Chapter 6

Energy-Efficient communication of river water pollution data using LoRa and Game Theory

Within this chapter, we introduce an energy-efficient system for transferring the river water pollution data to the base station. This system leverages deep neural networks and long-range communication technology to address a critical research challenge in the field—efficiently transmitting pollution data from the river to a central base station while minimizing energy consumption. In the preceding chapter, we delved into the estimation of Water Quality Index (WQI) using IoT devices. In this chapter, we present a novel approach rooted in game theory to determine the time duration for using suitable spreading factor of the long range network for transmitting river water pollution data to the remote host or base station. This strategic approach not only conserves energy but also ensures the reliable transmission of data, ultimately leading to more effective environmental monitoring. To validate our approach, we provide both experimental and real-world assessments that highlight the system's performance in terms of accuracy and energy efficiency. These evaluations underscore the practicality and effectiveness of our proposed system in the context of river water pollution monitoring.

6.1 Introduction

Water play a crucial role in sustaining human activities. Contamination in river water has the potential to disseminate harmful bacterial, viral, and parasitic diseases. Continuous monitoring of river water pollution is vital for improving water quality. However, performing regular on-site inspections to assess pollution can be a demanding task, especially in areas with extensive river basins. This process also requires reliable and expensive water quality analysis equipment, along with skilled personnel [73]. Furthermore, it demands dependable and costly water quality analysis equipment, as well as a skilled workforce, as emphasized in [27]. Rivers serve as vital sources of drinking water. Ongoing monitoring is essential to maintain the cleanliness of river water; otherwise, the supply of drinking water could become scarce. The swift advancement of the Internet of Things (IoT) holds the promise of interconnecting a multitude of cost-effective, compact, and energy-efficient devices [74, 75]. IoT devices, furnished with sensors, oversee environmental conditions and relay this data to a remote host. These devices initially gather and process extensive volumes of sensory data regarding water quality at designated collection points. Subsequently, the processed sensory data is promptly and efficiently dispatched to the remote host. A fundamental challenge inherent to IoT devices lies in their constrained energy, storage, and processing capabilities.

An important factor responsible for energy consumption in IoT is the selection of communication protocols. Most of the IoT applications support short-range and long-range communication protocols. The short-range communication protocols, including, Zigbee, Bluetooth, and WiFi [80, 81], consume low energy. The long-range communication protocols (*e.g.,* 3G and LTE) [82] support wide coverage and high transmission rates but suffer from colossal energy consumption. The Long Range Wide Area Network (LoRaWAN) justifies its suitability for IoT by providing long-range communication with low energy consumption [83]. The LoRaWAN architecture consists of LoRa Nodes (LNs), LoRa Gateway (LG), and Network Server (NS). The LNs work as an IoT device for estimating the WQI of river water. LoRa supports different Spreading Factors (SFs), ranging from 7 to 12 for flexible long-range communication with low power consumption. Such SFs have different communication ranges and consume unequal power during transmission of the data.

6.1.1 Motivation

Previous studies have following major limitations which motivated this work:

- Authors in [49] proposed a model for measuring the turbidity of the river water using hyper-spectral remote sensing data. They applied the random forest-based ensemble model for analyzing the hyper-spectral data.
- In [50] authors developed a low-cost sensor node for estimating the quality of water. The authors deploy the sensors inside the pipe with running water to analyze its suitability for drinking. The authors also developed an event detection algorithm for recording the data of different parameters associated with the quality of river water.
- Mondal *et. al.* [51] built a battery-free pH sensor tag for passively monitoring the pH of the water. The developed pH sensor tag can wirelessly transmit the data at a data rate of 0.595 Kbps.
- Wang *et. al.* [52] proposed a low-cost designing principle for developing a turbidity sensor using transmitted light and scattered light passed through the water. Authors have justified the cost-effectiveness and high order accuracy in estimating the turbidity of the water.
- Authors in [53] designed and developed a water monitoring system by analyzing different parameters.
- In [54] authors developed a system for collecting garbage from the river water.

They proved the effectiveness of their proposed system by designing a prototype robot, which incorporates low powered IoT devices for coordination and control.

A detailed explanation of the specific characteristics of the river water data analysis motivated the development of a new energy-efficient protocol, rather than relying on a standard off-the-shelf low-energy utilization algorithm for data communication.

- Latency Sensitivity:

  Our identification of river water pollution application is particularly sensitive to communication delays, a new protocol might be required to minimize latency in data transmission. Standard algorithms could prioritize energy efficiency at the expense of real-time responsiveness, which might not suit for our requirements. 0 In Chapter 6, we indeed assert our proposed approach to energy-efficient communication of river water pollution data. The specific characteristics of this sequential and time-dependent data—such as its high variability, real-time processing requirements, and the need for accurate, timely transmission—differ significantly from standard data types typically addressed by off-the-shelf low-energy algorithms.
- Node Mobility or Network Dynamics:

  In cases where data sources (e.g., IoT devices, sensors, etc.) are highly mobile or the network topology changes frequently, off-the-shelf protocols might struggle to maintain connectivity and ensure reliable data transfer under dynamic conditions. In our research, the nature of pollution data is highly dynamic, with spikes in readings that require immediate attention. Traditional algorithms may not account for this variability, leading to delayed responses.
- Environmental Factors:

  External conditions, such as high interference, multi-hop networks, or extreme temperatures, may affect communication. A customized protocol could offer better resilience to these environmental factors than generic algorithms. The environmental factors influencing water quality data necessitate a tailored protocol that prioritizes critical data over less urgent information, ensuring that essential pollution alerts reach the base station promptly.

- Energy Constraints:

  Although existing low-energy utilization algorithms focus on energy savings, our specific hardware or power constraints might demand a more aggressive energy optimization tailored to our data, balancing energy efficiency with performance better than generic algorithms. Our proposed protocol incorporates adaptive transmission strategies, which are crucial for optimizing energy use based on real-time data conditions. This contrasts with generic algorithms that may operate uniformly regardless of the data context.

By addressing these specific characteristics, our protocol enhances both the efficiency and reliability of data communication in this domain. We believe this approach significantly contributes to effective monitoring and response to water pollution, and we appreciate your inquiry into this important aspect of our research.

