policy optimization, MAPPO) against two baseline models (greedy and random) in choosing the number of sensed data to be updated to collect high-quality sensed data to achieve high resilience against attacks. We demonstrate through extensive theoretical and sensitivity analysis that MAPPO with the uncertainty maximization technique incorporated performs the best, ensuring a high monitoring quality and a low system overload.

Index Terms—Smart farm, energy-adaptive, deep reinforcement learning, solar sensors, uncertainty, cyberattacks.

1 Introduction

According to the Food and Agriculture Organization (FAO) of the United Nations [12], the food production rate would increase by a factor of up to 70 percent to absorb increase in the population, estimated as more than 9 billion people by 2050 [33]. To support farms' productivity, flexibility, or availability, smart farm technologies have developed by leveraging sensors, Internet-of-Things (IoT), edge and cloud computing technologies. Smart farm research is applied to develop agriculture business [7], improve monitoring animal welfare [36], and support the farmers for data sensing and environmental controls [36]. Smart farm research also investigated efficient data transmission considering CPU usage, signal strength, and battery operation time [18] for wireless sensor networks [28].

However, existing research lacks efforts to develop secure solutions for wireless sensor networks with energy constraints such as ones powered by solar energy harvesting. According to the World Health Organization (WHO), over half a million people died due to food contamination caused by bacteria, viruses, toxins, or chemicals. Cyber attacks on farms, transportation systems, and food processing industrial control systems to distort and disrupt the handling of correct data, can worsen the problem. Any distortion in the data received from the livestock monitoring systems can lead to a serious situation such as the spread of disease, possible pandemic, and provision of wrong information to potential customers of the livestock [14].

In this work, we are interested in improving the accuracy of the livestock monitoring system in farms in the presence of cyber attacks that can forge, modify, or drop sensed data from sensors to gateways or edge devices, or inject false data. Most wireless sensor networks (WSNs) are unable to accurately record biometrics for cattle because due to cattle's large size, battery-powered sensors (typically on three AA batteries) are attached to the collar of the cattle and can last only two to three days. Replacing such sensors and recharging their batteries every few days is an exacting task and not ideal for a typical farm. To address this problem, we consider a solar-powered sensor that can be attached to the livestock's ear. Due to the small size of the solar-powered sensor, the amount of energy harvested is low. Moreover, the amount of the harvested energy level fluctuates contin-

- *This research was partly sponsored by the National Science Foundation (NSF) under Grant Number CNS-2106987 and III-2107450. In addition, this research is also partly supported by Virginia Tech's Institute for Critical Technology and Applied Science (ICTAS) The Engineering Faculty Organization (EFO) Opportunity Seed Investment Grant and The Commonwealth Cyber Initiative (CCI) Southwest Virginia Node Research, Innovation, & Workforce Development Funding. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein (*Corresponding author: Jin-Hee Cho*). Qisheng Zhang, Dian Chen, Yash Mahajan, Ing-Ray Chen, and Jin-Hee Cho are with the Department of Computer Science, Virginia Tech, Falls Church, VA, USA. Dong Sam Ha is with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA. E-mail: {qishengz19, dianc, yashmahajan, irchen, dha, jicho}@vt.edu.*

uously as the livestock moves and the weather conditions change, all of which create uncertainty in the quality of sensed data.

In a given WSN environment, a low-energy sensor sends its sensed data to a nearby sensor with excess energy via BLE (Bluetooth Low Energy), which propagates this data to a LoRa (Long Range) gateway along with its own sensed data. The gateways upload the data to the cloud server accessible to the user. In this scenario, since there may be multiple sensors with low energy asking to transmit their sensed data to the sensor with high energy, a decision about which sensed data to transmit to the gateways can significantly impact the accuracy of monitored data. Instead of continuously transmitting sensed data for data that are already received sufficiently and show high quality (i.e., low uncertainty), sensors with high energy can select sensed data from other sensors with low energy, which request the transmission of their sensed data via BLE, to significantly increase certainty for the overall monitoring accuracy of all animals in the farm. To this end, we introduce an uncertainty-aware transmission policy based on the assessment by LoRa gateways. Specifically, a LoRa gateway can request sensors to send sensed data of certain animals whose monitored data have trended high uncertainty (i.e., low certainty). In this work, we leverage deep reinforcement learning (DRL) to identify what data need to be transmitted by sensors to improve the overall monitoring accuracy.

This work makes the following **key contributions**:

- We propose an energy-adaptive monitoring system for WSN-based smart farms with solar sensors attached to cattle. This is the first work that considers how WSN-based smart farms can maintain high monitoring quality under limited and fluctuating energy availability due to the use of solar sensors in the smart farm.
- We develop two algorithms based on *Deep Reinforcement Learning* (DRL) [11] and *Subjective Logic* [19] (SL) to identify an optimal set of sensed data of animals in a farm for maximizing the overall monitoring quality of the cattle in the smart farm while maintaining an acceptable level of energy maintained at sensors (i.e., not being overcharged or energy-depleted). More specifically, we develop uncertainty-aware DRL algorithms to minimize uncertainty in aggregated sensed data in the gateway, with the uncertainty being measured in two dimensions based on SL, i.e., *vacuity* due to a lack of evidence and *dissonance* due to conflicting evidence.
- We consider various types of cyberattack behaviors (i.e., non-compliance to the data request by a gateway, false data injection, and denial-of-service) to evaluate the robustness of the proposed uncertainty-aware DRL-based monitoring system for the smart farm.
- We validate the performance of the proposed uncertainty-aware DRL-based monitoring system using real datasets obtained from Virginia Tech's Smart Farm Innovation Network. Furthermore, we design a framework where healthy sensors generate synthetic datasets similar to real datasets and compromised sensors are modeled as attackers following the attack model for testing the robustness of our uncertainty-aware DRL-based algorithms against adversarial attacks. We conduct a comparative performance analysis of two proposed uncertainty-aware DRL-based algorithms (deep Q-learning, DQN, and multi-agent proximal policy optimization, MAPPO) against two baseline models (greedy and random) in choosing the number of sensed data to be updated to collect high-quality sensed data to achieve high resilience against attacks.

A preliminary version of the paper was published in [42]. This paper substantially extends the preliminary version with the following additional contributions:

- We newly devise a "monitoring error rate" metric that can evaluate the monitoring quality independent of monitoring data distributions. The developed monitoring error rate metric enables our proposed monitoring system to handle multi-dimensional heterogeneous data simultaneously.
- We provide mathematical proof that can justify how the SL uncertainty maximization technique contributes to reducing monitoring errors. From the theoretical analysis, we found that using the uncertainty maximization technique can lead to using more recent evidence than old evidence and thus reflecting recent network dynamics more appropriately.
- We enhance our attack model by considering the fast gradient sign method (FGSM) [13] state manipulation attack. No prior work in the literature has considered FGSM in monitoring smart farm systems.
- We identify an optimal deployment setting of LoRa gateways on which DRL agents run to maximize the chances for solar sensors to deliver their sensed data within the gateway wireless radio range.
- We provide the asymptotic complexity analysis of our proposed uncertainty-aware DRL-based algorithms. This analysis reveals a critical tradeoff between robustness/effectiveness vs. efficiency.
- We add extensive sensitivity analyses to investigate the effect of key designs and environment factors on performance, including the attack severity, the initial sensor node energy level, the number of solar sensors, and the chance for sensor nodes to be exposed to sun.

The rest of this paper is structured as follows. Section 2 provides a brief overview of the related work. Section 4 describes the network model, node model, and attack model considered in this work. Section 5 provides a detailed description of our proposed uncertainty-aware DRL-based algorithms for smart farm animal monitoring. Section 6 explains the experimental setup including datasets, parameterization of key design parameters, performance metrics, and baseline schemes considered for a comparative performance analysis. Section 7 demonstrates the key experimental results and their overall trends along with the physical interpretations of the observed results. Section 8 concludes the paper and suggests future work directions.

2 Related Work

In this section we provide a brief overview of related work in DRL-based optimization of WSNs, energy-adaptive smart environments, and uncertainty-aware smart environments.

2.1 DRL-based Optimization of WSNs

Algorithms to achieve energy-aware wireless sensor networks (WSNs) have been proposed in various WSN applications, including routing [21, 22], resource management [31], power control [3, 4, 27] and system/hybrid design [44, 41, 37]. A cluster based routing protocol was proposed based on Q-learning approach called *QL-Cluster* [21]. The QL-Cluster was designed to identify the best routes between individual nodes and remote healthcare stations to efficiently monitor a patient's health. Qi et al. [31] proposed an adaptive energy management strategy for a solar-powered WSN with hybrid storage, consisting of both super-capacitors and batteries, based on avoiding high current charging/discharge of the batteries and making full use of the supercapacitors.

Chen et al. [5] proposed a sleep scheduling algorithm for rechargeable sensors based on a DRL algorithm. The authors developed a precedence operator-based group formation algorithm to ensure the desired area coverage and a Q-leaning-based active node selection algorithm to maximize the network lifetime while achieving an acceptable coverage. Chen et al. [3, 4] leveraged DRL with Q-learning to control power for communications between the in-body sensor and Wireless Body Area Networks (WBANs) coordinator to build jamming attack-resistant, healthcare applications. Their research aimed to develop a WBAN coordinator that chooses the sensor to transmit the data in the next time slot and decides the transmitting power of these sensors, which is then sent to the sensor. The WBAN coordinator uses Q-learning to achieve an optimal power control strategy.

