# Chapter 2

# Preliminaries and related work

The preliminaries of the various methods utilized in this thesis are presented in this chapter. We also describe state-of-the-art work covering the strategies to gather different challenges faced while working on enhancing the capabilities and efficiency of devices in long-range IoT using machine learning.

## 2.1 Preliminaries

This section covers the overview of the LoRa network, followed by the knowledge distillation and federated learning techniques used in this thesis.

### 2.1.1 LoRaWAN overview

Semtech invented LoRa, a patented technology. There are public specifications for LoRaWAN, which are made available or promoted by the LoRa alliance. The physical (Radio) layer technology of the LoRaWAN protocol IEEE 802.15.4 is known as LoRa. Any protocol can be used on top of the LoRa physical layer technology [1]. However, the LoRa alliance recommend the LoRaWAN protocol. LoRaWAN requires radios to have a duty cycle of no more than 1 percent. The working of the LoRa modulation technique is based on the Chirp Spread Spectrum (CSS). In LoRa, information is encoded using frequency chirps with linear frequency variations over time. Chirp modulation is a technique for transmitting symbols. The symbol encodes into multiple signals of decreasing (down chirp) or increasing (up-chirp) radio frequencies. The spreading factor varies from 7 to 12, depending on the communication range. The high data rate, less communication range, and less packet loss in lower spreading factors compare to a higher spreading factor. A network server optimizes data rate and transmission power using the Adaptive Data Rate (ADR) bit. To control the uplink transmission of the

LoRa End Node (LEN), LoRaWAN defines an ADR scheme using the network server. Table 2.1 illustrates the different parameters in LoRaWAN.

**Table 2.1**: Illustration of different parameters, symbols and their values in LoRaWAN.

| Parameter | Symbol | Value |
|---|---|---|
| Spreading Factor (SF) | s | $\{7, 8, \cdots, 12\}$ |
| Signal Bandwidth (BW) | b | 125 kHz, 250 kHz or 500 kHz |
| Coding Rate (CR) | $c^r$ | 4/5,4/6,4/7,4/8 |
| Transmission power (TP) | P | {2,5,8,11,14} |

Three factors affect data rate (DR) in LoRa networks: BW, SF, and CR.

$$DR = b \times \frac{s}{2^s} \times c^r. \tag{2.1}$$

Calculating the symbol duration, denoted as $T_{sym}$, is as follows:

$$T_{sym} = \frac{2^s}{b}. \tag{2.2}$$

Using a higher SF implies a lower data rate as well as a longer symbol duration, which means more Time on Air (ToA). Each Device in LoRaWAN has a transmission duty cycle of 1 percent, which translates to 36 seconds per hour. Furthermore, LoRaWAN bandwidths can be used at 125kHz, 250kHz, and 500kHz. LoRaWAN has used the logic of the Media Access Control (MAC) protocol specification of the data link layer, which allows the end node to exchange information with the network server through a gateway. After getting a message from the end node, the network server sends that message to the application server, which is directly connected to the user. To secure the data, LoRaWAN uses two encryption techniques, Advance Encryption Standard and AES128. First, the AES128 security protocol applies between the LoRa end node and the LoRa network server. The second AES128 security protocol was used between the LoRa end node and the LoRa application server. Spreading factor, bandwidth, transmission power, and channel control are the transmission parameters in LoRa technologies. These transmission parameters impacted the qualities of received signal strength, power consumption, and coverage range.

### 2.1.2 Knowledge distillation

Knowledge Distillation (KD) technique improved the performance of the lightweight model using the generalization ability of the large-size model [2, 3]. KD uses keywords *teacher* and *student* for large-size and lightweight models, respectively. KD trains a student (lightweight model) under the guidance of a teacher (large-size model) so that the student can mimic a similar output pattern as the teacher. The trained student requires lower memory and task execution delay than the teacher with minimal or negligible performance compromise. Most of the existing KD approaches utilized the knowledge limited to the pre-trained teacher model and did not consider the knowledge from the training process of the teacher model. Different from the existing work, Zhao *et al.* in [4] employed one teacher trained from scratch to force the student to approach the optimal path toward the final logits and another pre-trained teacher to focus on a critical region that is useful towards the given task. The authors in [5] proposed a framework where a large-size model supervised the whole training process of the lightweight model. In addition, the lightweight model shared parameters with a large-size model to provide low-level representation directly from the teacher.