None of the existing work have considered simultaneously energy-efficiency, accuracy, and low cost as its primary objective. In this chapter, we address the problem: *How to facilitate successful transmission of estimated WQI (water pollution level) to the central base station with minimum energy consumption?* To solve this problem, this work proposes, a game theory based approach to estimate the time duration for using the suitable spreading factor (virtual channel) to transmit the pollution data from a river to the base station using long range communication technology. The utilization of only selected spreading factors (SFs) could potentially lead to network congestion in the LoRa network. Therefore, in the system design phase, we must confront the task of efficiently assigning SFs to LoRa Nodes (LNs) to ensure the successful transfer of WQI data to the remote host.

6.1.2 Major Contributions:

To the best of our knowledge, this is the first work to address the energy-efficient communication of river water pollution data to the base station using game theory and long range communication technology. Our major contributions are as follows:

- **Energy Efficient Communication of WQI:** The system uses LoRa network that provides long-range communication with low energy consumption. Each LN uses a SF as a virtual channel for transferring the WQI to the LG. However, such SFs are the resource constraints in the network. Data transmitted using high SFs suffers from higher interference problem. Sometimes, the interference problem hampers the successful transmission of the crucial data to the LG. To avoid such problem, the crucial data should send using low SF. Since low SFs consume high power, LNs do not transmit their data on low SF voluntarily. Therefore, we propose a game theory-based approach to estimate the time duration for using the suitable SFs to transmit the crucial data by each LN. The LN gets incentive as allow to use the high SF once after using the low SF. Thus, the game theory-based approach helps to reduce the interference problem.
- **Experimental and Prototype Results:** Finally, we conduct various experiments on an open-source dataset to verify the effectiveness of the system. The experimental analysis demonstrates the ability of proposed system to preserve energy. Further, the impact of LNs on game parameters and the impact of resource constraint devices on training and testing performance is studied. We also build a prototype to demonstrate the impact of the compressed DNN parameters, energy consumption, and accuracy.

The rest of chapter is organized as follows. Section 6.2 describes the preliminaries. Next, Section 6.3 presents energy-efficient communication model using game theory. Next, Section 6.4 presents the holistic energy efficient water monitoring system. Further, the experimental and real world evaluations are presented in Section 6.5 and Section 6.6, respectively. Finally, the chapter is concluded in Section 6.7.

6.2 Preliminary

In this section, we first describe the different terminologies used in this research work. Later, this section covers a brief description of the problem associated with energy-efficient transmission of river water pollution level data and the overview of solution.

6.2.1 Preliminary

In this section, we present energy-efficient transfer of river water pollution data from river to the central base station using long-range communication technology and game-theory. The permissible limit of all the parameters are used by index developers to obtain various classes (e.g., water supply, irrigation, drinking and so on) depending upon the water quality [64]. For example, the acceptable (or required) limit and the permissible limit for turbidity is $1NTU$ and $5NTU$ respectively. Next, the acceptable (or required) limit and the permissible limit for pH value is $6.5-8.5$ and “No relaxation” respectively. If the permissible limit of a parameter used to classify water for various above-mentioned purposes is low, even minor variations in the parameter’s value can significantly impact water quality. As a result, higher weights are assigned to these parameters.

To determine a threshold for identifying crucial sensory data in Definition 6.1, we first consider the key attributes of interest (e.g., temperature, pH, DO, etc) and collect both existing and new sensory data. Next, we analyze the existing data to understand its distribution and identify points where significant changes in water quality index (WQI) occur. Further, we calculate the differences between new and existing data samples at these change points, then establish a threshold using 95^{th} percentile statistical method. Finally, we validate this threshold with a separate dataset, adjusting as needed to optimize its sensitivity and specificity, and continuously monitor its effectiveness as new data is collected.

Definition 6.1 (Crucial data) *A sensory data is said to be crucial data if the difference between the values of the attributes of sensory data samples and existing data samples at a given location is more than a given threshold value. Usually, the crucial data are acquired by the sensors just after the change in the class of WQI.*

Since we are using parameters such as pH, dissolved oxygen (DO), electrical conductivity, *etc.*, the threshold value depends on the specific location of data collection in the river. We have collected the data from the river where the channels are having waste of industry, agriculture, and anthropogenic activities mixing with the river water. Therefore, the threshold varies based on GPS coordinates, and we have already mapped the relationship between each location and its corresponding crucial threshold value.

Definition 6.2 (Crucial time duration) *The duration of transferring the crucial data from a LN i to the NS by using the high power (low SF) is known as Crucial Time Duration (CTD). The CTD of LN i is denoted by t_i^h, where h indicates the high power.*

Definition 6.3 (Reward time duration) *After transmission of the crucial data from a LN i to NS using low SF, the NS assigns the low power (high SF) to the LN i as a reward. The duration of the time when the LN i using high SF is known as Reward Time Duration (RTD). The RTD of LN i is denoted by t_i^l, where l indicates low power.*

Definition 6.4 (Nash equilibrium (NE) *Let $t^{h^*} = [t_1^{h^*}, t_2^{h^*}, ..., t_N^{h^*}]$ be the strategy of LNs and $U^D(\cdot)$ is the utility of the LN, respectively. Then NE in the game follows the following condition*

$$U_i^D(t_i^{l^*}, t_i^{h^*}) \geq U_i^D(t_i^{l^*}, t_i^h),\ \forall i \in N.$$

6.2.2 Problem statement and overview of solution

In this section, we focus on the river water pollution data transmission to the base station using game theory and long range communication technology. The major research problem is how to successfully transmit the correctly estimated river water pollution data to the central base station. The SFs in the long range communication network are virtual channels which are used to transmit the data. Further, the use of only selected SFs may increase network congestion in the LoRa. Therefore, while designing the system, we need to address the following challenge assign the SFs to the LNs in such a way that the WQI successfully transfers to the remote host. This work therefore addresses the problem of how to transmit the pollution data from river to the base station successfully. **Overview of Solution:** In response to this challenge, we have introduced a game-theory based approach.The utilization of only selected Spreading Factors (SFs) could potentially lead to network congestion in the LoRa network. Therefore, in the system design phase, we must confront the task of efficiently assigning SFs to LoRa Nodes (LNs) to ensure the successful transfer of Water Quality Index (WQI) data to the remote host. To address this challenge, our approach involves the estimation of the optimal time duration for employing the appropriate spreading factor within the long-range communication network for the transmission of water pollution data to the base station. First, we formulated a game. We used Nash equilibrium to provide the solution.