Based on Q-learning and the application of transfer learning for learning the Q-learning parameters (to avoid random exploration at the start of the learning process), Chen et al. [3] achieved an optimal power control strategy that can help the WBAN coordinator choose the sensor for transmitting the data along with the transmitting power for these sensors. Similarly, to address the issues of transmission reliability, energy efficiency and Quality of Service (QoS) in WBANs, Chen et al. [4] proposed a sensor access control scheme based on DRL for the WBAN coordinator to choose the access time and transmit power of the sensor based on the state of the sensor, including signal-to-interference plus noise ratio, transmission priority, battery level and transmission delay. Furthermore, to accelerate the learning process and address the high dimensionality problem due to the increasing number of sensors, a convoluted neural network (CNN) is used to estimate the Q-values according to an approximate Q-function.

Since transfer learning often raises privacy concerns in a small feature space application, Zhuo et al. [44] propose a novel federated DRL framework, called *FedRL*, to build models of high quality for agents while also preserving their privacy. The FedRL framework aims to learn a private Q-network policy for each agent by sharing limited information, which is the output of the Q-network, amongst the agents. Similarly, in wireless distributed systems, distributed DRL approaches take much more time to converge than centralized DRL counterparts. Tehrani et al. [37] proposed a Federated Learning (FL) approach to DRL, referred to as *Federated DRL (F-DRL)*. In the F-DRL, the centralized DRL model is trained by sharing the model weights of the DRL agents at the base station in a FL fashion.

Yang et al. [41] conducted a comparative performance analysis for common DRL algorithms, including Deep Deterministic Policy Gradient (DDPG), Neural Episodic Control (NEC), and Variance Based Control (VBC), for the application of wireless network optimization. Through experiments on data gathered from a large cellular network, the authors showed the potential of DDPG and VBC. However, they found the limited action space of NEC because of the large Q-value table for this particular application. Naderializadeh et al. [27] used a multi-agent deep RL approach to tackle the problem of distributed user scheduling and downlink power control in multi-cell wireless networks. Compared against several decentralized and centralized baseline counterparts, the authors showed that the proposed algorithm outperforms two decentralized approaches while performing comparably to the centralized scheduling algorithms. Moreover, the agents are trained for a specific environment. However, the approach is only scalable to different network configurations by ensuring the DNN structure does not vary.

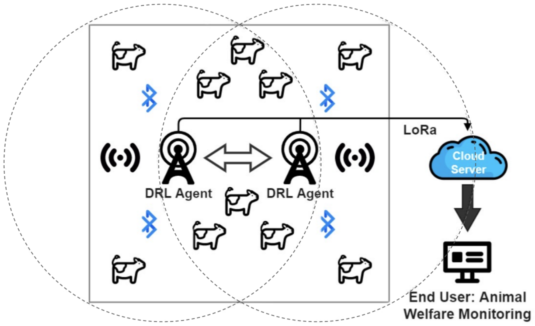


Fig. 1: Wireless Solar Sensor Node-based Smart Farm Network.

As discussed above, DRL has been applied to develop energy-efficient WSNs where privacy concerns are considered in applying deep neural networks (DNNs). However, no prior work has been done for energy-adaptive smart farms with solar sensors. We tackle this problem in this work. In particular, we develop uncertainty-aware DRL algorithms to maximize uncertainty-aware monitoring quality with sensors having limited and fluctuating energy harvesting.

2.2 Energy-Adaptive Smart Environments

Igder et al. [17] introduced an energy-adaptive Fog Server that could handle all requests at the same time under low-energy conditions with a limited number of requests. Popa et al. [29] proposed an intelligent platform and DNN-based models to achieve energy-efficient smart environments. The authors applied two techniques, namely, energy load forecasting and non-intrusive load monitoring, for learning while reducing energy consumption. In a smart home environment, the former was used to predict patterns of energy consumption and identify the unusual energy usage, while the latter was used to further find which appliance caused the anomaly to provide energy saving tips.

Modarresi and Symons [26] proposed a multidimensional framework with an instance of a high-level smart home network. They argued that the diversity of services affects routing level while providing routing service for both high and low energy consumption. Venkatesh et al.

[40] proposed a system using ultrasonic sensors with Naïve Bayes model for resilient activities tracking in a smart home. This approach enabled the system to accurately identify activities of residents, who were not seen in the training stage, without retraining the model. However, their evaluation did not include the evaluation of their approach in terms of resilience to various cyberattacks.

Relative to our work, the above cited works focused on learning and predicting energy usage to minimize energy consumption, while our work focuses on learning and deriving energy adaption strategies for LoRa gateways to obtain sensed data from solar sensors having limited and fluctuating energy, to maximize uncertainty-aware monitoring quality.

2.3 Uncertainty-Aware Smart Environments

Due to the dynamic nature of the multi-sensor smart environments, uncertainty and ambiguity can significantly introduce impact in the data prediction and monitoring of these smart environments. Zhang et al. [43] proposed learning the inhabitant's activity patterns in a smart home environment to learn under uncertainty caused by sensor malfunctions. Alemdar et al. [1] proposed uncertainty sampling-based active learning method that considers three different measures of uncertainty to select the most informative data points for activity recognition in smart homes. A number of research works [23, 2, 8, 9, 10, 15] have been conducted on uncertainty in context-aware systems where the environment is well defined. Various approaches have been proposed to model uncertainty, including semantic web [23], game theory [8], vector space model [30], asymptotic equipartition property (AEP) [34], signal processing and information-theoretic techniques [39], and Moore finite state machine (FSM) [32]. Machado et al. [23] proposed a contextual reference model based on semantic web to deal with uncertainty in a smart environment. Almeida and de Ipiña [2] developed an ontology for context-aware systems in smart environments to consider both ambiguity and uncertainty. To improve the accuracy of probabilistic inference systems for multi-sensor data-fusion, De Paola et al. [9] suggested that context information be included to prevent the increase of uncertainty and pointed out that the right mix of context information is fundamentally important. Rocher et al. [32] proposed a framework for estimating behaviour drift in smart-X systems at runtime. They leveraged Moore finite state machine (FSM) model combined with control theory and validated their approach based on a real dataset to ensure effectiveness and efficiency.

Unlike the above cited works, we consider multiple types of uncertainty and the uncertainty maximization technique in Subjective Logic [20] for monitoring data updates based on new evidence, thus maximizing monitoring quality while minimizing energy consumption.

3 Problem Statement

In this work, we aim to minimize monitoring error rate (i.e., a gap between the sensed data aggregated from sensors and the ground truth; see Eq. (14)) and system overload (i.e., a mean fraction of the failed requests of all requests sent from low-energy sensors; see Eq. (15)) in a sensor network by identifying an 'optimal policy'. An update policy $\mathcal{P}_T = \{p_1, p_2, \ldots, p_T\}$ contains monitoring actions p_i, where $i \in [1, T]$, $p_i \in \mathcal{P}_T$ and $\mathcal{P}_T$ is a set of monitoring actions available to the sensor in every monitoring step. Given a dynamic sensor network $\mathcal{G}_T = \{g_1, \ldots, g_i, \ldots, g_T\}$, the objective function is defined as follows:

$$\arg\max_{\mathcal{P}_T} \sum_{i=1}^{T} f(g_i(p_1, p_2, \ldots, p_i)), \quad s.t. \quad \forall i \in [1, T], p_i \in \mathcal{P}_T, \tag{1}$$

where $f(g)$ is based on the evaluation function $f : g \mapsto -\mathcal{ME}(g) - \mathcal{OL}(g)$, aiming to minimize the monitoring error rate $\mathcal{ME}$ and system overload $\mathcal{OL}$, to be detailed in Section 6. Determining an optimal update policy that aims to achieve multiple objectives is non trivial given the complexity involved in solving a multi-objective optimization problem [6]. This is discussed in detail based on experimental results discussed in Section 7.

4 System Model

This section discusses the network, node, and attack models.

4.1 Network Model

Our target WSN consists of sensors attached to the cattle, that continuously measures the bio-metric information and transmits the sensed data to the LoRa gateway, which then aggregates and transmits the clustered data to the cloud. Given the relative low cost of transmitting data over long ranges (LoRa) via the standard IP protocol for IoT devices, the LoRa gateways act as the optimal intermediary between the sensor nodes and the cloud server. In the given smart farm network (see Fig. 1), using BLE (Bluetooth Low Energy) a low-battery sensor (LBS) can relay its sensed data to a high-batter sensor (HBS). The high-battery sensor can then send the received sensed data along with its own data to LoRa gateway via LoRa. We assume that each sensor has a Microchip SAM R34/35 microcontroller with an embedded LoRa radio which dissipates 170 mW during transmission, while the microcontroller itself dissipates only 8 mW in active mode. The LoRa protocol is ideal for long distance communication with a distance coverage of several kms and a data rate of 27 $kbps$. Contrarily, the BLE protocol is purposed for short distance communication with a radius coverage of 100 $meters$ and a data rate of 2 $Mbps$. Furthermore, the BLE protocol drains considerably less power than the LoRa protocol. For example, only 11 mW of power is dissipated during transmission for a Texas Instruments CC2640R2F micro-controller chip with an embedded BLE radio [38]. Therefore, sending a single bit of data takes about 1,100 times lesser energy for the Texas Instruments CC2640R2F micro-controller chip when compared with the LoRa radio of SAM R34/35 microcontroller chip. We assume that the initial battery level of each deployed sensor is 5 kWs which is equivalent to full charge. Outdoor solar has a power density of about 10 mW/cm^2 whereas indoor light has a power density of 0.1 mW/cm^2 [24]. For a solar panel of 5cm, the maximum harvestable power for indoor light is about 2 mW and outdoor light is about 200 mW.