### 2.1.3 Federated learning overview

Federated Learning (FL) is an emerging paradigm that facilitates collaborative learning among multiple participant devices without compromising data privacy [6–8]. It allows participant devices to train a shared model in a decentralized manner, keeping private training data confined to the local devices. The traditional distributed training framework requires consensus after each local iteration, either using server or peer communications. However, the participants in FL perform multiple local updates before aggregation at the server in each communication round. Hence, FL also minimizes frequent consensus among the distributed participants. The central server initiates the FL operation and broadcasts a randomly initialized model to all the participants. Each participant trains the received model using its local dataset and sends the model's Weight Parameter Matrices (WPM) to the server. Afterwards, the server aggregates the WPM received from multiple participants and sends back the aggregated WPM. This process generates robust and generalized models for each participant. FL is a key enabler technique to preserve data privacy and minimize communication delay via local training and global aggregation [8, 9].

## 2.2 Related work on federated learning with imperfect labels

This section provides an overview of the existing research on FL, focusing on the available resources on participant devices and the requirements for transmitting data to the central server.

In their study, the authors in [10] proposed a Traffic Aware Data Offloading (TRADING) technique for processing large-scale traffic data, which aimed to enhance Intelligent Transportation System (ITS) capabilities and enable more precise decision-making in connected autonomous vehicle environments. TRADING emphasised efficiently distributing data traffic among gateways while taking network conditions nearby and automobile traffic. Furthermore, it points out the issue of gateway marketing overhead, freeing up transmission channels for efficient large data traffic transfer. The article evaluated TRADING's performance through realistic simulations in areas such as gateway access overhead, load distribution, data offloading delay, and success rates. The findings indicated promising developments in activating data-centric ITS solutions for handling substantial traffic patterns and trends based on data.

FL is a method of collaborative training that stores or maintains data on local devices. The author in [11] explored the use of various FL techniques on a multiple-access edge computing server, focusing on resource distribution for CPU and communication across mobile devices for learning. Based on the hyper-learning rate parameter, these learning services converged to a collaborative problem called Multi-Source Federated Learning (MS-FEDL). The author emphasized both decentralized and centralized methods to solve MS-FEDL. The results demonstrated the impact of the proposed methods with decentralized algorithms, providing independent management of resources and learning methods while preserving secrecy compared to a heuristic technique.

Due to the high communication cost, the article explored the limitations of employing FL for traffic forecasting using deep learning-based models. The authors in [12] suggested Clustering-based hierarchical and Two-step-optimized Federated learning (CTFed) as a feasible FL technique to address these limitations. CTFed implemented a two-step optimization approach using the particle swarm optimization algorithm to improve local models and group clients based on the similarity of their local model parameters. This approach decreased communication overhead by transmitting only one representative local model improvement from each cluster to the central server. In traffic forecasting, the proposed system was validated to enhance training efficiency, prediction accuracy, and robustness to network instability.

The author in [13] explored the difficulties of implementing FL in a volatile environment where client populations, client data, and training status could all change at any time. They proposed a new learning technique called volatile FL that considered set, statistical, and training volatility. To solve the client selection problem in volatile FL, they introduced the idea of Cumulative Effective Participation Data (CEPD) as an optimization objective and created the Upper Confidence Bound-based Greedy Selection

(UCB-GS) approach. The UCB-GS method significantly decreased the number of training rounds and enhanced the global model's accuracy in volatile FL, as demonstrated through experiments.

The author in [14] proposed FL, a decentralized learning strategy that enabled numerous edge computing nodes to collectively develop a shared learning model without sharing their raw data with a central server, thereby reducing communication costs. It addressed issues associated with non-independent and non-identically distributed data and User Equipment (UE) heterogeneity. The authors presented an enhanced FL technique that employed stochastic gradient descent on a sampled subset of UEs during each global round and included a weight-based proximal term in the local loss function. They provided an upper convergence bound that balanced convergence speed and the number of global rounds. The proposed FL algorithm was applied in wireless IoT networks to decrease energy use or completion time. The paper presented numerical results showing its improved performance and robustness compared to the FedAvg method, indicating its feasibility for learning under bandwidth constraints in wireless IoT networks.

The author in [15] outlined a survey of recently published papers in the field of FL. Due to concerns about data privacy and the growing significance of deep learning, the discipline of machine learning has undergone a transformational shift in recent years. This shift has given rise to FL, a decentralized technique of ML implementation that maintains data privacy by storing raw data on local devices, performing localized ML training, and reducing data transfer costs. FL involves members collaborating on shared models via a central server, providing an innovative solution to the challenges of traditional centralized systems. The article provided a thorough overview and analysis of this rapidly developing field, making it an invaluable resource for researchers and practitioners interested in cutting-edge technologies and related subjects. It examined various FL-related topics, such as system models, application areas, privacy and security, and resource management.

