DESIGN AND IMPLEMENTATION OF AN IOT PLATFORM FOR FLOOD PREDICTION

MEET DIVYESH MEHTA

A THESIS

IN

THE DEPARTMENT

OF

COMPUTER SCIENCE AND SOFTWARE ENGINEERING

PRESENTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF COMPUTER SCIENCE

CONCORDIA UNIVERSITY

MONTRÉAL, QUÉBEC, CANADA

AUGUST 2023

© MEET DIVYESH MEHTA, 2023

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Meet Divyesh Mehta**

Entitled: **Design and Implementation of an IoT Platform for Flood Prediction**

and submitted in partial fulfillment of the requirements for the degree of

Master of Computer Science

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining commitee:

______________________ Chair
Dr. Sandra Cespedes

______________________ Examiner
Dr. Juergen Rilling

______________________ Supervisor
Dr. Brigitte Jaumard

______________________ Co-supervisor
Dr. Tristan Glatard

Approved by ______________________
Dr. Leila Kosseim, Graduate Program Director

__________ 2023

____
Dr. Mourad Debbabi, Dean
Gina Cody School of Engineering and Computer Science

Abstract

Design and Implementation of an IoT Platform

for Flood Prediction

Meet Divyesh Mehta

Flooding, a major natural calamity, severely threatens communities and infrastructures in areas susceptible to floods. Consequently, implementing an Internet of Things (IoT)-based flood monitoring system becomes crucial. Existing flood monitoring systems lack a comprehensive and scalable IoT platform to collect real-time data from diverse sensors efficiently, visualize flood information, and provide accurate water level forecasts. This thesis proposes a complete system designed to address the challenges associated with efficient data collection and flood monitoring from diverse Internet of Things (IoT) sensors.

Our proposition involves creating and deploying a centralized system known as HYDROSIGHT, which facilitates the real-time gathering, monitoring, and visualization of flooding-related sensor data. HYDROSIGHT system also provides a log monitoring feature for effective debugging and troubleshooting. The IoT environment for flood monitoring and prediction system was designed to promote sustainability and autonomy by preferring sensors with minimal footprints and compatibility with solar panels. The system architecture leverages a 4G network for seamless data transmission.

To validate the practical applicability of the proposed design, HYDROSIGHT system was tested at two municipalities of Quebec, namely Terrebonne, and Lac-Supérieur. In addition, the platform was also deployed at the Ericsson facility in Montreal to test the 5G capabilities. The deployment in these locations allowed us to evaluate the performance and effectiveness of the HYDROSIGHT system in a real flood monitoring environment.

In addition to implementing the IoT testbed, a preliminary machine learning tool was developed on water level forecasting. In this experiment, we opted for an online machine-learning approach, recognizing the significance of real-time updates and low computational resources of IoT devices. Leveraging the constantly updating data from HYDROSIGHT, we trained and tested our online machine-learning model, enhancing its forecasting capabilities.

We conducted a comparative analysis to understand the advantages of online machine learning over traditional batch learning. This analysis involved examining the water level forecasting results obtained from both methods using time series data from the HYDROSIGHT system deployed at Lac-Supérieur in Quebec.

Keywords: Internet of Things (IoT), MQTT protocol, data collection, centralized monitoring, real-time analytics, online machine learning, flood prediction, water level forecasting.

Acknowledgments

I want to thank my research supervisor, Dr. Brigitte Jaumard, for her unwavering support, invaluable guidance, and understanding throughout this work. Her mentorship played a pivotal role in the successful completion of this thesis. I would also like to thank Dr. Tristan Glatard for accepting to be a co-supervisor for this project.

I am deeply thankful to Mitacs and Concordia University for their financial support, which made this research endeavor possible.

I am also indebted to my exceptional team at Ericsson, particularly Emmanuel Thepie Fapi, Pierre Thibault, and Jean Michel Sellier, for their outstanding collaborations. Their support, insightful ideas, and fruitful discussions have been instrumental in shaping this collaborative thesis project. Their constructive feedback has been invaluable and significantly contributed to the overall quality of the work.

I want to extend my heartfelt gratitude to the two municipalities of Quebec, Terrebonne and Lac Supérieur, for their invaluable support and assistance in providing arrangements to conduct the necessary tests for this research.

I appreciate Bhargav Yagnik, with whom I had the privilege of collaborating on this research project.

Finally, I would like to thank my family and friends for their unwavering encouragement and support. Their belief in me has been a constant source of motivation, and without them, this project would not have been possible.

Contents

List of Figures

List of Tables

Acronyms

AMQP Advanced Message Queuing Protocol. 7, 8

CoAP Constrained Application Protocol. 7, 8

HTTP Hypertext Transfer Protocol. 7

IoT Internet of Things. iii, iv, 2–4, 6, 10, 12, 16

MQTT Message Queue Telemetry Transport. iv, 7, 8, 15, 23

MQTT-SN MQTT for Sensor Networks. 8

Chapter 1

Introduction

1.1 General Background and Motivation

The National Weather Service describes floods as the ”situation when water surpasses the natural or artificial boundaries of rivers, streams, or other water bodies, or accumulates in low-lying regions due to drainage” [54].

This can happen for various reasons, such as heavy rainfall, melting snow, or the failure of manufactured structures such as dams or levees [55]. Flooding can cause significant damage to property and infrastructure and pose a severe risk to human life.

Flooding can have a range of impacts, depending on the severity and duration of the event. In the short term, flooding can disrupt transportation, communication, and other essential services, contaminate water supplies, and cause power outages. In the longer term, flooding can lead to property damage, loss of crops and livestock, and financial losses for individuals and businesses. Floods can also have broader societal impacts, such as displacement of communities and strain on emergency services.

Flooding significantly impacts Canada, with the Insurance Bureau of Canada reporting that it is the country’s most common and costly natural disaster. In Canada, flood losses totaled $405 million between 1983 and 2008 and cost $1.8 billion between 2009 and 2017 [26].

In Canada, the rise in population and infrastructure development in regions prone to flooding, combined with climate change, are expected to lead to a higher frequency of flooding incidents in coastal and urban regions [58]. Approximately 80% of Canadian cities are situated on floodplains along rivers, resulting from historical settlement patterns that favored locations near waterways for trade and natural resources access [18].

There is a strong need for an IoT system for flood monitoring and prediction, as floods are among the most destructive and costly natural disasters. An IoT system can provide early warning of potential flooding, enabling emergency responders and residents to take necessary precautions to protect themselves and their property. This can save lives and reduce the economic and societal impacts of floods.