Fig. 1 describes the considered network model in this work describing a smart farm environment with solar sensors attached to cows.

With the main objective of minimizing the monitoring error rate and system overload, a DRL agent is deployed at every LoRa gateway to shortlist, select and prioritize which animal's sensed data is required, at regular intervals. The process of identifying the important data, by the DRL agent is described in Section 5. We assume that for energy saving there is no encryption when the sensors communicate with each other via BLE and hence, malicious entities can intercept the data in transmission and modify/forge data. Additionally, an attacker can imitate a sensor by using its authentication key with the gateway and sending false data for the sensor itself as well as for other low battery sensors. We assume that the communication between the LoRa gateway and the cloud server is secure and encrypted based on traditional secret cryptography. As shown in Fig. 1, multiple LoRa gateways each running a DRL agent can collaborate to each other in sharing collected sensed data received from sensors.

4.2 Node Model

Sensor nodes in a given smart environment are assumed solar-powered and deployed as implants and can transmit data on request. Depending upon the animal's movement and the varying weather condition from day to day, the battery levels of the sensor may fluctuate throughout the day. Therefore, it is essential to utilize the energy wisely for high availability, consistency, and sustenance. Each sensor node i is characterized by $\mathrm{sn}_t^i = [\mathrm{temp}_t^i, \mathrm{hb}_t^i, \mathrm{ma}_t^i, \mathrm{bl}_t^i]$, where temp_t^i refers to sensor node i's temperature at time t in Celsius, hb_t^i is the number of i's heartbeat at time t, ma_t^i is i's speed at time t and bl_t^i is i's battery life at time t scaled in $[0, 100]$ in percent. Most of the sensor's energy will be used to transmit the sensed data to the LoRa gateway. In contrast, considerably less battery will be used for communication between the sensor and other nearby sensors via BLE as it consumes roughly 1100 times less (see Section 4.1).

Utilizing the data reported by the sensors to the LoRa gateway, each DRL agent will try to maximize the monitoring quality by selecting what data is needed with priority to accurately estimate the well-being of all the animals on the farm. To this end, each sensor node in the WSN is categorized into high battery-level sensor (HBS) and low battery-level sensor (LBS) based on the recommended battery level T_M. Since we are only interested in transmissions from LBS to HBS, we model the sensor network as a directed bipartite graph. Section 5 describes the actions performed by the DRL agent running on every LoRa gateway. The end-user will get the efficient monitoring results of the smart farm from the cloud server, which aggregates data on individual animals from various LoRa gateways. This work aims to evaluate how the DRL agent on LoRa gateways can enhance the quality of animal monitoring in the presence of cyberattacks and fluctuating sensor energy levels.

4.3 Attack Model

This work considers the following attack behaviors:

- *Non-compliance to the protocol*: A sensor node can be compromised and exhibit non-compliant behavior to the request by the DRL agents on LoRa gateways. For example, when an animal A's sensed data is requested by a DRL agent, the attacker can either not send A's data or send other sensor's data to the LoRa gateway. We model this by considering the attacker's non-compliance probability, denoted by P_{NCA}.
- *False data injection*: An attacker (e.g., a compromised sensor) can transmit forged/modified data or inject false data to gateways. In addition, man-in-the-middle attackers (MIMAs) can intercept data being transmitted in the middle and replace it with forged/modified data. The attackers can inject false data during the training phase (i.e., poisonous attacks) or the testing phase (i.e., evasion attacks). We call the compromised sensors 'internal attackers' while calling the external attackers intercepting sensed data for forgery/modification or injecting false data 'external attackers'. These attacks are modelled based on the probabilities of forging/modifying data by an internal or external attacker, denoted by P_{IDA} and P_{EDA}, respectively.
- *Denial-of-Service (DoS)*: A compromised high-battery sensor can send a request to nearby sensors requesting them to send its faked sensed data. As it is a type of internal attacks, we also model this DoS attack probability by P_{IDA}. This can make other sensors' energy levels drained quickly but wasted in sending false data. Note that to avoid an infinite loop, we assume the attacker will request sending its faked sensed data to legitimate sensors, not compromised sensors.
- *Fast Gradient Sign Method (FGSM)*: This state manipulation attack model is firstly proposed in [13] to generate adversarial examples in image classification tasks. To apply it in the context of DRL-based algorithm execution, we generalize DRL settings by considering actions as class labels. To make a fair comparison, we use the original loss function of each DRL algorithm to compute the gradients. Since we have multiple DRL agents in our smart farm system, we assume the attack will happen in both local agent states and global agent observations. We define the perturbation strength as P_{FGS}.

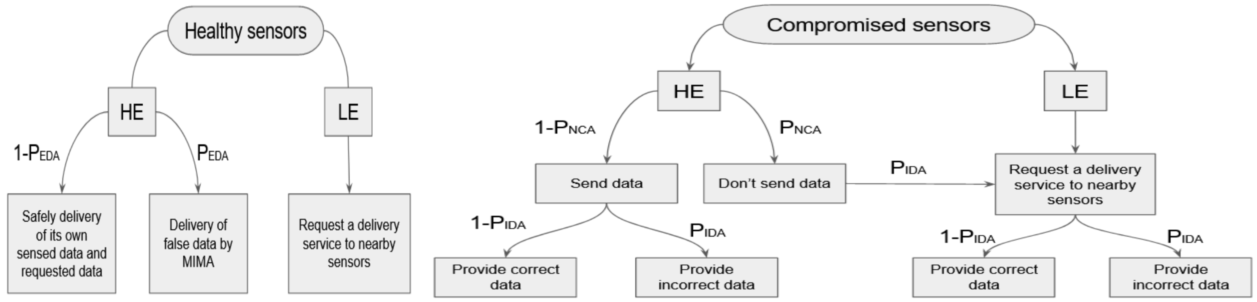
We summarize the above four types of attacks in Fig. 2.

5 DRL-based, Uncertainty-Aware Animal Monitoring

In this section we provide a detailed description of our proposed uncertainty-aware DRL-based algorithms for smart farm animal monitoring

5.1 Uncertainty-Aware Animal Monitoring

First, based on received data from sensors in the past, a gateway can estimate uncertainty in each animal's condition, such as heart rate, average temperature, minimum/maximum temperature, average activity, battery level of a sensor worn by the animal, and timestamp. Recall that a sensor with low energy will send its sensed data to a nearby sensor having high energy, which include sensors attached to animals within 100 meters. Hence, we distinguish direct sensed data from indirect sensed data in terms



(a) Outsider attacks by

(b) Insider attacks by false data injection, DoS, or non-compliance

Fig. 2: Attack scenarios by both outsider and insider attacks. Recall that P_{EDA} is the probability of an external attacker performing external data attacks, P_{NCA} is the probability of an internal attacker performing non-compliance attack, and P_{IDA} is the probability of an inside attacker performing false data injection attacks.

of whether a sensor sent its own sensed data or other sensor's sensed data. If the sensor with high energy transmitted other sensor's sensed data, it is treated as indirect sensed data. Otherwise, it is direct sensed data. The gateway periodically reports its collected data from sensors to the cloud server. Note that since the given network has multiple gateways, the corresponding multiple DRL agents will share information about sensed data and estimate each animal's conditions (see Table 1) and associated confidence level on each observation item (i.e., HR, average temperature, average activity, etc.). A database on the gateway keeps recording for all animals' condition data where sensed data by sensor i (i.e., ID) for an animal are stored.

Each observation item's condition (e.g., average temperature) will be reported as one of the K classes of the range, e.g., for the temperature reading, $K = 5$ meaning that there are 5 classes of ranges: 35 or below, 36-37, 38-39, 40-41, 42 or above. The end user can then easily determine if the temperature is normal based on the cloud server's received data. Since the gateway will periodically report average conditions for all animals to the cloud, it will aggregate sensed data from sensor nodes and measure their average with the probability of a condition being with K classes and multiple types of uncertainty values. To utilize the concept of uncertainty, we will apply Subjective Logic (SL) [20] to compute an opinion on each animal's condition in a given attribute (e.g., temperature, heart beats, activity, or battery level).

5.1.1 SL-based Formulation of a Multinomial Opinion

SL can explicitly express uncertainty caused by a lack of evidence, called *vacuity* in its opinion representation. In addition, SL can consider base rates as prior probabilities in a Bayesian way to formulate a second-order opinion and corresponding uncertainty estimates, where a second-order opinion is represented by Dirichlet distribution. We will use a Dirichlet probability density function (PDF) to model the distribution of class probabilities and corresponding uncertainty masses. In SL, a multinomial opinion in a given proposition x (e.g., an animal condition in our smart farm context, such as a temperature or heart beats) is represented by $\omega_X = (\boldsymbol{b}_X, u_X, \boldsymbol{a}_X)$ where a random variable $X \in \mathbb{X}$ (a subject domain) and $K = |\mathbb{X}| > 2$ and the additivity requirement of ω_X is given as $\sum_{x \in \mathbb{X}} \boldsymbol{b}_X(x) + u_X = 1$. To be specific, each parameter indicates,

- $\boldsymbol{b}_X$: *belief mass distribution* over $\mathbb{X}$;
- u_X: *uncertainty mass* representing *vacuity of evidence*;
- $\boldsymbol{a}_X$: *base rate distribution* over $\mathbb{X}$.