FL involves a large number of participant devices with varying resource levels. Running the same model on all participant devices and updating the WPM for global aggregation simultaneously has resulted in poor performance and longer convergence times due to the dissimilar resources [16–19]. The authors in [16] proposed a system to select participants for the adaptive global aggregation of the WPM. The system chose participants who could simultaneously contribute to the WPM, excluding stragglers from the overall aggregation. The authors in [17] developed a method to lower the CPU frequency of the faster members of the federation to match the slower computing performance of the stragglers, thereby reducing resource shortages for participants with slower processing. In addition, the authors suggested an algorithm for resource optimization in [18], which distributed participants among several sub-networks based on the availability of their resources.

The authors in [20] focused on the impact of noisy labels in Fl. The author derived maximum generalization error in the data as a noisy label. The author experimented on CIFAR-10 and MNIST datasets using different Fl algorithms. The experimental result showed that global model accuracy degraded while noise levels in data increased. The performance of the Fl model degraded in training over noisy fading wireless transmission channels. The authors in [21] presented two methods to overcome the gap in existing work. The author first tackled the effect of transmission channel noise. In this process, the author incorporated an innovative tracking-based stochastic approximation method in the federated averaging in the pipeline, which averages the effect of the noise in the channel. It also provided convergence guarantees without the increased transmission power gain. The second part was to take action to channel fading and further improve power consumption at the client level using optimization techniques.

In wireless FL, due to the lack of the spectrum, only some of the clients collaborate in the FL training in each round. Conversely, the performance degradation of FL is due to noisy labels, which existed naturally in the dataset. To reduce the noise label issue, the author [22] proposed an algorithm that measures the noise ratio for each client. The authors then proposed an online participation selection technique supported by Copeland score and multi-arm bandits. In this technique, the author selected low-noise ratio participants for every round of training. These proposed techniques depicted the improved result.

Distributed resources in Mobile Edge Computing (MEC) systems enable users to maintain their data privacy locally. Due to an unsecured wireless transmission environment and constraints on resources in the MEC system. Because of these issues, FL's performance and training efficiency have degraded. To solve this problem, the author [23] proposed the optimization algorithm to formulate the tradeoff between training cost and model accuracy. In the second part, the author proposed a joint optimization technique to optimize the sample selection, model compression and user strategies, providing the optimal solution for computationally improved efficiency. Finally, the result showed the cost and accuracy loss of Fl in the MEC system were significantly reduced.

## 2.3 Related work on interference problem in LoRaWAN

Generally, IoT applications rely on low-power devices, which communicate and interact to collect and transmit data from the environment to the Network Server (NS) [24]. LoRaWAN is an emerging paradigm in Low Power Wireless Area Networks (LPWAN) family [25]. The popularity of LoRaWAN in IoT is strengthened by its low power, long-range communication, and cost-effectiveness. LoRaWAN offers tradeoffs between communication range, power consumption, and data rate, making LoRa an efficient communication protocol for improving the sustainability of IoT applications. There are many popular LPWAN technologies like LoRa, SigFox, and NB-IoT. LoRaWAN is the most popular among all LPWAN technologies because its deployment is easy and cost-effective [26]. Sensor devices are equipped with a small LoRa chipset to enable various IoT applications by transmitting LoRa packets. The Chirp Spread Spectrum (CSS) modulation technique is used in LoRa to send data from the LoRa Node (LN) to the LoRa Gateway (LG) [27]. The LG then transmits the data to the Network Server (NS), which manages the entire network and allocates resources (such as spreading factor (SF) and bandwidth (BW)) to the LN. The NS subsequently sends the data to the application server for further interpretation.

LoRaWAN technologies find their applications in various domains, including smart home, smart agriculture, smart metering, smart parking, pollution monitoring, fire detection, industries, health care, *etc* [28–32]. LoRa provides a flexible range of communication with low power consumption via various SFs ranging from 7 to 12. The high value of SF provides a more extended communication range and, hence, a long air time. Due to comprehensive communication range, low power consumption, and ease of implementation, LoRa network applications grow exponentially in real-time settings [33]. With the exponential growth in the LoRa Network, the researchers explore and analyze issues like scalability, collision, capture effect, and LoRa parameter allocation [34]. Lora parameters include SF, BW, and CR to orthogonalize the transmissions.