In addition to providing early warnings, an IoT system can also help improve our understanding of the factors contributing to flooding. By collecting and analyzing large amounts of data from sensors in real-time, the system can identify patterns and trends that may not be immediately apparent. This can help us better understand the causes of floods and develop more effective strategies for mitigating their impacts. An IoT system can also be integrated with other technologies to provide a more comprehensive view of the situation, enabling more effective flood management efforts.

This project is a multidisciplinary effort that involves three main components: the design of an IoT system, HYDROSIGHT (the focus of this thesis), developing a machine learning prediction model for making accurate forecasts of potential flooding, and creating a digital twin of the target area to enable real-time visualization and analysis of flood-related data. The integration of these components will allow the development of a comprehensive and effective system for predicting and managing floods.

1.2 Research Project: Design and Implementation of an Energy Autonomous IoT Platform for Flood Prediction

Flood monitoring is crucial for effective disaster management and resilience in flood-prone regions. However, current flood monitoring systems lack a comprehensive and scalable Internet of Things (IoT) system that can efficiently collect data from diverse sensors and provide real-time flood data visualization. The sensors used in the current systems may not provide precise and reliable data, leading to inaccurate flood forecasts and hindering the ability to make informed decisions during flood events. Furthermore, these systems lack a comprehensive, scalable solution with a robust IoT environment. The lack of scalability limits the system's ability to handle a large volume of sensor data, especially during extreme flood events when data collection is crucial. In addition, the current IoT system's limitations are in data collection efficiency, which hampers its overall performance in flood monitoring. Inefficient data collection mechanisms may result in loss of data, leading to incomplete datasets and reduced predictive accuracy. This gap in the existing technology hinders timely decision-making during flood events and compromises the accuracy of flood predictions, leading to potential challenges in disaster preparedness and response.

The proposed research aims to develop an advanced IoT system called HYDROSIGHT, which will contribute to the existing knowledge base by solving the shortcomings in current flood monitoring systems. By seamlessly integrating various sensors and leveraging the 4G network for efficient data transmission, HYDROSIGHT will enhance data collection efficiency and enable real-time monitoring and visualization of flood data. Additionally, the research will explore an online machine learning approach for water level forecasting, improving prediction accuracy.

The significance of this research lies in its potential to transform flood monitoring and disaster management strategies. With an efficient and scalable IoT system like HYDROSIGHT, decision-makers can access accurate and timely data, enabling proactive responses to flood events and enhancing community safety. The research findings will fill the gap in current knowledge, improving flood resilience and disaster preparedness in vulnerable regions, ultimately contributing to the advancement of flood monitoring technology and its practical applications.

Three deployments of the HYDROSIGHT platform have been carried out: two in the municipalities of Terrebonne and Lac Supérieur and one in the premises of Ericsson in St Laurent. As only 4G was available at the locations of these two municipalities, motivation of one on the roof of the parking of Ericsson was to perform tests with 5G, unfortunately, we could not find any 5G data logger available on the market, which would satisfy our design requirements.

There are currently at least two different initiatives in Quebec in order to watch water levels in rivers in order to monitor flood events.

The first initiative is with Hydro Meteo and Geosapiens and the E-nundation product [17]. This last one relies on water sensors only which are powered with an electrical cable, requiring a dedicated electrical pole. In addition, water level prediction only relies on historical water levels.

The second initiative is with CMM - Communauté Métroplitaire de Montréal and the Grand Crues project [6]. There is a plan to deploy a total of 29 measuring stations will collect water level data from all over the metropolitan territory and will make it possible to monitor the behavior of the main watercourses and to know, through a website, forecast for 3 days. Again, these water measuring stations are powered with electrical cables, and only use historical water levels to predict the water levels.

1.3 Contributions of the Thesis

In this thesis, we have significantly contributed to flood monitoring by designing and developing an integrated IoT environment and a scalable monitoring platform.

The following contributions summarize the key achievements of this study:

- Developed an integrated IoT environment that seamlessly integrates various sensors (Ultrasonic Snow Depth Sensor(260-700-D), Water Level Sensor(MX2001-01-SS-S), and Rain Metric(RXW-RGF-900) Sensor) while considering factors like cost, connectivity, sustainability, and accuracy.

- Created a scalable and robust data transmission and collection system leveraging the 4G network for efficient communication between IoT devices.

- Design a centralized system for monitoring, visualizing, and analyzing real-time flood data from the integrated IoT sensors.

- Validated the practical applicability of the proposed system through testing in flood-prone areas (Terrebonne, Lac-Supérieur).

- Explored online machine learning approach for water level forecasting using real-time data collected from HYDROSIGHT.
- Conducted a comparative analysis between online and traditional batch learning approach for water level forecasting.

1.4 Organization of the Thesis

Chapter 2 presents a literature review of related subjects, including recent studies on flooding prediction environments, the Internet of Things(IoT), Communication Protocols for IoT, and Online Machine Learning. Chapter 3 puts forward the system architecture of HYDROSIGHT system. Chapter 4 ensembles and discusses online machine learning for water level forecasting, using some of the data collected by the IoT platform. Conclusions and future work are discussed in the last chapter.

Chapter 2

Literature Review

In this chapter, we present related works on three items: first, on the IoT platforms already designed or developed for flood prediction, and second on the communication protocols used for IoT environments. In addition to that, we discuss work related to Online Machine Learning.

2.1 IoT platform for Flood prediction

Several researchers have explored using Internet of Things (IoT) in flood monitoring platform, presenting innovative approaches to integrate sensors, communication networks, and cloud computing for data collection and processing. Satria *et al.* [50] and Ancona *et al.* [2] propose flood monitoring platforms that leverage IoT devices and networks to collect real-time data on water levels, flow rates, and rainfall. The advantages highlighted in their works include improved flood response, better flood management, and timely alerts to stakeholders such as emergency responders and local authorities. Rani *et al.* [43] and Bande and Shete *et al.* [3] focus on developing smart flood monitoring platforms that incorporate IoT devices and sensors, with an additional emphasis on the integration of machine learning and neural networks for flood prediction. Rani *et al.* [43] employ machine learning algorithms to analyze collected data and provide accurate flood predictions. On the other hand, Bande and Shete *et al.* [3] used neural networks to predict flood severity based on data from water level and flow rate sensors. Both studies highlight the advantages of real-time data processing and accurate flood predictions that could significantly contribute to effective emergency planning and response strategies. A study introduces a novel flood monitoring and warning system (FMWS) that leverages the capabilities of long-range wide area networks (LoRaWAN) to monitor water levels in catchment areas[62]. The system uses smart sensing units to collect and transmit data to the cloud. The system is designed to be low-cost and easy to deploy in catchment areas.