6.3 Energy-efficient transmission of river water pollution data

Game theory is employed for detection of water pollution level in the river. Game theory is a mathematical and strategic framework used to analyze and model interactions between rational decision-makers, often referred to as players, in situations where the outcome of each player's choice depends on the choices made by others. The LNs support multiple SFs, ranging from 7 to 12. The energy consumption of a LN is high when it uses low SF. The system uses a game theory-based approach to estimate the time duration for using the suitable SFs to transmit the crucial data by each LN. The communication unit sends the normal and crucial WQI to the connected LG by using suitable SFs. A period in LoRa protocol is the completion time of a cycle for a signal. A duty cycle in LoRa protocol, denoted as δ, is the fraction of one period in which a signal or system is active. This section proposes a system to detect the pollution level of the river water using LoRa. In the previous chapter, we studied the estimation of WQI in IoT using knoweldge distillation. Let $\mathbf{M}^t$ and $\mathbf{M}^s$ denote the cumbersome DNN and compressed DNN, respectively. The $\mathbf{M}^t$ and $\mathbf{M}^t$ are called as teacher and student models in knowledge distillation, respectively. Let $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^n$ represents the dataset having n number of instances, where, $\mathbf{x}_i$ is the i^{th} vector instance having class label y_i. If there are p sensors, each having q axes that are used for recording $\mathcal{D}$ at a sampling rate λ, then length of i^{th} data instance $|\mathbf{x}_i|$ is calculated as, $|\mathbf{x}_i| = \lambda \times p \times q$. For example, if we record a dataset at the sampling rate of 100 Hz using tri-axial sensor, then the length of data instance is $100 \times 3 \times 3 = 900$ per second. The $\mathbf{M}^t$ is trained on dataset $\mathcal{D}$, whereas, for $\mathbf{M}^s$, we utilize the dataset $\mathcal{D}' = \{\mathbf{x}_i{}', y_i\}_{i=1}^n$ with reduced length of instances. Further, to obtain $\mathbf{x}_i{}'$ from $\mathbf{x}_i$, such that $|\mathbf{x}_i{}'| < |\mathbf{x}_i|$, we incorporate following two techniques: i) Despite using the value of all q axes, use average value, then $|\mathbf{x}_i{}'| = \lambda \times p \times 1$; and ii) Identify the most distinguishable axis from q axes and only use value of that axis. Here, also we obtain $|\mathbf{x}_i{}'| = \lambda \times p \times 1$.

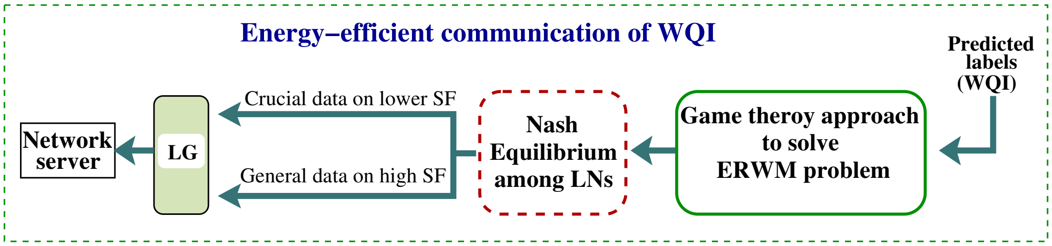


Figure 6.1: Block diagram of energy-efficient pollution data communication system.

The DNN system consists of $\mathcal{N}$ LNs. They periodically transmit data to LG. Let a LN $i \in \mathcal{N}$ uses SFs belongs to $s \in \mathcal{S}$ with coding rate $c \in \mathcal{C}$ for transmitting the data to the LG. Let the LN i consists of high, average, and low power levels denoted by h, a, and l, respectively. The s_i^j denotes the SF uses power level j for transmitting the data of LN i, where $1 \leq i \leq N$ and $j \in \{h, a, l\}$. Let b is the bandwidth used between LNs and LG then the transmission rate from LN i to LG [13] is given as

$$r_i^j = b \times \frac{s_i^j}{2^{s_i^j}} \times \frac{4}{4+c}. \tag{6.1}$$

The ERWM system consists N LNs, where, the CTD is denoted by $\mathbf{t^h} = [t_1^h, t_2^h, \cdots, t_N^h]$ and RTD is denoted by $\mathbf{t^l} = [t_1^l, t_2^l, \cdots, t_N^l]$. Each LN competes with each other to minimize its energy consumption. The utility function of a LN i consists of the following terms:

6.3.1 Energy consumption during CTD

Let P_i^h and r_i^h denote the power consumption and transmission rate during CTD of a LN i, respectively. The energy consumption per data of the LN i during CTD is given as

$$E_i^h = t_i^h \times r_i^h \times P_i^h. \tag{6.2}$$

6.3.2 Energy consumption during RTD

The Nash equilibrium estimates the CTD of a LN i denoted by t_i^{h*}, which provides the stable state of the system. RTD at low power is given to the LNs based on the CTD of LNs. RTD also depends on the strategy of other LNs also. As we increase the CTD of LNs, competition among LNs to get RTD increases which reduces the RTD of LN i. Let α and β is the bias constants such as $\alpha >> \beta$, then the RTD of LN i is given as

$$t_i^l = \alpha t_i^h - \beta \sum_{i \in \mathcal{N}} t_i^h \tag{6.3}$$

Let P_i^l and r_i^l denote the power consumption and transmission rate during RTD of a LN i, respectively, then the energy consumption per data of the LN i during RTD is given as

$$E_i^l = t_i^l \times r_i^l \times P_i^l. \tag{6.4}$$

6.3.3 Energy consumption during remaining duty cycle

LN i is active during the duty cycle. During RTD and CTD, the SF is assigned by the network server using the equilibrium. The LN i during remaining duty cycle uses random available SF for transmitting the sensory data to the LG. Let P_i^a denotes the power consumption when the LN i uses randomly available SF. Let m is the number of times the trigger occurred in one duty cycle due to changing in the class of water quality. Let t_i^a is the time duration of LN i to transmit data on random SFs which is defined as