The projected probability distribution of multinomial opinions is given by:

$$\boldsymbol{P}_X(x) = \boldsymbol{b}_X(x) + \boldsymbol{a}_X(x) \cdot u_X, \quad \forall x \in \mathbb{X} \tag{2}$$

The base rate for belief $\boldsymbol{b}_X(x_i)$, which is $\boldsymbol{a}_X(x_i)$, means the prior preference over the x_i belief (e.g., a class). If no preference is given, we consider the base rate equally for each belief mass, i.e., $\boldsymbol{a}_X(x_i) = 1/K$ for any x_i.

Given the amount of evidence supporting belief x_i is $\boldsymbol{r}(x_i)$, the observed evidence in the Dirichlet PDF can be mapped to the multinomial opinions as:

$$\boldsymbol{b}_X(x) = \frac{\boldsymbol{r}(x)}{W + \sum_{x_i \in \mathbb{X}} \boldsymbol{r}(x_i)}, \; u_X = \frac{W}{W + \sum_{x_i \in \mathbb{X}} \boldsymbol{r}(x_i)}, \tag{3}$$

where W refers to the amount of uncertain evidence. Commonly, W is set to the number of belief masses (i.e., $W = K$).

5.1.2 Estimation of Multiple Types of Uncertainty

SL categorizes uncertainty into two primary sources [20]: (1) basic belief uncertainty derived from single belief masses, and (2) intra-belief uncertainty based on the relationships between different belief masses. These two sources of uncertainty can categorize the two uncertainty types, *vacuity* and *dissonance*, respectively, that correspond to vacuous belief and contradicting beliefs. In particular, vacuity of an opinion ω_X is captured by uncertainty mass u_X while dissonance of an opinion, $\boldsymbol{b}_X^{\text{Diss}}$, is formulated by [19]:

$$\boldsymbol{b}_X^{\text{Diss}} = \sum_{x_i \in \mathbb{X}} \left(\frac{\boldsymbol{b}_X(x_i) \sum\limits_{x_j \in \mathbb{X} \setminus x_i} \boldsymbol{b}_X(x_j) \text{Bal}(x_j, x_i)}{\sum\limits_{x_j \in \mathbb{X} \setminus x_i} \boldsymbol{b}_X(x_j)} \right), \tag{4}$$

where the relative mass balance between a pair of belief masses $\boldsymbol{b}_X(x_j)$ and $\boldsymbol{b}_X(x_i)$ is expressed by:

$$\text{Bal}(x_j, x_i) = 1 - \frac{|\boldsymbol{b}_X(x_j) - \boldsymbol{b}_X(x_i)|}{\boldsymbol{b}_X(x_j) + \boldsymbol{b}_X(x_i)}. \tag{5}$$

The dissonance estimation is useful to measure the 'inconclusiveness' of an opinion even under a large amount of evidence which almost equally support each singleton belief.

In this work, we regard each reported data from sensor nodes to a gateway as evidence. For instance, in a temperature report, if 38 C is reported, b_2 (i.e., b_1 = lower than normal, b_2 = normal, b_3 = higher than normal) should be updated based on Eq. (3). When uncertainty becomes zero, an opinion will not be updated anymore, which makes new evidences cannot be properly utilized in the latest opinion. To avoid this, we will deploy the *uncertainty maximization* technique [20] to reduce the impact of conflicting evidence while transforming the amount of conflicting evidence into vacuity of an opinion.

Given opinion $\omega_X = (\boldsymbol{b}_X, u_X, \boldsymbol{a}_X)$ where $\boldsymbol{P}_X = \boldsymbol{b}_X + \boldsymbol{a}_X \cdot u_X$ for a domain $\mathbb{X}$, the corresponding vacuity-maximized opinion is denoted by $\ddot{\omega}_X = (\ddot{\boldsymbol{b}}_X, \ddot{u}_X, \boldsymbol{a}_X)$ where $\ddot{u}_X$ and $\ddot{\boldsymbol{b}}_X$ are computed by:

$$\ddot{u}_X = \min_i \Big[\frac{\boldsymbol{P}_X(x_i)}{\boldsymbol{a}_X(x_i)}\Big], \tag{6}$$

$$\ddot{\boldsymbol{b}}_X(x_i) = \boldsymbol{P}_X(x_i) - \boldsymbol{a}_X(x_i) \cdot \ddot{u}, \text{ for } x_i \in \mathbb{X}.$$

Note that we use a threshold ρ to trigger the vacuity maximization. That is, Eq. (6) above can be triggered only when $u_X < \rho$ where ρ is sufficiently low (e.g., 0.05). The purpose of updating ω_X to $\ddot{\omega}_X$ is to allow the opinion to be further updated by receiving new evidence or being combined with other opinions which are possible only when $u_X > 0$.

In a given category X, the animal condition is estimated as an opinion, ω_X where the corresponding uncertainty types, vacuity and dissonance, are estimated, respectively, at the gateways that aggregate sensed data and transmit the average condition value in a given category along with the belief masses and uncertainty masses associated with ω_X.

5.2 DRL-based Monitoring Update

5.2.1 State Space ($\mathcal{S}_t$)

We assume a partially observable environment where each DRL agent can only access information in the dataset of the local gateway it is running on. We formulate the state space of each agent i at time t indicating the total number of local reports for the duration of t, $\{\ell_1^i, \ell_2^i, \ldots, \ell_{t-1}^i, \ell_t^i\}$, where $\ell_{t^*}^i$ refers to the number of local reports in $[t^*-1, t^*]$. To be specific, assume $t \in [0, T]$, the overall state space is given by $\mathcal{S}_t = \{s_t^1, \ldots, s_t^i, \ldots, s_t^m\}$, where m is the number of DRL agents (i.e., LoRa gateways) and s_t^i is the state space for agent i at time t, which is given by $s_t^i = \{\ell_1^i, \ell_2^i, \ldots, \ell_t^i, 0, \ldots, 0\}$.

5.2.2 Action Space ($\mathcal{A}_t$)

For each DRL agent, it will choose k animals whose data is more helpful in terms of improving monitoring quality and reducing system overload. Note that a certain amount of redundant information is desired since there is a possible situation that sensors fail to transmit data due to limitations of their energy level or topology. Based on agent i's local gateway dataset, the utility of animal j is given by:

$$\text{utility}_{ij} = (1 - \text{vac}_t^{ij}) + (1 - \text{diss}_t^{ij}) + \text{fr}_t^{ij} + f(\text{bl}_t^{ij}), \tag{7}$$

where vac_t^{ij}, diss_t^{ij}, fr_t^{ij} are vacuity, dissonance, and degree of freshness of animal j's sensed data at time t by DRL agent i. To calculate vacuity and dissonance, we use each evidence (i.e., a report of animal conditions from sensor node) to hold a categorical class (i.e., below normal range, normal range, above normal range). We initialize the opinion for a given animal with one evidence (i.e., $\mathbf{r}(x_l) = 1$) for each class l and $K = 3$. fr_t^{ij} is formulated by $\text{fr}_t^{ij} = e^{-\phi t}$, where t is time elapsed from the last update and ϕ is a constant to normalize the freshness. $f(x)$ is defined by $f(x) = -(x - T_M)^2$ where x is set to bl_t^{ij}, the battery life of sensor j at time t by DRL agent i. By scaling $\text{vac}_t^{ij}, \text{diss}_t^{ij}, \text{bl}_t^{ij}$, and fr_t^{ij} in $[0, 1]$, we set each component of utility_{ij} to $[0, 1]$ as a real number. Here T_M denotes the recommended level that the battery of a sensor node should be maintained in, which always have some remaining energy or not too full under sun. A list of animal IDs will be calculated based on Eq. (7) in ascending order, so each agent will request data for top k animals. The action space has three actions selecting first k animal IDs such that $k \in \{0, \lfloor \frac{n}{2} \rfloor, n\}$, where n is how many LBS in the current environment. In this way, the size of action space is not dependent on n (i.e., 3), which is able to reduce the computation load raised by infinite action spaces and make this monitoring system possible for applying to larger-scale sensor networks as generalization. For a lower k, it may impact the monitoring quality. At the same time, a higher k will obtain a larger amount of unnecessary data transmission and result in the system overload. Therefore, the DRL agent aims to identify a best action, which is the optimal k for this setting.

5.2.3 Immediate Reward (r_t)

This is formulated by $r_t^i = f(g_t(k_1, k_2, \ldots, k_t))$ based on $f(g_t) = -\mathcal{ME}(g_t) - \mathcal{OL}(g_t)$ given in Eq. (1), where k_i is an action taken in step i.