The research of this thesis presents the design and the implementation in a real environment of an advanced flood monitoring IoT system, HYDROSIGHT, that ensures enhanced robustness and reliability through rigorous data checks for accurate sensor readings. The HYDROSIGHT features a centralized real-time monitoring platform, offering stakeholders a comprehensive and easily accessible overview of flood-prone areas. The key differentiator of the system is its high scalability by leveraging containerization techniques, allowing for efficient adaptation to varying data loads and potential future expansion. This makes it a versatile and adaptable solution suitable for different environments and regions.

2.2 Communication Protocol

The Internet of Things (IoT) has attracted considerable attention as a transformative technology with far-reaching applications One crucial aspect of the IoT is the selection of appropriate communication protocols, which enable efficient device-to-device and device-to-internet communication. In this context, several research papers have conducted comparative evaluations of various IoT protocols to assess their performance characteristics and suitability for different applications. Several authors, e.g., Elhadi *et al.* [12], Gemirter *et al.* [16], Pohl *et al.* [41] and Fysarakis *et al.* [15], conducted comparative studies on IoT protocols, including CoAP, MQTT, AMQP, and HTTP. These studies assessed metrics like message latency, size, delivery time, and network utilization to determine the strengths and weaknesses of each protocol. Among the findings, Message Queue Telemetry Transport (MQTT) stood out for its overall strong performance with low message latency, high message throughput, and suitability for real-time applications. Constrained Application Protocol

(CoAP) and Advanced Message Queuing Protocol (AMQP) showcased advantages in specific scenarios, such as constrained devices and applications with limited network bandwidth. De Caro *et al.* [8] and Chen and Kunz *et al.* [59] focused on evaluating the MQTT protocol against other alternatives. De Caro *et al.* [8] compared MQTT-SN and CoAP for smartphone-based sensing, with MQTT-SN demonstrating lower message latency and smaller message size than CoAP. On the other hand, Chen and Kunz *et al.* [59] assessed MQTT and CoAP under a constrained wireless access network, finding MQTT to have the highest message delivery rate, making it preferable for real-time applications.

In contrast to the reviewed literature that covered various IoT protocols and their performance comparisons, our research centered on leveraging the Message Queue Telemetry Transport (MQTT) protocol for sensor data transmission. Our specific focus on MQTT was driven by its lightweight design and efficient message delivery features, making it highly suitable for resource-constrained devices and real-time data communication.

2.3 Online Machine Learning

Online learning is closely related to other areas like adaptive learning, continual learning, incremental learning, sequential learning, and these terminologies may often be confused with one another. The learner in adaptive learning attempts to adapt the learning model for dynamically changing environments and models can utilize online learning in such environments but it is not limited to online learning and heuristic adaptation or modifications of batch learning algorithms can be utilized which evolve to respective environment changes [20]. Continual learning is often termed as 'lifelong learning' and inspired from humans ability to continuously learn new tasks while being good at other tasks through our lifetime. While contiual learning is closely related to online learning, existing studies follow paradigm of batch training to learn on existing tasks or new tasks. Incremental learning refers to learning from stream of data samples in constrained spaces to address efficiency and scalability. Incremental learning can be viewed as a branch of online learning and extension for adapting traditional offline learning techniques in data-stream settings

[44]. Sequential learning revolves around the learning from sequential training data in which the order of data is very important [9]. Sequential learning can be solved using either batch or online learning algorithms.

Several Neural networks-based learning approaches [30, 57] followed this branch of online learning, and the Perceptron could be viewed as the simplest form of online learning. One extensively studied approach in online/stochastic gradient descent utilizes the efficient back-propagation algorithm for online learning [30]. With advancements in machine learning, several attempts have been made to make deep learning compatible with online learning [31, 63]. Sahoo *et al.* [49] proposed Hedge backpropagation to learn deep neural networks in an online setting and address slow convergence of deep networks through dynamic depth adaptation. Hoi *et al.*[20] provides a comprehensive overview of online learning techniques, covering various algorithms and methodologies with a main focus on approaches in online supervised learning and online learning with partial feedback.

In this section, we highlight the developments mainly focused on online machine learning, and their applications in IoT-based environments

2.3.1 Online learning in time series

Online learning is used for time-series problems where data arrives sequentially and needs to be processed and predicted in an ongoing manner, such as IoT sensor data [24, 27], stock prices [51], and network traffic [36], among others. Kraemer *et al.*[27] simulated online machine learning training to predict 1-day photovoltaic energy. Improvement of 56% is observed on the online machine learning model using Random Forest Regressor(RFG), compared to persistent predictor (termed LAG). Singh *et al.* [51] compared incremental learning using linear regression with online-offline learning approach utilizing a wide range of deep learning models like LSTM, Stacked LSTM, Bi-LSTM, CNN, and CNN-LSTM. Results showed that BiLSTM performed the best compared to other Incremental learning approaches. The online-offline model used by Singh *et al.*[51] was trained at the end of every trading session. Melgar *et al.*[32] devised a novel forecasting algorithm for streaming time series called StreamWNN. StreamWNN utilizes K-Nearest neighbors and starts with an offline stage in which a forecasting model is created based on historical data, and further, the model is incrementally updated in the online stage with buffer data. One-day updates on model-trained electrical energy consumption performed better than no updates or updating the model monthly or quarterly. Multiple studies have demonstrated improvements in training time-series data using online machine-learning approaches, motivating further experimentation to evaluate the performance of online machine-learning on our specific dataset.

2.3.2 Flood forecasting systems using online learning

Yu *et al.*[61] used an ensemble method comprising dynamic evolving neural-fuzzy inference system or DENFIS. DENFIS utilizes Evolving Clustering Method (ECM) where clusters get regularly modified during online learning. This approach fails when the observed water-level values have never occurred in the training or update phase, and time-order not being maintained in cluster organization. Thus prediction would be inaccurate compared to other supervised models. [60] compared the performance of Support Vector Machines (SVM) and Gated Recurrent Unit(GRU) model on the historical dataset (1981-1986) with online learning (termed as incremental updates) and showed that improvement was observed in the rainfall-runoff prediction.