Recent research in the LoRa network focused on scalability and reliability. In this process, the LoRa network faced issues like capture effect and packet collision, especially in high-density LoRa nodes. Many techniques, including power regulation and dynamic scheduling, were investigated to solve the issues. In this article, the author [35] presented RS-LoRa, a novel MAC layer protocol that dynamically manages transmission parameters using a lightweight scheduling strategy. This strategy improved packet error ratio, throughput and fairness, especially in dense LoRa nodes.

IoT has attracted much attention recently. The LoRa network plays a crucial role in IoT application expansion. In deploying the Dense LoRa node, it faces issues like packet collision, scalability throughput, *etc.* The author [36] suggested that the Traffic aware Channel and Backoff window size Allocation (TCBA) system seeks to close this gap by reducing latency and enhancing network capacity. For the LoRa network, the author derived optimal packet transmission probabilities and support capacities by integrating a statistical latency-aware network model. The TCBA technique reduced end-to-end latency and achieved higher throughput than previous methods, as confirmed by experimental results.

The authors in [37] improved the Packet Delivery Ratio (PDR) model that considered the accumulated interference power from several colliding frames and took these dependencies. The author also investigated the reliability optimization of inter-packet Error Correction Codes (ECCs) without any change in network capacity. The improved technique leverages a framework for choosing the best transmission parameter. An ad hoc Python discrete-event simulator improved the theoretical model by evaluating the effect of near-far and inter-SF interference and showing the benefit of power regulation.

Due to its multi-device management capabilities, LoRa is widely recognized as a low-power wide-area network (LPWAN) solution for the Internet of Things (IoT). However, allocating Spreading Factors (SF) remains challenging due to complexities and scalability issues. In addressing this, the author in [38] employs a game-theoretic technique to calculate Bayesian Nash equilibria and formulate SF allocation as a Bayesian game.

In the existing literature, SF allocation in LoRaWAN networks is typically based on distance analysis, which assesses the scalability of the network by considering the distance between end nodes and the gateway. The LoRa network employs the un-slotted ALOHA protocol for access control, as mentioned in [39]. ALOHA is a random access control protocol used in LoRaWAN networks, lacking contention techniques like CSMA/CD, which makes it susceptible to collisions. Packet collisions occur when multiple packets attempt to access the shared channel using the same Spreading Factor (SF), which can significantly impact the scalability of LoRaWAN. Therefore, SF allocation schemes play a crucial role in determining the scalability of LoRaWAN networks. The author has focused extensively on SF allocation strategies in addressing collision issues and enhancing scalability, as discussed in references [40–42].

## 2.4 Related work on UAV-enabled aerial networks

IoT-enabled monitoring and control systems are proliferating due to the easier availability of low-cost and small sensors [43]. Unmanned Aerial Vehicle (UAV) based monitoring is one of the potential applications of IoT [44]. The coordination of AI and IoT remarkably improves the computation speed and range of IoT devices used in industries, facilitating human decision-making at network edge [45]. AI and IoT jointly form a new term, namely Artificial Intelligence of Things (AIoT). AIoT simultaneously involves human and machine intelligence working together on the network edge of IoT devices. AIoT encourages faster predictive analytics in different aspects like scientific calculations, medical urgency, activities and environment monitoring [46]. UAV-enabled aerial networks act as an intermediatory component between the data acquisition units and Cloud [47]. The sensors at the acquisition units produce a vast amount of data. UAVs periodically collect and process this data and generate results for transferring to the Cloud for further action. The data processing and result generation on UAV is preferred over off-the-site (e.g., Cloud) to avoid colossal communication delays. The battery-operated UAVs provide service for pre-specified (or limited) time intervals due to faster battery depletion. However, adverse situations (like search and rescue operations during natural calamities, floods, etc.) demand an extended lifetime of a UAV-enabled aerial network. It is tedious to deploy DNN on UAV due to excessive resource demand. Researchers made significant efforts to reduce such resource demand of DNN [48,49].

However, this reduction often leads to compromises in accuracy. KD is a technique aimed at enhancing the performance of lightweight Deep Neural Networks (DNNs) by leveraging the generalization capabilities of larger models, as discussed in [50]. Furthermore, Artificial Intelligence of Things (AIoT) enables different power-saving modes in smartphones, allowing users to select modes that optimize operability on low power. Similarly, the longevity of aerial networks can be extended by dynamically selecting an appropriate DNN based on residual power levels. However, UAVs face constraints such as limited storage capacity, making it impractical to store multiple versions of DNNs in advance.