$$t_i^a = \delta - m \times \left(t_i^h + t_i^l \right). \tag{6.5}$$

The energy consumption of the LN i during remaining duty cycle, denoted by E_i^r, is given as

$$E_i^a = t_i^a \times r_i^a \times P_i^a. \tag{6.6}$$

6.3.4 Energy consumption to maintain the stable state

LN i is able to send $\hat{d}_i^h$ data which is equals to $\alpha t_i^h r_i^h$. However, the equilibrium provides $d_i^h = (\alpha t_i^h r_i^h - \beta \sum_{i=1}^{N} t_i^h r_i^h)$ data to be transmitted at the LG for maintaining the stable state in the system. The energy consumption of LN i for maintaining the stable state is defined as the following Taguchi loss function:

$$E_i^m = \theta_i \left(\hat{d}_i^h - d_i^h \right)^2 \times P_i^a. \tag{6.7}$$

The non-linear relationship between data size and energy consumption arises from several factors: as data size increases, the complexity of processing grows, leading to higher energy demands due to more sophisticated algorithms and data management overhead. Additionally, larger datasets often result in diminishing returns on insights, causing energy costs to rise disproportionately. These elements contribute to a non-linear increase in energy consumption as data size expands.

The total energy consumption of LN i is the sum of E_i^h, E_i^l, E_i^a, and E_i^m, where $1 \leq i \leq N$. The utility of a LN i is the residual energy of i, denoted by U_i^D. Form Eqs. 6.2, 6.4, 6.6, and 6.7, the utility U_i^D is given by

$$\begin{aligned} U_i^D =& E - \left(E_i^h + E_i^l + E_i^a + E_i^m \right), \\ =& E - \left\{ t_i^h r_i^h P_i^h + (\alpha t_i^h - \beta \sum_{i \in \mathcal{N}} t_i^h) r_i^l P_i^l + \Big(\delta - m \left(t_i^h + t_i^l \right) \Big) r_i^a P_i^a + \theta_i \left(\hat{d}_i^h - d_i^h \right)^2 P_i^a \right\}. \end{aligned} \tag{6.8}$$

6.3.5 ERWM problem formulation

The concept of game is a game involves multiple players inherently. The game is a competitive activity which involves a strategy and multiple players contest or compete to maximize their profit or utility according to a fixed set of rules. There is no charity in the game. While we are playing a game we are looking to win the game. What you derive from the game is a utility. Here the utility is the reward time refers send the data with low power. The aim of the game is to maximize the profit or maximize your profit. Each game should be contrained by a set of rules or played according to the set of rules. The players are rational. There is no communication among the players. This is the assumption of the game. The players are different LNs, NS, LG. The actions are RTD(low power), CTD (high power). The players can not communicate with each other. The utility of player is not only determined by his action but by other players also. That is what make it a competition. So, this is the strategic interaction. The game for LN i can be formulated as

$$\textbf{ERWM Problem}: \max_{t_i^h} U_i^D(t_i^l, t_i^h) \tag{6.9a}$$

$$\text{s.t., } t_i^h + t_i^l + t_i^a \leq \mathcal{T} \tag{6.9b}$$

$$\{t_i^x\}_{\{h,a,l\}} \geq 0, 1 \leq i \leq N. \tag{6.9c}$$

where, $U_i^D(t_i^l, t_i^h)$ denotes the utility function of i which optimizes its CTD to maximize its utility function. Next, we derive Nash equilibrium of the game [84], [85] which determines the optimal strategy of LN i (*i.e.* t_i^{h*}) for given RTD t_i^l.

6.3.6 ERWM problem solution

Lemma 6.1 *Given t_i^l provided by network server to the LN i, there exist optimal CTD of LN i can be expressed as*

$$t_i^{h*} = \frac{mP_i^a r_i^a(1+\alpha-\beta) - P_i^l r_i^l(\alpha-\beta)}{a_{i,1} r_i^h} - \frac{(P_i^h r_i^h + a_{i,1}\sum_{j\neq i} t_j^h r_j^h)}{a_{i,1} r_i^h}, \tag{6.10}$$

where, $a_{i,1} = 2\theta_i P_i^a \beta^2 \sum_{j\neq i} {t_j^h}^2$.

Proof: Using Lagrangian multipliers $\gamma_{n,1}$ and $\gamma_{n,2}$ for constraints defined in Eq. 6.9a

$$
\begin{aligned}
\mathcal{L}_i^D(t_i^h, t_i^l) =& E - \Big\{ t_i^h r_i^h P_i^h + (\alpha t_i^h - \beta \sum_{i \in \mathcal{N}} t_i^h) r_i^l P_i^l + \Big(\delta - m(t_i^h + t_i^l)\Big) r_i^a P_i^a + \theta_i \left(\hat{d}_i^h - d_i^h\right)^2 P_i^a \Big\} \\
& + \sum_{x=\{h,a,l\}} \gamma_{i,x} \mathbf{t}_i^x - \sum_{x=\{h,a,l\}} \gamma_{i,2} \mathbf{t}_i^x - t_i^{max} = 0, \\
\text{s.t.,} \quad & \gamma_{i,x} t_i^x = 0, \\
& \sum_{x=\{h,a,l\}} \gamma_{i,2} \mathbf{t}_i^x - t_i^{max} = 0, \\
& \gamma_{i,x}, t_n^x \geq 0, \gamma_{n,2} > 0. \qquad (6.11)
\end{aligned}
$$

The first and second order derivatives of the utility function of a LN defined in Eq. 6.11 with respect to t_i^h is given as

$$
\frac{d\mathcal{L}_i^D(t_i^h, t_i^l)}{dt_i^h} = m P_i^a r_i^a (1 + \alpha - \beta) - P_i^l r_i^l (\alpha - \beta) - P_i^h r_i^h - 2\theta_i \beta^2 r_i^h P_i^a \sum_{i \in \mathcal{N}} t_i^h r_i^h + \gamma_{n,1} - \gamma_{n,2}. \qquad (6.12)
$$

$$
\frac{d^2\mathcal{L}_i^D(t_i^h, t_i^l)}{d(t_{i_h})^2} = -2\theta_i \beta^2 (r_i^h)^2 P_i^a. \qquad (6.13)
$$

We can see from Eq. 6.13 that $\frac{d^2\mathcal{L}_i^D(t_i^h, t_i^l)}{d(t_{i_h})^2} < 0$.