2.3.3 IoT with online learning

Internet of Things (IoT) based machine learning systems integrate connected devices and sensors with machine learning algorithms to enable data collection, analysis, and real-time decision-making. Online Machine Learning has been tested on microcontrollers integrated into the IoT in diverse domains, including agriculture [38], mobile and wearable devices [24], energy sensors [27], and other areas. Due to Microcontrollers' limited storage capacity and low energy consumption, online machine learning is a suitable option for these devices. Tiny Online Machine Learning(TinyOL) [45] libraries are developed for incremental training on-device for streaming data. Such libraries aim to run training of ML models on low compute hardware of 64MHz CPU with 256 KB RAM along with battery consumption of 0.1W.

Despite the progress in algorithms and libraries for IoT-based online machine learning systems, there is a limited amount of research in the literature that integrates real-time online machine learning methods using water-level data from IoT sensors for flood forecasting. Our aim is to address this gap with our work.

Chapter 3

Design of HYDROSIGHT system

This chapter focuses on the design and architecture of the HYDROSIGHT system aimed at addressing the challenges of real-time data collection, monitoring, and visualization of Internet of Things (IoT) sensors for flooding. The primary objective of this chapter is to provide a detailed explanation of the system's architecture and its various components.

By thoroughly exploring the HYDROSIGHT architecture, we aim to understand how the system operates and how data flows among its components.

3.1 System Requirements

The system requirements for the proposed IoT system, named HYDROSIGHT, consist of functional and non-functional requirements that must be addressed for successful implementation.

3.1.1 Functional Requirements

The functional requirements of the HYDROSIGHT system encompass crucial aspects such as data collection and processing, real-time monitoring, data storage, integration with the 4G network, and logging functionality for troubleshooting and debugging. The system must efficiently collect and process data from the snow, water level, and rain metric sensors to provide comprehensive flood-related insights. HYDROSIGHT collects parameters like rainfall, snow depth, water temperature, water flow, and water level. It should offer real-time monitoring and visualization of the collected data, presented through a user-friendly interface for straightforward interpretation. Reliable data storage capabilities must be provided to retain historical sensor readings for future analysis. Additionally, the system should have a logging functionality to record and store system activities, aiding in troubleshooting and debugging any issues that may arise during operation. Lastly, the system should seamlessly integrate with the 4G network to facilitate efficient data transmission, ensuring a robust and responsive flood monitoring mechanism.

3.1.2 Non-Functional Requirements

In addition to the functional aspects, the system's non-functional requirements are equally crucial for its success. The system must demonstrate scalability to accommodate the increasing number of stakeholders while maintaining optimal performance. Reliability is paramount to ensuring continuous operation with minimal downtime, contributing to practical disaster preparedness and response. To minimize energy consumption and promote sustainability, the system should use clean energy sources, such as solar panels, for power supply and low-power consumption devices. A user-friendly interface is essential to enable stakeholders to interact with and comprehend flood-related data and visualizations seamlessly. Reduced data transmission and processing latency are crucial for real-time monitoring and rapid response during flood events. Lastly, the system's compatibility with various IoT sensors, data loggers, and communication protocols will facilitate seamless integration and interoperability, enhancing the overall efficiency of the flood monitoring system.

3.2 High-Level Architecture

The proposed IoT system's high-level architecture, HYDROSIGHT, is designed to enable efficient data collection from diverse IoT sensors and real-time flood data monitoring and troubleshooting. The major components of the architecture include the IoT Sensors, Data Transmission Modules, Data Collection Framework, Data Storage Technology, and Monitoring platform.