In [51], the authors employed random dropout to eliminate insignificant connections, but this approach often compromises performance by removing relevant connections. On the other hand, in [52], layer factorization was introduced to reduce floating-point operations. Additionally, in [53], the authors emphasized the concept of Once-For-All (OFA) networks, which generate various DNNs with different specifications to minimize training costs. To enhance the performance of compressed DNNs, the authors in [50] proposed a KD technique. This technique leverages the differences between the logits of a large-size DNN (teacher) and a compressed DNN to improve the student's generalization ability. Furthermore, the idea of layer sharing between teacher and student networks was introduced in [54], which aids in improving student performance while avoiding random initialization.

Innovative approaches to data processing are increasingly crucial due to the explosive expansion of the IoT, especially considering IoT devices' often constrained battery life and computing power. While significant research has been conducted on UAVs' deployment, trajectory planning, and resource allocation, less attention has been given to integrating MEC with UAVs to address IoT constraints. The author in [55] emphasizes a UAV-aided MEC system to reduce energy consumption for IoT devices and UAVs. This system optimizes job offloading, resource allocation, and UAV trajectory planning. The approach utilizes a Block Sequential Upper-bound Minimization (BSUM) method to address the non-convex nature of the problem. Simulations are conducted to demonstrate the effectiveness of the algorithm proposed.

IoT and smart devices have led to intelligent applications at the edge of wireless networks. This has enhanced the subject of edge learning, which combines wireless communication and machine learning. Using mobile edge computing and the massive amount of data dispersed over numerous devices, edge learning overcomes the restricted processing power and data at individual devices. The author [56] described design principles for learning-driven communication or wireless communication in edge learning.

Massive IoT is an emerging technology driven by the proliferation of sensors and wearable devices that continuously monitor and transmit data. Fifth-generation (5G) wireless networks are expected to offer enhanced capabilities and open up new markets across various use cases. In their work, the author in [57] proposes a novel approach to improve data collection in dense Wireless Sensor Networks (WSNs) by integrating projection-based Compressive Data Gathering (CDG) with Unmanned Aerial Vehicles (UAVs). CDG optimizes the lifetime of WSNs by reducing transmissions and energy consumption through data consolidation at projection nodes (heads). UAVs are then used to efficiently transmit the aggregated data to a distant sink, avoiding the need for multi-hop or long-distance communications. The study formulates a joint optimization problem that includes head selection, routing tree construction, clustering, and UAV trajectory planning to maximize the efficiency of the data-gathering process.

## 2.5 Conclusion

This work has considered various problems and challenges with the advancement of IoT applications using machine learning. The main objective of this thesis was to create efficient, effective, and secure long-range communication in smart IoT applications. This work focuses mainly on two tasks, i.e. solving imperfect labels in federated learning in intelligent transportation systems based on LoRa communication and interference communication in the LoRa network. While distributed training model in the federated learning environment. While training the model for locomotion modes recognition, we face the problem of imperfect label data in federated learning using LoRa communication protocol. We also solve interference problems based on Signal noise ratio(SNR) in LoRa networks in smart building applications. Federated learning-based IoT application is robust and secure and also improve bandwidth utilization. These problems are common and frequently occur in IoT applications in the distributed training learning environment. This thesis emphasized three major research contributions and one research enlist challenge in agriculture from an IoT perspective. Some challenges with existing techniques are mentioned as follows: Long distances, erratic wireless connectivity, and restricted bandwidth can all make it difficult for vehicles (participants) to send huge volumes of transportation data to the central server. This is addressed by long-range communication technologies such as LoRaWAN; however, they still have interference issues. The dynamic nature of vehicular environments faces challenges to manage stable and reliable communication between moving vehicles and the central system. As vehicles move, network connectivity may become unstable, leading to data loss or delays. In ITS, the data collected from different vehicles and infrastructure sensors may be highly heterogeneous (Non-IID), as each vehicle may encounter different traffic conditions, routes, or environments. ITS Systems require real-time data processing to ensure prompt action, such as managing incidents, modifying traffic signals, or providing route suggestions. In the FL system, sharing data among devices and participants may concern of privacy and security problems, particularly when it comes to sensitive transportation data like speed, location, and travel habits.