Since the value of second order derivative of $\mathcal{L}_i^D(t_i^h, t_i^l)$ is negative, the payoff function $U_i^D(t_i^h, t_i^l)$ in Eq. 6.8 is concave in t_i^h and **ERWM Problem** is convex optimization problem. Therefore the game has at least one Nash Equilibrium [86]. The value of t_i^h can be obtained from the coefficient matrix by setting the first order derivative of payoff function as calculated in Eq. 6.12 equals to zero. Let A and B are the coefficient matrices. Eq. 6.12 can be rewritten as

$$\underbrace{\begin{bmatrix} a_{1,1} & a_{1,1}(t_2^h r_2^h) & \dots & a_{1,1}(t_N^h r_N^h) \\ a_{2,1}(t_1^h r_1^h) & a_{2,1} & \dots & a_{2,1}(t_N^h r_N^h) \\ \vdots & \vdots & \ddots & \vdots \\ a_{N,1}(t_1^h r_1^h) & a_{N,1}(t_2^h r_2^h) & \dots & a_{N,1} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} t_1^h \\ t_2^h \\ \vdots \\ t_N^h \end{bmatrix}}_{\mathbf{t}} = \underbrace{\begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_N \end{bmatrix}}_{B}, \tag{6.14}$$

where, $B_i = mP_i^a r_i^a(1 + \alpha - \beta) - P_i^l r_i^l(\alpha - \beta) - P_i^h r_i^h \lambda' r_n^c - {b'_n}^c + \gamma_{n,1} - \gamma_{n,2}$, and $1 \leq n \leq N$ and $a_{i,1} = 2\theta_i P_i^a \beta^2 \sum_{j \neq i} {t_j^h}^2$.

From the strictly diagonal dominant theorem [87], the coefficient matrix A is non-singular. Since matrix A is non-singular, inverse of the matrix A is possible. Therefore Best response strategies of LNs $\mathbf{t}^h = [t_1^h, \cdots, t_n^h, \cdots, t_N^h]$ as the time duration to transmit data to the LG on SF c can be calculated as $A^{-1}B$, which is defined as

$$\begin{aligned} t_i^{h*} = & \frac{mP_i^a r_i^a(1 + \alpha - \beta) - P_i^l r_i^l(\alpha - \beta)}{a_{i,1} r_i^h} \\ & - \frac{(P_i^h r_i^h + a_{i,1} \sum_{j \neq i} t_j^h r_j^h)}{a_{i,1} r_i^h}. \end{aligned} \tag{6.15}$$

□

A Nash equilibrium is a set of strategies one for each player, such that no player has incentive to change her or his strategy given what the other players are doing. Nash equilibrium is a law no one wants to break even in the absence of police force. In the Nash equilibrium, each player's strategy is optimal when considering the decisions of other players. The Nash equilibrium is used to analyze the outcome of the strategic interaction of several players. In a strategic interaction, the outcome for each player depends on the decisions of the other players as well as their own. In a Nash equilibrium, every player in a group makes the best decision for himself, based on what he thinks the others will do. The Nash equilibrium of game returns optimal RTD and CTD of the LNs. The ERWM system initially uses the available SF for transmitting the sensory data. It uses Eq. 6.10 for calculating the optimal value of CTD. The ERWM system repeats the CTD till the stable state is not reached. The complete procedure is shown in Algorithm 6.1.

Algorithm 6.1: Nash Equilibrium among the LNs

```
/* Run the following at each LN $i \in \mathcal{N}$. */
Input: Precision threshold $\eta$, $\tau \leftarrow 0$, $t_i^h[0] \leftarrow \eta$ ;
Output: Best response strategy $t_i^h$ of $i$ ;
do
    $\tau \leftarrow \tau + 1$;
    /* Using Eq. 6.10 for estimating $t_i^h[\tau + 1]$. */
    $t_i^h[\tau + 1] = \frac{mP_i^a r_i^a (1+\alpha-\beta) - P_i^l r_i^l (\alpha-\beta) - P_i^h r_i^h - a_{i,1}[\tau] \sum_{j \neq i} t_j^h r_j^h}{a_{i,1}[\tau] r_i^h}$;
while $\left(\left\|t_i^h[\tau + 1] - t_i^h[\tau]\right\| > \eta\right)$;
```

Complexity Analysis for the message parsing in Algorithm 6.1: To analyze the complexity of message passing in Algorithm 6.1 for reaching consensus in Nash Equilibrium among the LNs (LoRa Nodes), we need to focus on the communication steps involved in updating the best response strategy $t_h[i]$.

1. **Initialization:** Each LoRa node i initializes its strategy with an input message related to the precision threshold η. This requires no communication as it is a LoRa operation.
2. **Iterative Updates:** The algorithm enters a loop where the LoRa node iteratively updates its strategy:
   - At each iteration τ, node i calculates its new strategy $t_h[i][\tau + 1]$ based on:
   - Its own parameters.
   - The strategies $t_h[j]$ of all other nodes $j \neq i$.
3. **Message Passing:** For each update, node i must communicate with all other nodes to obtain their current strategies $t_h[j]$. Therefore, the number of messages passed at each iteration is proportional to $|N| - 1$ (where $|N|$ is the total number

of nodes, excluding the node itself). If the algorithm requires k iterations to converge to a consensus (i.e., when the difference $|t_h[i][\tau+1] - t_h[i][\tau]| \leq \eta$), then the total number of messages exchanged will be $k \cdot (|N| - 1)$.

The complexity of message passing in terms of the number of required messages is $O(k \cdot |N|)$, where k is the number of iterations needed for convergence, and $|N|$ is the number of nodes involved in the consensus.

The message passing complexity of Algorithm 6.1 is $O(k \cdot |N|)$, reflecting the need for each node to communicate with all others across multiple iterations until convergence is achieved.