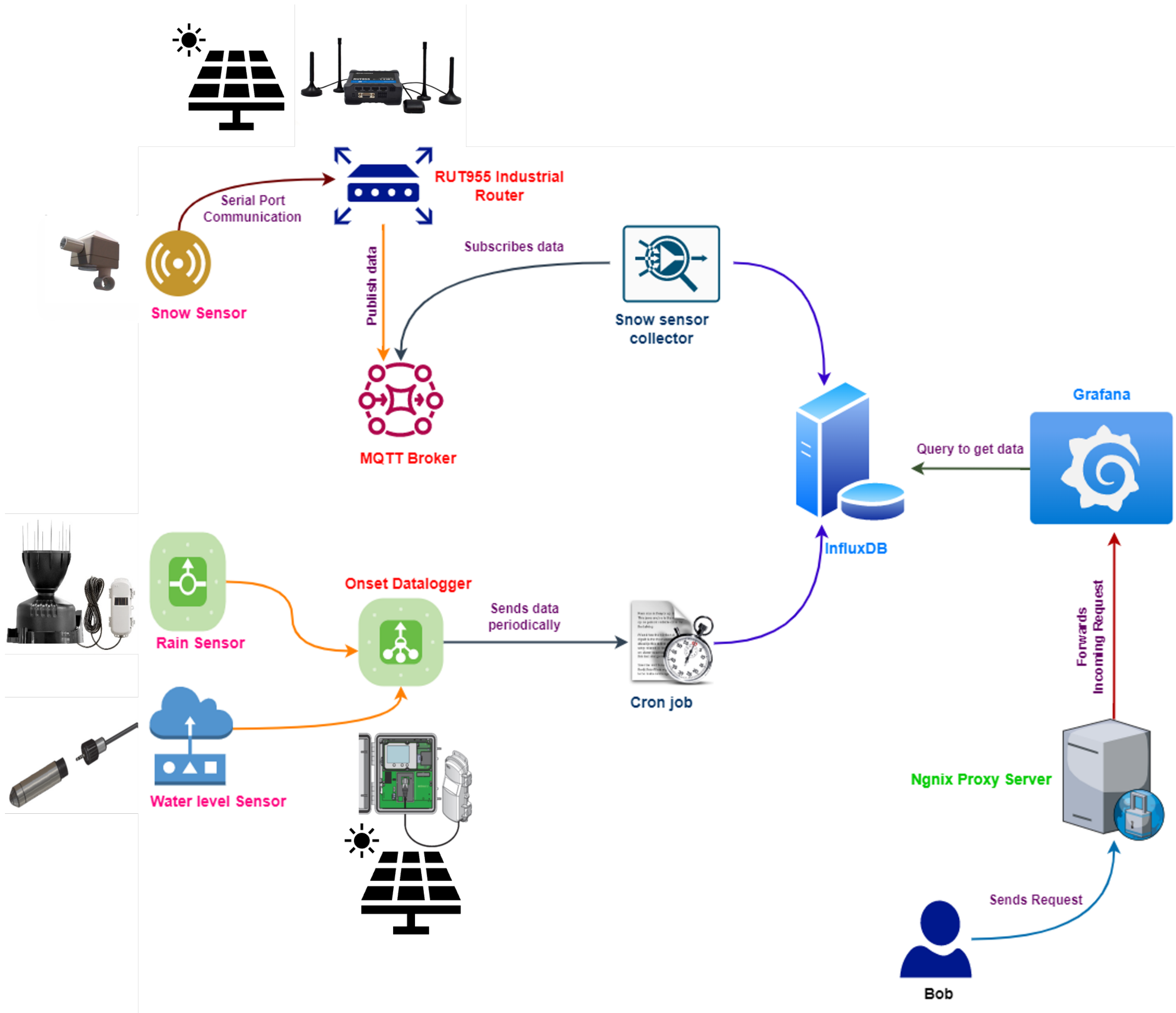


Figure 1: High-Level Software Architecture of HYDROSIGHT

- **IoT Sensors**

  IoT sensors play a vital role in collecting essential environmental data such as snow depth, water levels, temperature, and rainfall measurements. These parameters contribute to the inputs of flooding events. HYDROSIGHT system encompasses sensors like Ultrasonic Snow Depth Sensor(260-700-D), Water Level Sensor(MX2001-01-SS-S), and Rain Metric Sensor(RXW-RGF-900). Snow sensor communicates with data transmission component, RUT955, through RS232 serial port communication, while water level sensor and rain metric sensor communicate wirelessly through OnSet 4G Data logger.

- **Data Transmission Module**

  This component ensures efficient and reliable data communication from IoT sensors in the system in real time. The data transmission component comprises a RUT955 Industrial router and a 4G Data Logger Station(RX3004). The implementation of utilizing the 5G network for data transmission was not feasible due to the unavailability of a suitable data logger with 5G capabilities that also met the requirements.

  The router transmits data from the snow sensor, and the data logger transmits data from rain and water level sensors. These components leverage a 4G network to transmit data. Message Queue Telemetry Transport (MQTT) communication protocol is used for data transfer from the RUT955 router.

- **Data Collection Module**

  Data in the IoT Platform is collected through two methods. Firstly, the Snow Sensor data is transmitted in real-time using the MQTT protocol through the RUT955 Industrial Router. The server's **Snow Sensor Collector** program receives and processes this data for continuous updates and flood monitoring. Secondly, the Datalogger data is collected daily through a scheduled cron job. The Datalogger station stores the data locally on the server, and the cron job retrieves and transfers it to the database, providing historical data.

- **Data Storage**

  To efficiently handle the real-time data coming from the sensors in the HYDROSIGHT system, we have chosen to leverage the capabilities of InfluxDB, a time series database [25]. InfluxDB is well-suited for handling large volumes of time-stamped data and provides excellent performance for storing and querying time series data. For each deployment site of HYDROSIGHT system, we implement partitions in InfluxDB. These partitions are created to ensure data security and maintain isolation in data storage. By partitioning the data, we can independently isolate and manage data streams from different deployment sites, enhancing data organization and retrieval efficiency.

- **Monitoring Platform**

  In the HYDROSIGHT system, we utilize Grafana [19] as our monitoring and visualization tool. Grafana effectively queries the InfluxDB to obtain the required data, ensuring efficient and real-time data retrieval. The pull-based mechanism of Grafana contributes to reduced latency, allowing for timely updates and analysis of flood-related data. We implement a reverse proxy server to enhance security using Nginx [37]. This reverse proxy server handles incoming user requests and securely forwards them to Grafana. By implementing this security measure, we ensure that user data remains protected and accessible only to authorized personnel, enhancing the overall reliability and confidentiality of the HYDROSIGHT system.

3.3 Detailed Component Description

This section explains each significant component within the HYDROSIGHT system, outlining its purpose, functions, and responsibilities.

3.3.1 IoT Sensors

The sensors are critical in gathering real-time data on various environmental parameters such as snow depth, water levels, and rainfall. This data is essential for assessing flood risks; hence, selecting appropriate sensors is crucial in designing an effective and reliable Internet of Things (IoT) environment.

The selection process for IoT sensors involves carefully considering key factors to ensure the system's effectiveness. Connectivity requires sensors with 4G and 5G capabilities for seamless data transmission. Due to the limited availability of 5G connectivity in municipalities outside Montreal, sensors equipped with 4G capabilities were deemed suitable for the system. Power supply plays a vital role, favoring sensors compatible with solar panels to promote sustainability and autonomy. Robustness is essential, necessitating sensors with a wide operating temperature range and high environmental rating to withstand challenging conditions like harsh winters with -40°C temperature.

Ensuring high accuracy and a wide operating range is paramount to achieve precise measurements. Additionally, cost-effectiveness and accessibility are integral factors considered during the consideration process.

3.3.1.1 Ultrasonic Snow Depth Sensor

The 260-700-D Ultrasonic Snow Depth Sensor (Figure 2) is a cost-effective device that provides accurate and non-contact snow depth measurements. Employing ultrasonic technology, the sensor emits sound waves toward the ground and calculates the time it takes to reflect from nearby objects within a defined area. The sensor delivers precise and reliable range readings by processing this time of flight information, enabling accurate snow depth.

The sensor performs two consecutive measurements during each cycle and compares them to enhance measurement accuracy. Upon completing each measurement cycle, the sensor outputs four essential values in ASCII format: temperature, time of flight, distance, and the number of retries made. The data is formatted for straightforward interpretation during transmission, with tab characters (ASCII 09) used to separate the values, and the message is completed with a carriage return (ASCII 13) and line feed (ASCII 10). During each measurement cycle, the 260-700-D Snow Sensor facilitates data transmission using RS232 serial port communication.

A detailed explanation of the working of 260-700-D is provided in Appendix A.1.

Figure 2: 260-700-D Ultrasonic Snow Depth Sensor [39]

3.3.1.2 Water Level Sensor and Water Level Sensor Module

The MX2001-01-SS-S water level sensor (Figure 3) is designed for measuring water level, water flow, water temperature, barometric pressure, and differential pressure. The sensor is equipped to measure atmospheric and water pressure at its location. By taking the difference between the atmospheric pressure and the water pressure reading obtained from the sensor, the system can determine the hydrostatic pressure caused explicitly by the water column's height. This pressure difference serves as a direct indicator of the water level.