5.3 Data Aggregation at LoRa Gateways

For each sensor node, it will send its sensed data to LoRa gateways or a high-energy sensor close to it, as the information shown in Table 1. After receiving the reports from all sensor nodes capable of transmitting their data, a LoRa gateway will compute an opinion based on the received data. The opinion is composed of belief and uncertainty in terms of vacuity and dissonance masses. We define the opinion a *monitoring opinion* (MO) below, of measured temperature, heart beats, and activity in the period of monitoring time Δ. Δ is computed based on the time interval of current time and last reported time. When the amount of sensed data gets large enough, the MO may not be updated further because vacuity approaches closely to zero, which an opinion cannot be updated from new sensed data based on Eq. (3). To update the MO properly from received new evidence, we will deploy the vacuity (uncertainty) maximization technique in Eq. (6). We will use a threshold ρ (i.e., $0 < \rho < 1$) to determine when to update the MO based on Eq. (6). That is, if $u_X < \rho$, this evidence will update an opinion based on Eq. (6). Animal's average conditions will be computed from multiple sensed data and MOs gained based on different sensor nodes at time t by the DRL agent on each gateway. As opinions are independent of each other, we can compute the joint opinion using multinominal multiplication technique in SL [20]. The resulting sensed data in temperature, activity, and heart beat will be their average value while the resulting MO will be represented by $\omega = \{\boldsymbol{b}, u, d, \boldsymbol{a}\}$ where a gateway receives sensed data packets and it aggregates them to formulate an MO from multiple, different sensors, $\boldsymbol{b}$ is a vector of belief masses, u is a vacuity mass, d is a dissonance mass, and $\boldsymbol{a}$ is a vector of base rates corresponding to $\boldsymbol{b}$.

5.4 Mathematical Proof of Effectiveness Using Uncertainty Maximization

In this section, we formally prove the effectiveness of the *uncertainty maximization* technique [20] by a mathematical proof. We observe that uncertainty measures and monitoring error rate can be viewed as functions of evidence. Based on Eq. (3), the uncertainty (*vacuity*) drops when the amount of received evidence increases. Furthermore, given Eq. (4), *dissonance* solely depends on the distribution of belief masses without vacuity being involved. Thus, the dissonance can be viewed as a constant when there is enough evidence from the same distribution.

Note that the uncertainty (vacuity) maximization in Eq. (6) reinitializes the vacuity by transforming previous evidence from the belief masses to the uncertainty mass. Given the following [20],

$$j = \arg\min_{i} \Big[\frac{\boldsymbol{P}_X(x_i)}{\boldsymbol{a}_X(x_i)}\Big], \tag{8}$$

where $\boldsymbol{P}_X(x_i)$ is the projected probability of having x_i and $\boldsymbol{a}_X(x_i)$ is the base rate (i.e., prior belief) that support a belief mass x_i. Then, we have the updated belief masses and the uncertainty (vacuity) mass given by:

$$\ddot{\boldsymbol{b}}_X(x_k) = \frac{\boldsymbol{r}(x_k) - \boldsymbol{r}(x_j)}{W + \sum_{x_i \in \mathbb{X}} \boldsymbol{r}(x_i)}, \ddot{u}_X = \frac{W + K\boldsymbol{r}(x_j)}{W + \sum_{x_i \in \mathbb{X}} \boldsymbol{r}(x_i)}, \tag{9}$$

where W is an amount of non-informed evidence (i.e., uncertain evidence), K is the number of belief masses (e.g., classes), and $\boldsymbol{r}(x)$ is the amount of evidence supporting belief mass x. Since the amount of non-informed evidence, W, increases to $W + K\boldsymbol{r}(x_j)$, we replace W with $W + K\boldsymbol{r}(x_j)$ in the denominators of both $\ddot{\boldsymbol{b}}_X(x_k)$ and $\ddot{u}_X$ as

$$\ddot{\boldsymbol{b}}_X(x_k) = \frac{\boldsymbol{r}(x_k) - \boldsymbol{r}(x_j)}{(W + K\boldsymbol{r}(x_j)) + \sum_{x_i \in \mathbb{X}}(\boldsymbol{r}(x_i) - \boldsymbol{r}(x_j))}, \tag{10}$$

$$\ddot{u}_X = \frac{W + K\boldsymbol{r}(x_j)}{(W + K\boldsymbol{r}(x_j)) + \sum_{x_i \in \mathbb{X}}(\boldsymbol{r}(x_i) - \boldsymbol{r}(x_j))}. \tag{11}$$

The above implies that the updated vacuity only considers partial history evidence, indicating more recent evidence than the past evidence.

TABLE 1: EVD Dataset Description

Metric	Description
serial	A unique animal identifier
HR	Heart Rate of the animal
average-temperature	Average body temperature in Celsius
min-temperature	Minimum temperature in Celsius
max-temperature	Maximum temperature in Celsius
average-activity	Average activity recorded by the number of steps taken
battery-level	Residual battery life
timestamp	Date and time of transmission

As shown in Eq. (6.4), the monitoring error rate is closely related to the amount of evidence. Assume that at each time step $t \in [0, T]$, the information of n_t animal is being updated and each animal j's information is updated m_j times in total. Hence, we have the expected monitoring error rate given by,

$$\begin{aligned} E(\mathcal{ME}) &= \frac{E(\sum_{t \in T} \sum_{x \in X} \mathrm{me}_t^x)}{NT|X|} \\ &= \frac{\sum_{t \in T} \sum_{x \in X} E(\sum_{j=1}^{n} D(\mathrm{eo}_t^x(j), \mathrm{gt}_t^x(j)))}{NT|X|} \\ &= \frac{\sum_{x \in X, t \in T} E(\sum_{j=1}^{n_t} 0 + \sum_{j=n_t+1}^{n} 1)}{NT|X|} \\ &= \frac{\sum_{x \in X} \sum_{t \in T}(N - n_t)}{NT|X|} \\ &= 1 - \frac{\sum_{x \in X} \sum_{j=1}^{n} m_j}{NT|X|}. \end{aligned} \tag{12}$$

Here T is the total monitoring time, N is the number of solar sensors attached to cows, me_t^x is the overall monitoring error rate of all N cows' conditions of attribute x at time t, $\mathrm{eo}_t^x(j)$ and $\mathrm{gt}_t^x(j)$ is the estimated and ground truth observation of cow j's condition in x attribute at time t, respectively. The above derivation proves that $E(\mathcal{ME})$ is only dependent upon n_t and it increases when n_t decreases. In addition, each animal j's monitoring error rate is only related to the amount of corresponding evidence, m_j. Therefore, we prove that the order of two sensors remains invariant under the partial order relations of vacuity and monitoring error rate.

6 Experimental Setup

This section describes the datasets, parameterization, DRL schemes, and performance metrics used for the experiments conducted in this work.

6.1 Datasets

At Virginia Tech, we have a collection of interconnected data collection and analysis hubs called the *SmartFarm Innovation Network* (TM), which is designed to facilitate testing and demonstration of emerging technologies throughout the state. From the smart farm, we obtained sample datasets collected from four different sensors, namely, EmbediVet Implantable Temperature Device (EVD), Halter Sensor, Heart Rate Sensor, and Implantable Temperature Sensor. The dataset from the EVD consists of 8 components as descried in Table 1. We consider 6 components out of them, except a serial number and timestamp, as the sensed data to represent the physical conditions of animals. The temperature and the heart rate sensor provide us with temperature in Celsius and heart rate in beats per minute (bpm), respectively, where the Halter sensor helps in geolocation and assessing the animal's motion and posture to report on their activity level.

Since the existing dataset obtained from Virginia Tech's smart farm does not include any data compromised by attackers, we designed a framework where each sensor could generate synthetic datasets similar to real datasets and some compromised sensors are modeled as attackers following the attack model described in Section 4.3.

6.2 Parameterization

We consider 20 cows within a 40 acres ($\sim$ 160K square meters) square farm area (A) with 402 meters in length (a). We consider two gateways with the same circular coverage. We further assume that these gateways could cover the farm area and each of them is covered by the other. In general, for a given number of m gateways with the same radius r, we aim to find the minimum radius r_m such that the farm area is fully covered by the total gateway coverage and each gateway is covered by other gateways to enable mutual communications. To this end, we solve the following optimization problem to identify the minimum radius r_m:

$$r_m = \min_r \{r : \exists P_i = (x_i, y_i) \in \mathbb{R}^2\}, \tag{13}$$

$$\text{where } 1 \leq i \leq m, \;\; s.t. \;\; \forall P = (x, y) \in [-a/2, a/2]^2,$$

$$(\min_i d(P, P_i) \leq r) \wedge \forall (i, j) \in [1, m]^2, \;\; \mathrm{d}(P_i, P_j) \leq r,$$

where a is the length of farm side, P_i is the center of gateway, and $\mathrm{d}(\cdot, \cdot)$ is the Euclidean distance function. We only consider the case when $m = 2$, where gateways locate in $(-\frac{a}{4}, 0)$ and $(\frac{a}{4}, 0)$ respectively with the same radius $\frac{\sqrt{5}a}{4}$. Fig. 1 shows how two gateways are optimally deployed with the corresponding wireless radio ranges used in our smart farm network environment.

To model the availability of solar energy based on sun's movement in a day, we define a charging probability distribution $P(x, y, t)$ over the farm area as the probability of being charged if a sensor is located at (x, y) at time t. For simplicity, we assume that $P(x, y, t)$ has a quadratic form at time t and is represented by $P(x, y, t) = \max\{0, -\frac{1}{6}(t - t_{xy})^2 + 1\}$, where t_{xy} is a function of location (x, y) based on the farm's direction. We consider a square farm with its center at the origin and x axis towards west. Thus, t_{xy} is formulated as $t_{xy} = \frac{t_0}{a} \times (x - \frac{a}{2}) + 12$, where t_0 is a hyper-parameter. In general, to model different weather conditions, we can use a weight α to discount the charging probability as $\alpha P(x, y, t)$ with $0 \leq \alpha \leq 1$.