6.4 Holistic ERWM System

This section proposes the the overall energy-efficient river water pollution monitoring system by combining two parts. The first part incorporates knowledge distillation approach for estimation of river water pollution level whereas the second part comprises energy-efficient communication of pollution data to the base station. In the previous chapter, we studied that the compressed deep neural network after deploying on the IoT devices estimates the WQI (identifies the water pollution level) with acceptable accuracy. In this chapter, we studied the energy-efficient communication of river water pollution data. The overall ERWM system is as shown in the Fig. 6.2. The choice of using deep neural networks for sensor-based river water pollution assessment and game theory for monitoring through knowledge distillation was indeed carefully considered. After conducting extensive experiments, we found that the deep neural network approach effectively captured the complex patterns in the pollution data, significantly improving prediction accuracy compared to traditional methods. Similarly, the application of game theory provided a robust framework for understanding the interactions among various stakeholders, facilitating more effective monitoring strategies.

Our experiments demonstrated that both methods not only justified their selection

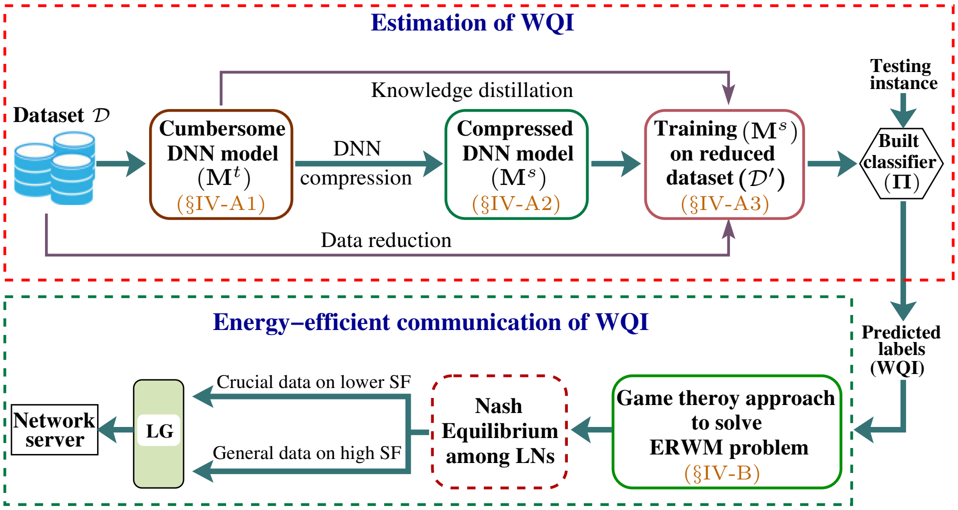


Figure 6.2: Block diagram of ERWM system.

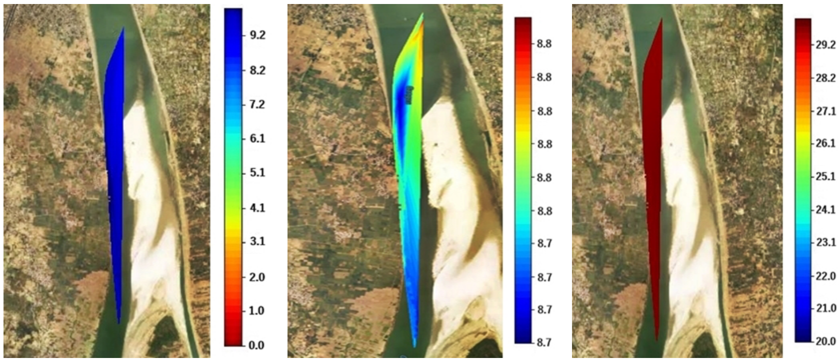
but also yielded meaningful insights and improvements in the assessment and monitoring processes. While there are alternative methods available, the results indicated that the chosen approaches were well-suited to the specific challenges of our application. Thus, we believe that the methodologies employed were appropriate and beneficial for the goals of our study.

6.5 Experimental evaluation and results

This section presents the experimental analysis for verifying the effectiveness of ERWM system. Firstly, we describe a publicly available dataset called River Water Monitoring (RWM) [1]. Next, we present the implementation details of ERWN system. Finally, we describe the experimental results to study the impact of different variables, devices, and game parameters on the performance of ERWM system.

6.5.1 RWM dataset

The RWM dataset [1] consists of the water parameters of the major Indian rivers. The dataset is collected to measure and analyze the river water pollution level. In the dataset, the water parameters that are considered to measure the water pollution level include, temperature, the potential of Hydrogen (pH), Electrical Conductivity (EC), Dissolved Oxygen (DO), total dissolved solids, turbidity, ammonia, *etc.* Further, the dataset is annotated with six labels, *i.e.,* "very-bad" , "bad", "medium", "good", "very-good", and "excellent", depending upon the pollution labels of the river. Such labels are denoted by $\mathbf{a_1}$, $\mathbf{a_2}$, $\mathbf{a_3}$, $\mathbf{a_4}$, $\mathbf{a_5}$, and $\mathbf{a_6}$ classes, respectively. Fig. 6.3 illustrates the heat map of river water parameters. The RWM dataset contains 100000 instances with their annotated class labels.



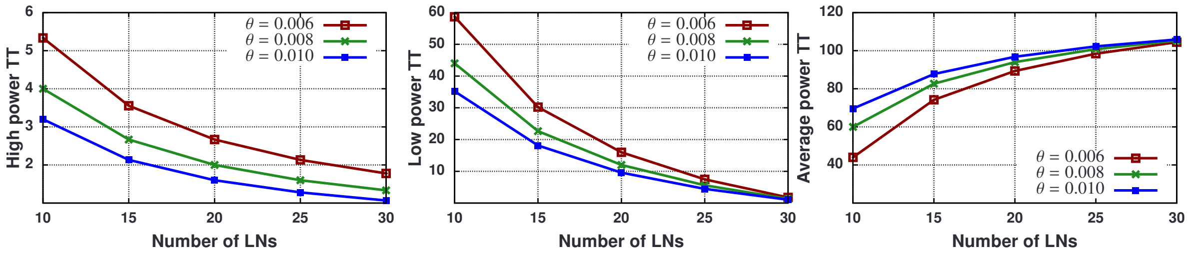
(a) DO. (b) pH. (c) Temperature.