Figure 3: MX2001-01-SS-S Water Level Sensor [34]

The RXMOD-W1 Water Level Sensor Module (Figure 4) complements the functionality of the MX2001-01-SS-S Water Level Sensor, enhancing its capabilities and ease of use. The water level sensor module interfaces with the MX2001-01-SS-S sensor, processing its data and transmitting it to RX3004 4G data logger.

Additional information on the MX2001-01-SS-S water level sensor and RXMOD-W1 Water Level Sensor Module is provided in Appendix A.2.

Figure 4: RXMOD-W1: Water level Sensor Module [47]

3.3.1.3 Rain Metric Sensor and Wireless Manager Module

The RXW-RGF-900 rainfall sensor (Figure 5) is designed to record rainfall in 0.2-millimeter increments. This rainfall sensor has a wireless feature enables it to communicate data directly to RX3004 datalogger over a 4G network.

Figure 5: RXW-RGF-900 Rain Metric Sensor [48]

When rain falls, it enters the sensor's collector cone and passes through a debris-filtering screen, ensuring that any unwanted particles or debris. The rainwater then collects in the chamber of the tipping mechanism. To transmit the collected rainfall data to the RX3004 4G data logger station, the rainfall sensor utilizes the RXMOD-RXW-900 Wireless Manager Module (Figure 6). This module facilitates wireless communication between the rain sensor and the data logger, enabling real-time and continuous monitoring of rainfall data.

Figure 6: RXMOD-RXW-900 Wireless Manager Module [48]

3.3.2 Data Collection and Transmission

This section focuses on the crucial data collection and transmission aspects fundamental to our IoT flood monitoring platform. The RUT955 industrial router and the RX3004 4G Data logger station facilitate smooth and efficient data exchange from the sensors. These reliable and advanced devices are essential for acquiring, transmitting, and monitoring real-time data, making them integral components of HYDROSIGHT system.

3.3.2.1 Industrial Router

The RUT955 Industrial router is a compact, high-performance device with multiple connectivity options, including 4G LTE. The industrial-grade design of the router makes it durable and resistant to damage.

This router was selected based on specific criteria to meet essential project requirements. Its RS232 port was a vital communication interface for the Novalynx 260-700-D snow depth sensor, ensuring efficient data transmission. The router's industrial-grade build and durability made it well-suited for harsh environments, guaranteeing reliable performance outdoors. With 4G technology support, it established a dependable uplink connection, enabling uninterrupted data transmission and real-time monitoring of the snow sensor's readings.

Figure 7: RUT955 Industrial Router [46]

3.3.2.2 Data Logger Station

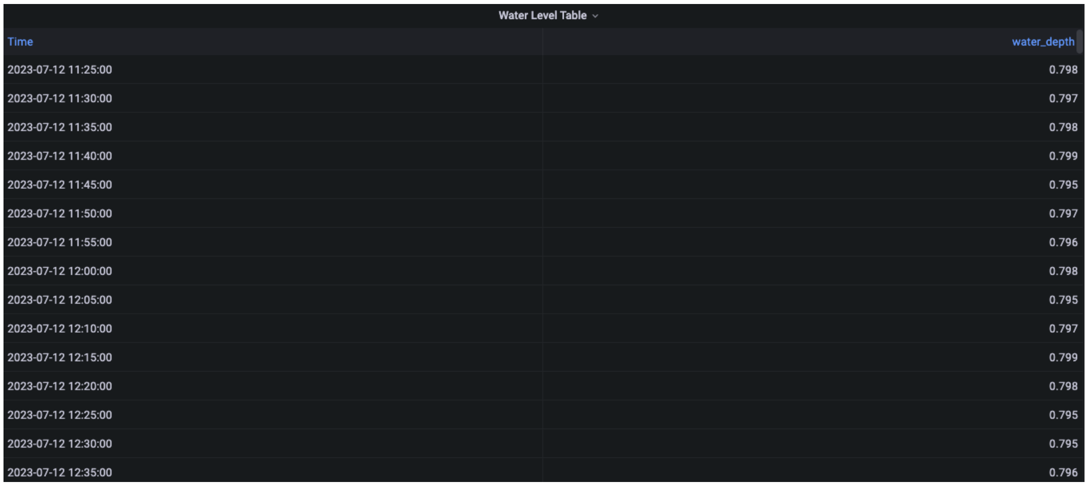
The RX3004 4G data logger station is pivotal in modern data acquisition and monitoring systems, offering advanced features that optimize data collection and analysis efficiency with high precision. This data logger station facilitates seamless communication and transmission as a crucial intermediary between sensors and the central data management system. Leveraging its 4G connectivity, the RX3004 efficiently handles multiple sensor inputs and processes and transmits them to the designated server.

Figure 8: RX3004 4G Data Logger [7]

3.3.3 Monitoring Platform

The monitoring platform of HYDROSIGHT serves as a central hub for real-time data visualization and analysis. It provides a comprehensive view of the flood monitoring system's performance and enables effective decision-making. The platform incorporates two key components: a tabular data display and a time series graph representation of water depth.

The tabular data, as shown in Figure 9, showcases the latest sensor readings and relevant information in a structured and easily interpretable format. This table presents essential parameters, such as water level(in meters), collected from the various sensors deployed in flood-prone areas. It offers an instant snapshot of the current environmental conditions, allowing users to monitor changes and take appropriate actions promptly.



Water Level Table

Time	water_depth
2023-07-12 11:25:00	0.798
2023-07-12 11:30:00	0.797
2023-07-12 11:35:00	0.798
2023-07-12 11:40:00	0.799
2023-07-12 11:45:00	0.795
2023-07-12 11:50:00	0.797
2023-07-12 11:55:00	0.796
2023-07-12 12:00:00	0.798
2023-07-12 12:05:00	0.795
2023-07-12 12:10:00	0.797
2023-07-12 12:15:00	0.799
2023-07-12 12:20:00	0.798
2023-07-12 12:25:00	0.795
2023-07-12 12:30:00	0.795
2023-07-12 12:35:00	0.796

Figure 9: HYDROSIGHT: Water Depth Table [19]

On the other hand, the time series graph representation, as shown in Figure 10, presents a graphical visualization of the water depth data collected over a specific period. This graph allows users to track water level variations over time, identifying trends and patterns in flood occurrences. The time series graph is an invaluable tool for understanding the dynamics of water levels, helping authorities and decision-makers make data-driven decisions for flood management and preparedness.