Each cow's attributes are collected by an attached solar-powered sensor. We adopt normal distributions $\mathcal{N}(38,\, 1^2)$ and $\mathcal{N}(1.5,\, 0.1^2)$ to describe a cow's temperature and velocity respectively. The cow's heart beat is modeled as two uniform distributions: $\mathcal{U}(60, 84)$ when it moves or $\mathcal{U}(48, 60)$ when it does not move. We use P_i^{mv} for cow i's moving probability.

For an opinion about a cow's attributes, we will simply categorize based on three beliefs, i.e., lower than normal, normal, and higher than normal. The normal ranges of a cow's temperature, heart rate, and moving activity are given $[37.8, 39.2]$ Celsius, $[48, 84]$ number of beats per min., and $[1, 2]$ meters per sec., respectively. We consider the number of uncertain evidence being three where each belief mass has the same base rate (i.e., 1/3) [20].

TABLE 2: Key Design Parameters, Their Meanings and Default Values

Param.	Meaning	Value
m	The number of gateways	2
N	The number of sensors(cows)	20
T_M	A minimum battery level to transmit sensed data by a sensor	30%
LBS/HBS	Low/High battery level sensors	/
P_i^{mv}	Cow i's probability to move	$[0.3, 0.7]$
P_A	Probability for an attacker or a compromised node to perform a certain attack (e.g., P_{NCA}, P_{IDA}, P_{EDA}, P_{FGS})	0.1
A	Area of a given smart farm	40 acres
a	length of a given smart farm	402 m
ρ	Uncertainty maximization threshold	0.05
t_0	Hyper-parameter used in sun model	0.2
T_u	Time interval for a sensor to send sensed data	30 s
T_a	Time interval for a gateway to take an action to adjust k	60 s
T_L	Initial battery level for low battery level sensors	30%
α	The probability for a cow to be exposed to sun depending on its position when sun is available	1

We consider 24 consecutive hours as the total monitoring session. For every $T_a = 60$ sec, each gateway would take an action to identify an optimal monitoring strategy. We assume 5 HBS with full initial battery level and 15 LBS with random initial battery levels below T_M. All LBS could only broadcast their own data to HBS via BLE. Each sensor can broadcast at most two sets of sensed data to each LoRa gateway within the wireless range per $T_u = 30$ sec. In this way, each HBS can send its own data and another set of data requested by LBS. The monitoring system would derive the consolidated priority list of update lists from gateways. Then the system would leverage the Hopcroft–Karp algorithm [16] to solve the maximum matching problem in bipartite sensor networks. In this way, the system could ensure the maximum number of transmissions being executed.

As for the energy consumption, message transmissions need 170 mW per sec and the sleep mode costs 2 mW per sec. We assume a sensor is only activated for message transmissions. A sensor can be charged by the outdoor solar with 200 mW per sec. In this way, a sensor can be charged 200 mW × 6 h = 4.32 kWs under 6 hours of sun.

We set all attack probabilities to P_A. For inside attackers, we initially pick them among the total number of sensors at random. For outside attackers (i.e., MIMAs), we pick a set of nodes at random transmitting messages intercepted by MIMAs with $P_{EDA}(P_A)$. The attackers launch attacks based on Fig. 2. Finally, we consider FGSM as a state manipulation attack to disturb the DRL training phase. Table 2 summarizes the key design parameters, their meanings, and default values.

6.3 DRL Algorithms

We develop two uncertainty-aware DRL algorithms (MAQN and MAPPO) whose performance is compared against two baseline schemes (Greedy and Random), described as follows:

- **Multi-Agent Deep Q-Learning (MADQN)** [25]: DRL agents learn a state-value Q function to select the optimal actions. In a multi-agent scenario, we extend DQN to MADQN, where each DRL agent learns an independent local Q function. We consider two variants of MADQN using UM or not and name them MADQN-UM and MADQN-NUM, respectively.
- **Multi-Agent Proximal Policy Optimization (MAPPO)**: MAPPO extends the PPO [35] to a multi-agent environment to mitigate non-stationarity by adopting a global critic value function to guide each local actor value function. We consider two variants of MAPPO with and without using uncertainty maximization. We name them MAPPO-UM and MAPPO-NUM, respectively.
- **Greedy Algorithm**: DRL agents make heuristic choices by enumerating all actions at each step and choosing the one with the optimal reward.
- **Random**: DRL agents randomly select an action, e.g., k animal IDs to receive their sensed data.

6.4 Metrics

We consider the following metrics to evaluate the performance of the four DRL schemes described in Section 6.3:

- **Accumulated reward ($\mathcal{R}$)**: This metric represents the sum of mean accumulated reward for all DRL agents through all simulation runs.
- **Monitoring error rate ($\mathcal{ME}$)**: This metric is measured based on the mean difference between the aggregated data of each animal's condition at each gateway and the ground truth data of the corresponding animal's condition. We measure $\mathcal{ME}$ by:

$$\mathcal{ME} = \frac{\sum_{t \in T} \sum_{x \in X} \mathrm{me}_t^x}{NT|X|}, \quad \mathrm{me}_t^x = \sum_{j=1}^{n} D(\mathrm{eo}_t^x(j), \mathrm{gt}_t^x(j)), \tag{14}$$

$$s.t. \quad D(a,b) = \begin{cases} 1 & \text{if } a \neq b; \\ 0 & \text{if } a = b. \end{cases}$$

  where T is the total monitoring time, N is the number of animals, me_t^x is the overall monitoring error rate of all N animals' conditions of attribute x at time t, $\mathrm{eo}_t^x(j)$ and $\mathrm{gt}_t^x(j)$ are the estimated and ground truth observation of animal j's condition in x attribute at time t, respectively.
- **Overload ($\mathcal{OL}$)**: This metric evaluates the system overload by the mean fraction of the failed requests over all sent requests from LBS. In specific, we have

$$\mathcal{OL} = \frac{1}{T} \sum_{t \in T} \frac{rq_t^f}{rq_t}, \tag{15}$$

  where T is the total monitoring time, rq_t^f and rq_t are the numbers of failed requests and total requests at time t, respectively.

7 Experimental Results and Analysis

7.1 Algorithmic Complexity Analysis

We first analyze the algorithmic complexity of the four DRL schemes described in Section 6.3. Table 3 shows the asymptotic complexities in Big-O notation for the four schemes discussed in Section 6.3. We notice that cost of MADQN/MAPPO only depends on the training episode n_e and training time per episode t_{train}. Greedy needs to enumerate the total action space and thus its complexity depends on the size of action space n_{action}. When the action space is large enough, greedy can incur more cost than MADQN/MAPPO. Table 3 shows that the Random is the most efficient algorithms among all while showing the worst performance (to be discussed further in the next section).

TABLE 3: Asymptotic Complexity Analysis of the Considered Schemes

Scheme	Complexity
MADQN/MAPPO	$O(n_e \times t_{train})$
Greedy	$O(n_{action})$
Random	$O(1)$

7.2 Sensitivity Analysis

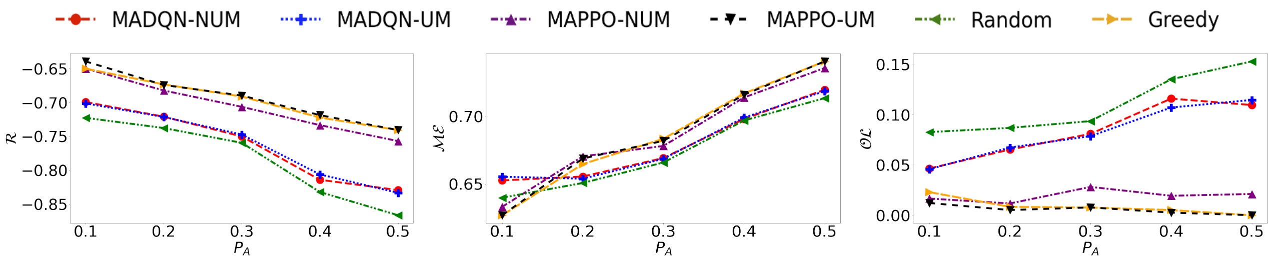
Below We conduct in-depth sensitivity analyses of the two baseline schemes (Greedy and Random) and the two uncertainty-aware DRL schemes (i.e., MADQN, MAPPO) with uncertainty maximization (i.e., MADQN-UM, MAPPO-UM) vs. without uncertainty maximization (i.e., MADQN-NUM, MAPPO-NUM) over a wide range of the attack probability P_A, the initial low battery level T_L, the number of cows (sensors) N, and the weather condition weight α. Note that there are 6 schemes in total in our analysis because of the distinction of with uncertainty maximization vs. without uncertainty maximization in the two uncertainty-aware DRL schemes.