Figure 6.3: Heat maps of different parameters in the sensory dataset [1], on a specific date (for river Ganges in Varanasi).

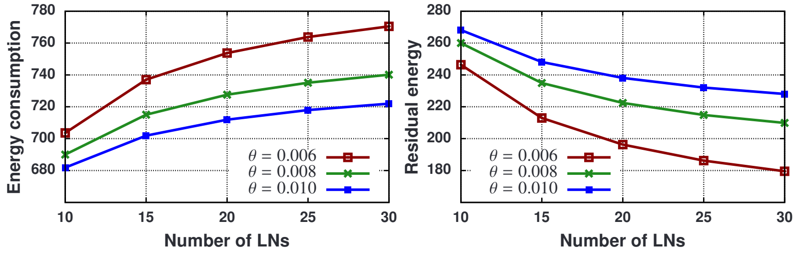
6.5.2 Implementation details

We incorporate Python programming language, where, we use deep learning library Tensorflow for implementing the energy efficient communication of river water pollution data.

Further, for randomly partitioning the dataset for training and testing sub-datasets, we incorporate *train_test_split* function in sklearn library of Python language. Here, the number of convolutional layers considered during the experiment is 5 having 128 filters each. This layer configuration is selected as we obtained the best performance out of several configurations during the experiments. We use Relu function for normalization on each layer other than softmax function at the output layer. The optimizer during



(a) High power transmission. (b) Low power transmission. (c) Average power transmission.



(d) Energy consumption. (e) Residual energy.

Figure 6.4: Illustration of the number of LNs on different game parameters, where TT is Transmission Time.

the experiment was "adam", and the learning rate is 0.005.

6.5.3 Experimental results

This section presents various experimental results on RWM [1] dataset to study the training performance, class-wise accuracy, impact of the number of LNs, and the devices used for deploying the DNN models.

6.5.3.1 Impact of number of LNs on game parameters

Further, we study the impact of the number of LNs on different game parameters, including, high power transmission time, low power transmission time, average power transmission time, energy consumption, and residual energy. We made the following observations from this result. Part (a) of Fig. 6.4 illustrates the rapid decrement in the value of high power transmission time as the number of LNs increases for all $\theta = \{0.006, 0.008, 0.010\}$. It is due to less availability of spreading factor for data transmission when the number of LNs increases. Similarly, the pattern of decrement in the value of low power transmission time is nearly the same as high power transmission time upon increasing the count of LNs from 10 to 30, as shown in part (b) of Fig. 6.4.

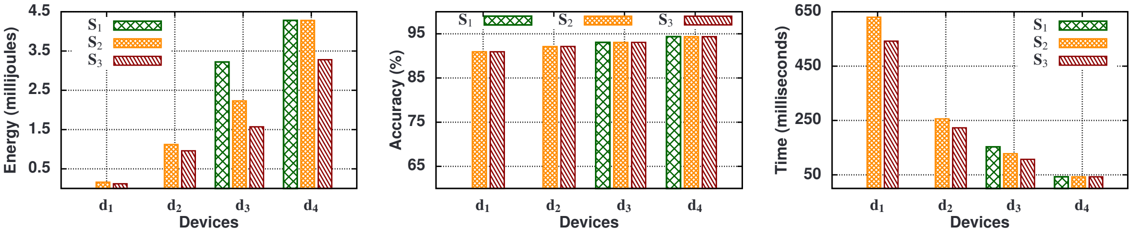
Further, part (c) of Fig. 6.4 shows that the value of average power transmission time increases by increasing the number of LNs. It is because the increment in the number of LNs increases the average transmission time over the LoRa network. Finally, parts (d) and (e) depict the increasing and decreasing behaviour of energy consumption and the remaining energy of the ERWM system with the increase in the number of LNs, respectively. Here, a similar pattern of increment and decrement is observed for all value of θ, *i.e.,* $\theta = 0.006$, $\theta = 0.008$, and $\theta = 0.010$.

6.5.3.2 Schemes for evaluation

Finally, for experimentally analyzing the energy saving achieved by ERWM system, we define three schemes as follows

1. **Scheme $\mathbf{S}_1$**: It is the baseline scheme, where the original teacher model is trained with its higher structural configuration. Here, no compression, data reduction, and Algorithm 6.1 for allocating the SFs are adopted.
2. **Scheme $\mathbf{S}_2$**: In this scheme, the DNN models are compressed to enhance its suitability for limited resources IoT devices. We preserve sufficient energy that becomes favourable for battery operated devices.
3. **Scheme $\mathbf{S}_3$**: This scheme is the replication of ERWM system, where DNN compression, data reduction, and Algorithm 6.1 for allocating the SFs to the LNs are used.

Fig. 6.5 illustrates the energy consumption, accuracy, and execution time on different devices under schemes $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{S}_3$. Accuracy is a measure of system performance that refers to the closeness of the system output with the actual output. In terms of classification and prediction, it is the fraction of correct number of predictions to the total predictions made. The accuracy of a system depends on the data input and



(a) Energy consumption. (b) Accuracy. (c) Execution time.

Figure 6.5: Illustration of energy, accuracy, and execution time on different schemes ($\mathbf{S}_1$, $\mathbf{S}_2$, and $\mathbf{S}_3$) for devices ($\mathbf{d_1}$, $\mathbf{d_2}$, $\mathbf{d_3}$, and $\mathbf{d_4}$).

the computational model used by the system for obtaining the output. The better computational model provides better accuracy for the input data but requires complex processing. Because of limited resources and low processing capability of IoT (LN) devices, only low computational models with limited data can be used, which leads to lower performance accuracy. To improve the accuracy, game theory based approach incorporated at the resource constrained IoT devices to estimate the time duration for using the suitable spreading factor to transmit the pollution data using long range communication network. Here, upon using scheme $\mathbf{S}_1$, the complete teacher model has to be deployed on different devices. So, under scheme $\mathbf{S}_1$, we cannot deploy the teacher model on devices $\mathbf{d_1}$ and $\mathbf{d_2}$ due to resource limitations. However, on-device $\mathbf{d_3}$, we manage to achieve the feasibility of deployment with increasing energy and time. For device $\mathbf{d_4}$, no compression or data reduction is required due to the efficacy of resources, so the teacher model is used under all three schemes on device $\mathbf{d_4}$. $\mathbf{S}_3$ uses Algorithm 2 for allocating the SFs to the LNs and therefore saves more energy then $\mathbf{S}_2$ as illustrated in Fig. 6.5.