HYDROSIGHT leverages Loki, a powerful logging solution, to enable efficient data logging and real-time visualization. Loki functions as a logging backend, allowing the system to store efficiently and index vast amounts of log data from flood monitoring sensors and devices. Loki utilizes a lightweight indexing system, significantly reducing the overhead associated with traditional logging solutions, making it ideal for the IoT flood monitoring platform.

Furthermore, Loki is seamlessly integrated with the monitoring platform, providing a user-friendly interface for querying and visualizing the log data. It enables users to search and filter the log data based on various parameters and time frames, facilitating quick and precise analysis of flood-related events.

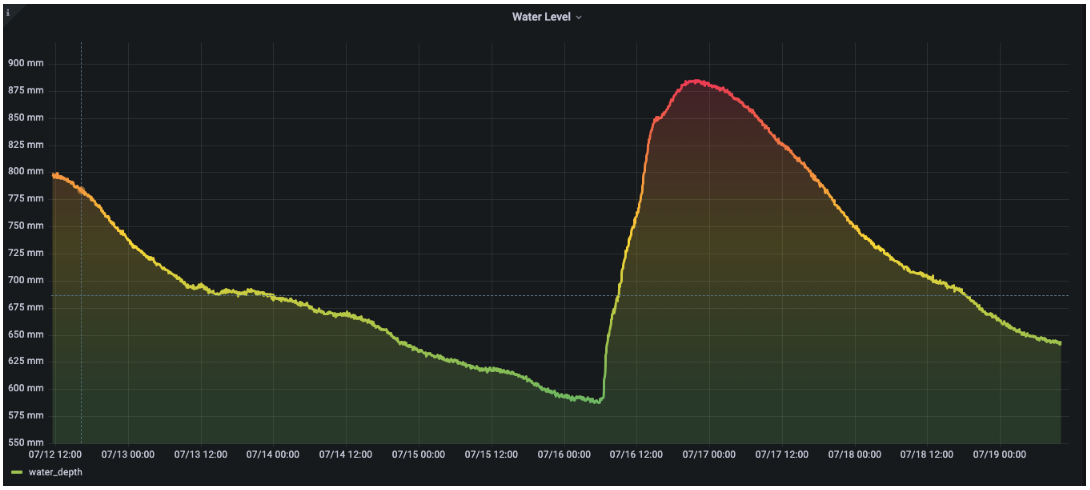


Figure 10: HYDROSIGHT: Water Depth Visualization [19]

3.4 Data Flow and Communication

This section delves into the detailed data flow and communication aspects within the HYDROSIGHT. A complete understanding of these processes is crucial for grasping the seamless operation and efficiency of the entire system. We will explore how data travels through the various components, from its initial collection at the deployed IoT sensors to its storage and transmission for real-time monitoring and visualization. Moreover, we will shed light on the data communication protocols and mechanisms employed between the components, emphasizing their role in facilitating smooth and reliable information exchange. By explaining the data processing, storage, and transmission procedures, we aim to provide a holistic view of how the HYDROSIGHT system efficiently handles and leverages the vast array of data for accurate flood monitoring and prediction.

3.4.1 Communication Protocol

This section explains the publish/subscribe communication model and briefly explores the Message Queue Telemetry Transport (MQTT) [33] communication protocol.

3.4.1.1 Publish/Subscribe Systems

The publish/subscribe (pub/sub) communication model operates based on a fundamental principle: components seeking specific information register their interest, a process known as subscription, thus becoming subscribers. On the other hand, components wishing to disseminate information do so by publishing their data, and they are referred to as publishers. Acting as a mediator, the Broker ensures that publisher data reaches the subscribed subscribers. The Broker plays a pivotal role in coordinating subscriptions, and subscribers typically need to contact the Broker to initiate their subscription process directly.

There are three primary categories of pub/sub systems: topic-based, type-based, and content-based [13]. The communication model of a topic-based publish/subscribe (pub/sub) system is depicted in Figure 11. A subscriber sends a subscription message (sub(topic)) to the Broker to express interest in a specific topic. In contrast, a publisher sends a publication message (pub(topic, data)) containing the data to be disseminated, along with the associated topic. When there is a match between the publisher's and the subscriber's topics, the Broker forwards the publication message (pub(topic, data)) to the subscriber. If multiple subscribers have matching topics, a single publication message (pub(topic, data)) can be distributed to all.

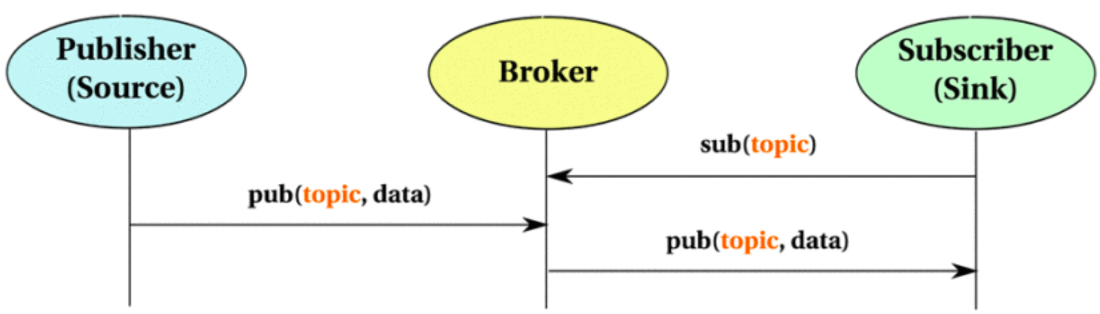


Figure 11: Topic-based Pub/Sub Communication Model [21]

3.4.1.2 Message Queue Telemetry Transport(MQTT)

In the context of IoT, effective communication between various devices holds paramount importance as it enables one IoT device to send instructions to another appliance for efficient system management [28].

Specifically, MQTT has found extensive application in various IoT devices and instant messaging systems due to its design for efficient operation on low-power machines, making it a lightweight protocol of choice [22].

3.4.1.3 MQTT Implementation in HYDROSIGHT

The MQTT architecture implemented in the HYDROSIGHT system comprises two essential components: the Client and the Broker[4]. The Client, which can act as a Publisher or a Subscriber, establishes a network connection with the Server (Broker). In the HYDROSIGHT system, the Snow Sensor is a Publisher, sending snow depth data in ASCII format via RS-232 serial port communication to the RUT955 Industrial router. The RUT955 router converts the received data to a string format and packages it as a message to be published to the MQTT Broker. The message is published under the specific topic ”snowsensor” at regular intervals of 5 minutes, ensuring real-time updates.