7.2.1 Effect of Varying the Attack Severity

Fig 3 shows the effect of varying the attack probability, P_A, on the performance of the six schemes in terms of the three metrics in the network. We observe that increasing the attack probability (P_A) decreases $\mathcal{R}$ while increasing $\mathcal{ME}$ and $\mathcal{OL}$. Note that when P_A increases, the monitoring system gets severely compromised, thus the payoff per monitoring update would drop. MAPPO and Greedy can successfully identify this change in the payoff and achieve better performances than other schemes. The overall performance order with respect to the three metrics is: MAPPO-UM ≥ Greedy ≥ MAPPO-NUM ≥ MADQN-UM ≥ MADQN-NUM ≥ Random.

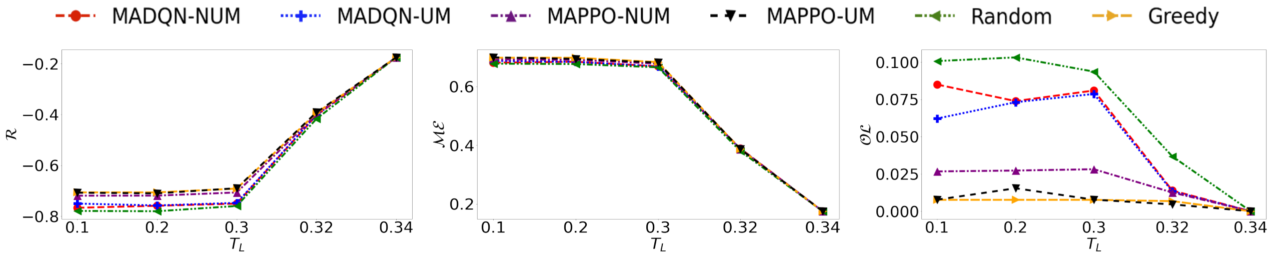
7.2.2 Effect of Varying the Initial Battery Levels (T_L)

Fig. 4 shows the effect of varying the initial battery level assigned to low-battery level sensors (LBS), T_L, on the performance of the six schemes in terms of the three metrics in the network. We observe that increasing attack probability (T_L) increases the accumulated reward ($\mathcal{R}$) while decreasing the monitoring error rate ($\mathcal{ME}$) and the degree of overload ($\mathcal{OL}$). We also observe that when T_L increases, all three metrics converge to one point due to the decreased number of LBS. Monitoring policies only apply to LBS and thus negligible differences are observed under high T_L. Overall, our proposed MAPPO-based schemes can achieve a low monitoring error rate with the lowest overload.



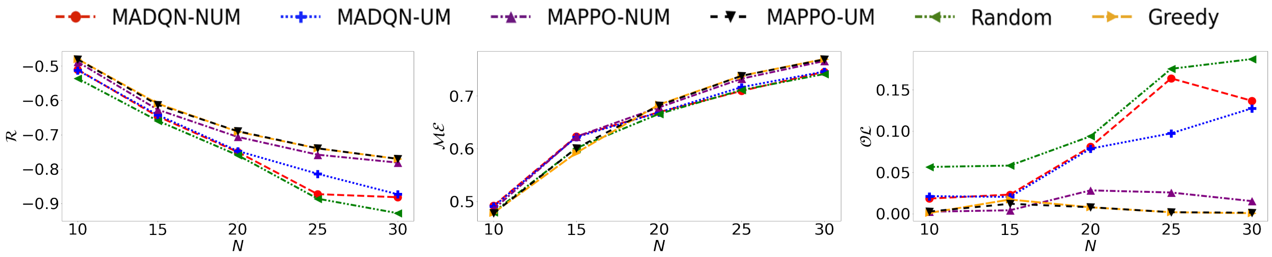
(a) Accumulated reward ($\mathcal{R}$) (b) Monitoring error rate ($\mathcal{ME}$) (c) Overload ($\mathcal{OL}$)

Fig. 3: Comparative performance with respect to varying attack probability (P_A).



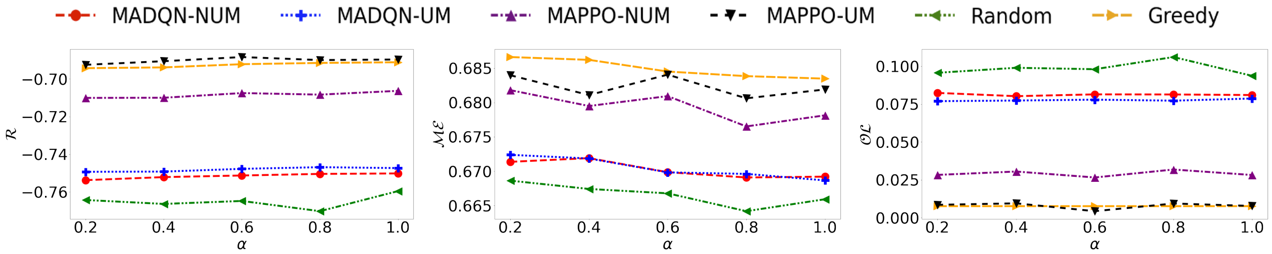
(a) Accumulated reward ($\mathcal{R}$) (b) Monitoring error rate ($\mathcal{ME}$) (c) Overload ($\mathcal{OL}$)

Fig. 4: Comparative performance with respect to varying the initial low battery level (T_L).



(a) Accumulated reward ($\mathcal{R}$) (b) Monitoring error rate ($\mathcal{ME}$) (c) Overload ($\mathcal{OL}$)

Fig. 5: Comparative performance with respect to varying a number of solar sensors (N) attached to cows.



(a) Accumulated reward ($\mathcal{R}$) (b) Monitoring error rate ($\mathcal{ME}$) (c) Overload ($\mathcal{OL}$)

Fig. 6: Comparative performance with respect to the different level of sun exposure (α).

7.2.3 Effect of Varying the Node Density (N)

Fig. 5 shows the effect of varying the number of solar sensors, N, on the performance of the six schemes in terms of the three metrics in the network. We observe that increasing the number of sensor nodes (N) decreases the accumulated reward ($\mathcal{R}$) while introducing higher $\mathcal{ME}$ and $\mathcal{OL}$. When N increases, there are not enough HBS (high battery level sensors) to transmit data for LBS. Consequently, the update requests from LBS may mostly fail. Our proposed MAPPO-UM scheme achieves the lowest monitoring error rate when N is low and the lowest overload when N is high, revealing the tradeoff that a lower monitoring error rate can incur a higher system overload.

7.2.4 Effect of Varying the Degree of Sun Exposure (α)

Fig. 6 shows the effect of α (the probability for a cow to be exposed to sun for energy harvesting) on the performance of the six schemes. We observe that a higher α contributes to boosting $\mathcal{R}$ while reducing $\mathcal{ME}$. Moreover, $\mathcal{OL}$ is not sensitive to varying α because the energy harvesting speed exceeds the battery consumption speed. We also observe that $\mathcal{ME}$ is equally reduced for every monitoring policy. Thus, when α is high, small changes in monitoring policies leads to insensitive $\mathcal{OL}$ while decreasing $\mathcal{ME}$.

8 Conclusions & Future Research

We summarize the **key findings** of this work as follows:

- This work is the first that proposed technologies to support an energy-adaptive monitoring system properly operating even in the presence of various adversarial attacks, including false data injection, DoS, and state manipulation (i.e., poisoning datasets in deep learning models) attacks. Unlike existing works that mainly focused on energy-aware approaches, our work achieved energy-adaptiveness and data security under energy-fluctuating, adversarial, and dynamic IoT environments.
- This work introduced uncertainty-aware data aggregation and update approaches to enhance the monitoring quality of the smart farm network. This approach was validated via mathematical proof and extensive experiments based on real datasets.
- We considered multiple deep reinforcement learning agents to identify optimal settings to maximize the monitoring quality of smart farms with solar-powered sensors. This design allowed high sustainability and scalability. Moreover, collaborative learning results in high performance in monitoring quality and system overload.
- We found from our experiments that the system overload does not always increase the monitoring error rate. Our proposed MAPPO-UM scheme can find monitoring policies to minimize both the monitoring error rate and the system overload.
- The payoffs to monitoring updates are vastly different under different scenarios. This discrepancy can result in different optimal monitoring policies being identified and applied in different scenarios.
- Our proposed MAPPO-UM scheme is showed to have an acceptable time complexity for which the major complexity comes from the training time and the size of the action space. MAPPO-UM outperformed other counterparts with an 8% reduction in both the monitoring error rate and the system overload.
- Among all schemes considered, MAPPO-UM can best adapt to different scenarios and identify the best monitoring policies for minimizing the monitoring error rate and the system overload.
- Our proposed MAPPO-UM scheme also showed strong robustness, particularly under harsh environments as demonstrated via extensive sensitivity analyses.

We also plan to further the research in the following **future research directions**:

- We will use more than two gateways to introduce more DRL agents for the proposed smart farm system to be applicable to larger-scale networks.
- We will leverage *transfer learning* algorithms to further improve the speed of learning convergence.
- We will identify an optimal energy level that can be used for low-energy solar sensors to request data transmission to nearby high-energy sensors.

References

[1] H. Alemdar, T. L. M. van Kasteren, and C. Ersoy, "Active learning with uncertainty sampling for large scale activity recognition in smart homes," *Journal of Ambient Intelligence and Smart Environments*, vol. 9, no. 2, pp. 209–223, Jan. 2017. [Online]. Available: https://content.iospress.com/articles/journal-of-ambient-intelligence-and-smart-environments/ais427

[2] A. Almeida and D. L. de Ipiña, "An approach to more reliable context-aware systems by assessing ambiguity - taking into account indetermination and vagueness in smart environments," in *Proceedings of the 2nd International Conference on Pervasive Embedded Computing and Communication Systems - Volume 1: PECCS,*, INSTICC. SciTePress, 2012, pp. 233–236.