Further, schemes $\mathbf{S}_2$ and $\mathbf{S}_3$ are suitable for all devices with limited resources. The proposed scheme ($\mathbf{S}_3$) requires both less energy and execution time in contrast with scheme $\mathbf{S}_2$ and achieves similar accuracy. Therefore, we can conclude that the proposed scheme $\mathbf{S}_3$ not only achieves significant performance abut also preserves considerable accuracy for its effective deployment on battery-operated IoT devices.

6.6 Real World Evaluation

We have deployed ERWM system in the real world to evaluate the performance of the system. We have selected Ganga river water dataset because our institute (IIT (BHU) Varanasi) is located near the banks of the river.

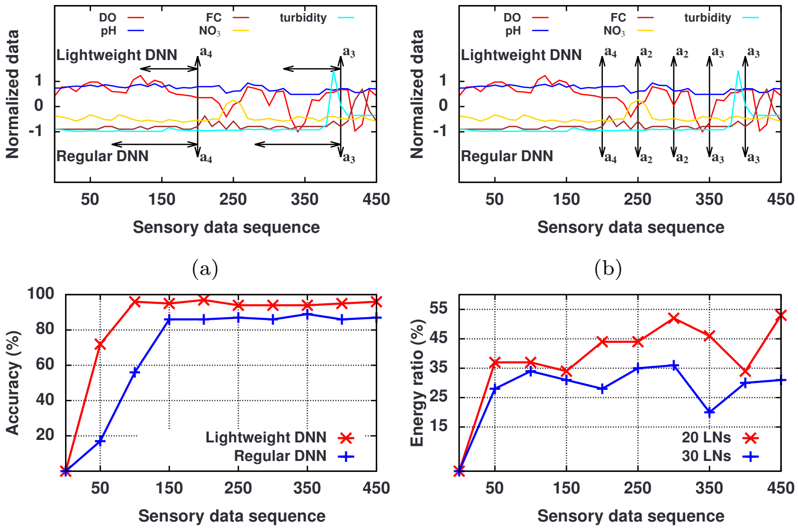
Experimental setup: The authors are working on a project for monitoring the river water [1]. The sensory dataset consists the following attributes: pH, temperature, electrical conductivity, dissolved oxygen, turbidity, nitrate, ammonia, chlorophyll-a, chromophoric dissolved organic matter (CDOM) fluorescence, tryptophan fluorescence, and channel flow. The complete details about the project, dataset, and video can be found at [1]. We have collected the sensory data by using HANNA HI-9829 Multiparameter pH/ISE/EC/DO/Turbidity Waterproof Meter with GPS, Turner C. 3 Fluorometric sensor, and global water FP101 flow probe. Fig. 6.6 illustrates FP101 probe, HI-9829, fixing the sensors, and LoRa Shield image of fieldwork. We use LoRa Shield for Arduino V95 for communicating the WQI to the base station. The Raspberry Pi 4 Model B is used for the processing of the sensory data and generates the WQI by using the proposed DNN. The Raspberry Pi and LoRa Shield together work as a LN in the system. Since sensor probes are not supported LoRa network, therefore we have installed the collected sensory data in the LN. The base station is set up in our lab (Department of CSE, IIT (BHU) Varanasi) which is nearly 5 Kilometres from the experimental site. We have selected 30 locations near the banks of the river for deploying the LNs.

Experimental results: The objective of this experiment is to illustrate the impact of the ERWM system on energy consumption and accuracy. We consider two solutions with and without ERWM system. In without ERWM system, we use large-size DNN (Regular DNN) and random allocation of SFs. Here, we have execuated DNN on Raspberry Pi for estimating WQI and LoRa for communicating it to the base station. Next, we have used the ERWM system, where compressed DNN runs (Lightweight) on Raspberry Pi for estimating WQI and proposed mechanism for allocating the suitable

Figure 6.6: Illustration of devices used for sensing the river water.

SFs to the LNs.

The large-size DNN requires more operations and therefore takes more time to estimate the WQI on LoRa node. To compute WQI at a specific time interval, without ERWM system uses less number of instances as compare with the proposed ERWM system. Parts (a) and (b) of Fig. 6.7 illustrate the consideration of instances and the generated WQI. Part (c) of Fig. 6.7 demonstrates that the ERWM system gives high accuracy because of the following reasons: i) ERWM system uses a large number of instances; therefore model learns more sophisticated features and provides better accuracy; and ii) Without using the proposed SF allocation, the system drops some estimated WQI packets; therefore base station takes average based on the limited results. This reason also increases the uncertainty of the results, as shown by the range graph.



(a) (b)

(c) (d)

Figure 6.7: Illustration of results of real world evaluation.

Part (d) of Fig. 6.7 illustrates that energy consumption ratio of with and without ERWM system. The energy consumption ratio is estimated as Energy ratio = $\frac{\text{Energy consumption with ERWM}}{\text{Energy consumption without ERWM}} \times 100$. Part (d) of Fig. 6.7 illustrates the energy consumption is highly reduced for ERWM system in both cases, *i.e.,* of 20 LNs and 30 LNs. It is because, in case of without ERWM, all LNs try to use best SFs for the successful transmission of all types of packets (crucial or normal); therefore, energy consumption is high. The energy consumption is also high because of the high packets collision and re-transmission.

6.7 Conclusion

In this paper, we proposed an Energy-efficient communication of River Water pollution data to the central base station in LoRa-based Internet of Things. The system estimated the required energy for transmitting the WQI. Further, this work established Nash equilibrium among the devices to find suitable SFs in LoRa network. The experimental and prototype results show that the proposed ERWM system can achieve significantly higher accuracy and preserve considerable energy during transmission of WQI. This work also provides a future direction to develop the LN deployment strategy that reduces the number of LNs and prolong the network lifetime.