[3] G. Chen, Y. Zhan, Y. Chen, L. Xiao, Y. Wang, and N. An, "Reinforcement learning based power control for in-body sensors in WBANs against jamming," *IEEE Access*, vol. 6, pp. 37 403–37 412, 2018.

[4] G. Chen, Y. Zhan, G. Sheng, L. Xiao, and Y. Wang, "Reinforcement learning-based sensor access control for wbans," *IEEE Access*, vol. 7, pp. 8483–8494, 2019.

[5] H. Chen, X. Li, and F. Zhao, "A reinforcement learning-based sleep scheduling algorithm for desired area coverage in solar-powered wireless sensor networks," *IEEE Sensors Journal*, vol. 16, no. 8, pp. 2763–2774, 2016.

[6] J. Cho, et al., "A survey on modeling and optimizing multi-objective systems," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1867–1901, 2017.

[7] M. Colezea, G. Musat, F. Pop, C. Negru, A. Dumitrascu, and M. Mocanu, "CLUeFARM: Integrated web-service platform for smart farms," *Computers and Electronics in Agriculture*, vol. 154, pp. 134–154, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168169917305112

[8] S. K. Das and N. Roy, "Learning, prediction and mediation of context uncertainty in smart pervasive environments," in *On the Move to Meaningful Internet Systems: OTM 2008 Workshops*, R. Meersman, Z. Tari, and P. Herrero, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 820–829.

[9] A. De Paola, P. Ferraro, S. Gaglio, and G. Lo Re, "Context-awareness for multi-sensor data fusion in smart environments," in *AI*IA 2016 Advances in Artificial Intelligence*, G. Adorni, S. Cagnoni, M. Gori, and M. Maratea, Eds. Cham: Springer International Publishing, 2016, pp. 377–391.

[10] A. De Paola, P. Ferraro, S. Gaglio, G. L. Re, and S. K. Das, "An adaptive bayesian system for context-aware data fusion in smart environments," *IEEE Transactions on Mobile Computing*, vol. 16, no. 6, pp. 1502–1515, 2017.

[11] H. Dong, Z. Ding, and S. Zhang, Eds., *Deep Reinforcement Learning Fundamentals, Research and Applications*. Springer, 2020.

[12] FAO. (2020) The food and agriculture organization (fao) of the united nations. [Online]. Available: http://www.fao.org/home/en/

[13] I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *International Conference on Learning Representations*, 2015. [Online]. Available: http://arxiv.org/abs/1412.6572

[14] M. Gupta, M. Abdelsalam, S. Khorsandroo, and S. Mittal, "Security and privacy in smart farming: Challenges and opportunities," *IEEE Access*, vol. 8, pp. 34 564–

34 584, 2020.

[15] Y. Hajjaji, W. Boulila, I. R. Farah, I. Romdhani, and A. Hussain, "Big data and iot-based applications in smart environments: A systematic review," *Computer Science Review*, vol. 39, p. 100318, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574013720304184

[16] J. E. Hopcroft and R. M. Karp, "An nˆ5/2 algorithm for maximum matchings in bipartite graphs," *SIAM Journal on computing*, vol. 2, no. 4, pp. 225–231, 1973.

[17] S. Igder, S. Bhattacharya, and J. M. H. Elmirghani, "Energy efficient fog servers for internet of things information piece delivery (iotipd) in a smart city vehicular environment," in *2016 10th International Conference on Next Generation Mobile Applications, Security and Technologies (NGMAST)*, 2016, pp. 99–104.

[18] A. Izaddoost, E. Ogodo, and S. Prasai, "Enhanced data transmission platform in smart farms," in *ACM Proceedings of the International Conference on Omni-Layer Intelligent Systems (COINS)*, New York, NY, USA, 2019, pp. 58–61.

[19] A. Jøsang, J. Cho, and F. Chen, "Uncertainty characteristics of subjective opinions," in *2018 21st International Conference on Information Fusion (FUSION)*, July 2018, pp. 1998–2005.

[20] A. Jøsang, *Subjective Logic: A Formalism for Reasoning Under Uncertainty*. Springer Publishing Company, 2016.

[21] F. Kiani, "Reinforcement learning based routing protocol for wireless body sensor networks," in *2017 IEEE 7th International Symposium on Cloud and Service Computing (SC2)*, 2017, pp. 71–78.

[22] F. Li, X. Song, H. Chen, X. Li, and Y. Wang, "Hierarchical routing for vehicular ad hoc networks via reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1852–1865, 2019.

[23] A. Machado, V. Maran, I. Augustin, J. a. C. Lima, L. K. Wives, and J. P. M. de Oliveira, "Reasoning on uncertainty in smart environments," in *Proceedings of the 18th International Conference on Enterprise Information Systems*, ser. ICEIS 2016. Setubal, PRT: SCITEPRESS - Science and Technology Publications, Lda, 2016, p. 240–250. [Online]. Available: https://doi.org/10.5220/0005866502400250

[24] C. O. Mathuna, T. O'Donnell, R. V. Martinez-Catala, J. Rohan, and B. O'Flynn, "Energy scavenging for long-term deployable wireless sensor networks," *Talanta*, vol. 75, no. 3, pp. 613–623, 2008.

[25] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[26] A. Modarresi and J. Symons, "Modeling technological interdependency in iot - a multidimensional and multilayer network model for smart environments," in *2019 11th International Workshop on Resilient Networks Design and Modeling (RNDM)*, 2019, pp. 1–7.

[27] N. Naderializadeh, J. Sydir, M. Simsek, and H. Nikopour, "Resource Management in Wireless Networks via Multi-Agent Deep Reinforcement Learning," *arXiv e-prints*, p. arXiv:2002.06215, Feb. 2020.

[28] J. O, D. Noh, and Y. Sohn, "Empirical test of Wi-Fi environment stability for smart farm platform," in *2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT)*, 2017, pp. 1–5.

[29] D. Popa, F. Pop, C. Serbanescu, and A. Castiglione, "Deep learning model for home automation and energy reduction in a smart home environment platform," *Neural Computing and Applications*, vol. 31, 05 2019.

[30] D. Preuveneers and Y. Berbers, "Architectural backpropagation support for managing ambiguous context in smart environments," in *Universal Access in Human-Computer Interaction. Ambient Interaction*, C. Stephanidis, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 178–187.

[31] N. Qi, K. Dai, F. Yi, X. Wang, Z. You, and J. Zhao, "An adaptive energy management strategy to extend battery lifetime of solar powered wireless sensor nodes," *IEEE Access*, vol. 7, pp. 88 289–88 300, 2019.

[32] G. Rocher, J.-Y. Tigli, and S. Lavirotte, "Probabilistic models toward controlling smart-* environments," *IEEE Access*, vol. 5, pp. 12 338–12 352, 2017.

[33] M. Roser. (2020) Future population growth. [Online]. Available: https://ourworldindata.org/future-population-growth

[34] N. Saxena, A. Roy, and J. Shin, "Chase: Context-aware heterogenous adaptive smart environments using optimal tracking for resident's comfort," in *Ubiquitous Intelligence and Computing*, J. Indulska, J. Ma, L. T. Yang, T. Ungerer, and J. Cao, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 133–142.

[35] J. Schulman, et al., "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[36] H. Sun, Q. Zhu, J. Ren, D. Barclay, and W. Thomson, "Combining image analysis and smart data mining for precision agriculture in livestock farming," in *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 2017, pp. 1065–1069.

[37] P. Tehrani, F. Restuccia, and M. Levorato, "Federated deep reinforcement learning for the distributed control of nextg wireless networks," 2021. [Online]. Available: https://arxiv.org/abs/2112.03465

[38] *CC2640R2F SimpleLink^{TM} Bluetooth® 5.1 Low Energy Wireless MCU*, Texas Instruments, 2016, rev. C. [Online]. Available: https://www.ti.com/product/CC2640R2F

[39] N. Twomey, T. Diethe, I. Craddock, and P. Flach, "Unsupervised learning of sensor topologies for improving activity recognition in smart environments," *Neurocomputing*, vol. 234, pp. 93–106, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231216315740

[40] K. Venkatesh, B. Barmada, V. Liesaputra, and G. Ramirez-Prado, "Resilient activities tracking in a smart home using ultrasonic sensors," in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 5274–5281.

[41] K. Yang, C. Shen, and T. Liu, "Deep reinforcement learning based wireless network optimization: A comparative study," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2020, pp. 1248–1253.

[42] Q. Zhang, Y. Mahajan, I. Chen, D. Ha, and J. Cho, "Deep reinforcement learning based wireless network optimization: A comparative study," in *IEEE Global Communications Conference (GLOBECOM)*, 2022, accepted.

[43] S. Zhang, S. McClean, B. Scotney, and C. Nugent, "Learning under uncertainty in smart home environments," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 2083–2086.

[44] H. H. Zhuo, W. Feng, Q. Xu, Q. Yang, and Y. Lin, "Federated reinforcement learning," *CoRR*, vol. abs/1901.08277, 2019. [Online]. Available: http://arxiv.org/abs/1901.08277