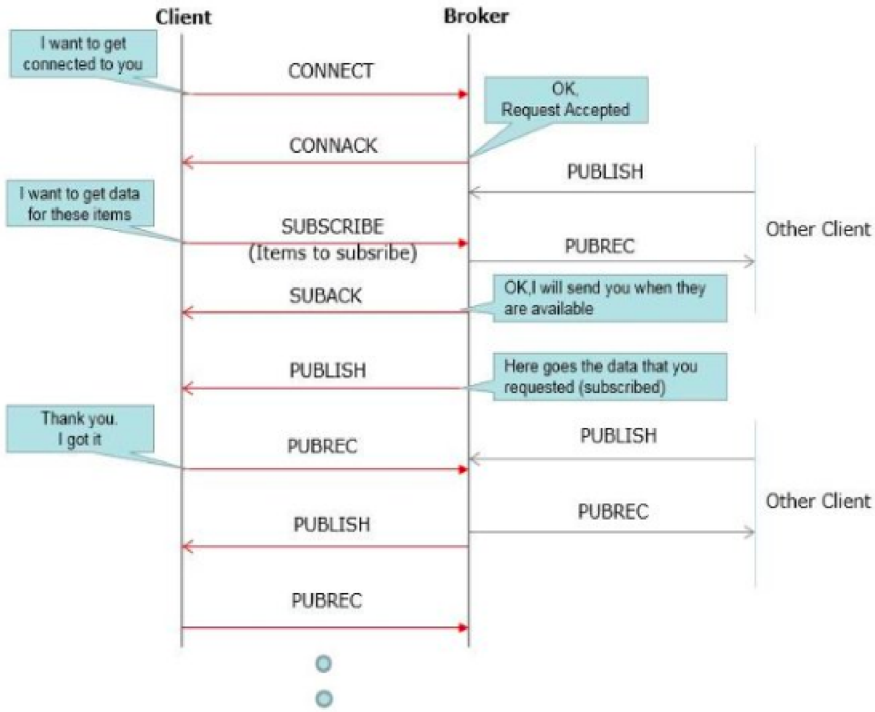


Figure 12: Working of MQTT [52]

The Broker plays a vital role in the HYDROSIGHT system's communication process. It receives messages from the Publishers, filters them, and identifies the Subscribers interested in specific information. In this case, the Snow Data Collector Service, a microservice acting as an MQTT Client, subscribes to the MQTT Broker for the "snowsensor" topic. As new data is received from the Broker, pre-processing of the data is performed to ensure accuracy and consistency.

3.4.2 Data Flow

This section provides an overview of the data flow within the HYDROSIGHT system, which is divided into two main parts: data coming from sensors to the server and the collection and storage of data in the persistent storage.

As shown in Figure 13, the first part of the data flow involves receiving data from the sensors through two different methods. The snow sensor sends data in ASCII format through RS-232 serial port communication to the RUT955 Industrial router. The data is converted to a string format in the router and packaged as a message to be published to the MQTT broker. The message is assigned to a specific topic, ”snowsensor,” and is transmitted every 5 minutes. The MQTT broker, implemented using Mosquitto, operates as a service in the backend server, receiving and distributing the messages to its subscribers.

On the other hand, the rain metric and water level sensors transmit various measurements (e.g., water level, rainfall, temperature, flow, and barometric pressure) every 5 minutes to the data logger station through the 4G network, using their respective sensor modules. The data logger station collects these sensor readings for 24 hours and exports them to the backend server in CSV format.

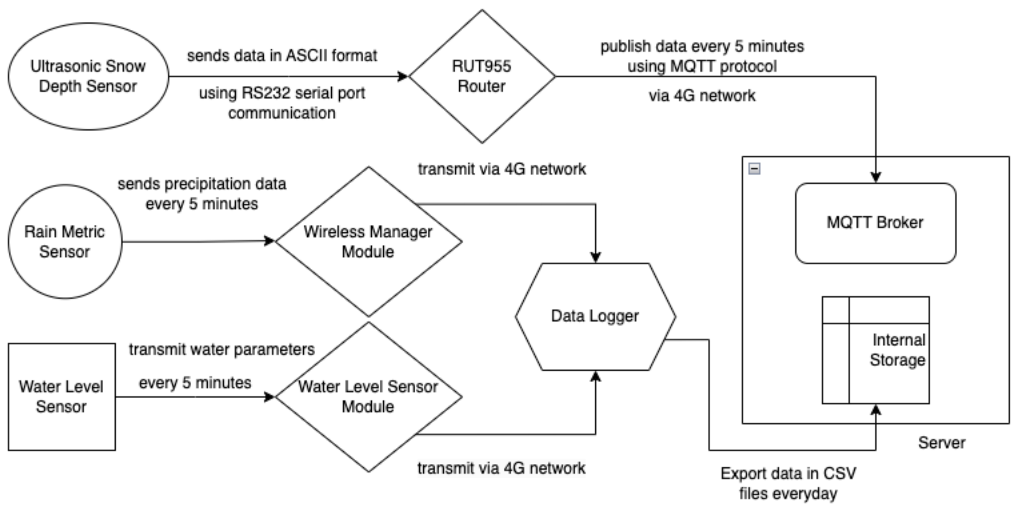


Figure 13: Data flow from IoT sensors to server.

As shown in Figure 14, the collected data is processed and stored in the persistent storage in the second part of the data flow. The system leverages microservice architecture, employing two distinct microservices: the Snow Data Collector Service and the Datalogger Data Collector Service. The Snow Data Collector Service acts as an MQTT client, subscribing to the ”snowsensor” topic in the MQTT broker. Upon receiving data from the broker, the service performs pre-processing, which involves conducting quality checks on the raw sensor data. This step is essential to ensure data consistency, as certain inconsistencies and garbage values have been observed in the dataset from Hydro Météo [23] and CMM [5]. Once the pre-processing is complete, the data is stored in the InfluxDB database, which serves as the persistent storage for the system.

Concurrently, the Datalogger Data Collector Service executes a scheduled cron job daily to process the data received from the data logger station. The service checks for new data files provided by the data logger station and processes them, applying quality checks to ensure data integrity. Once the data passes the quality checks, it is stored in persistent storage.

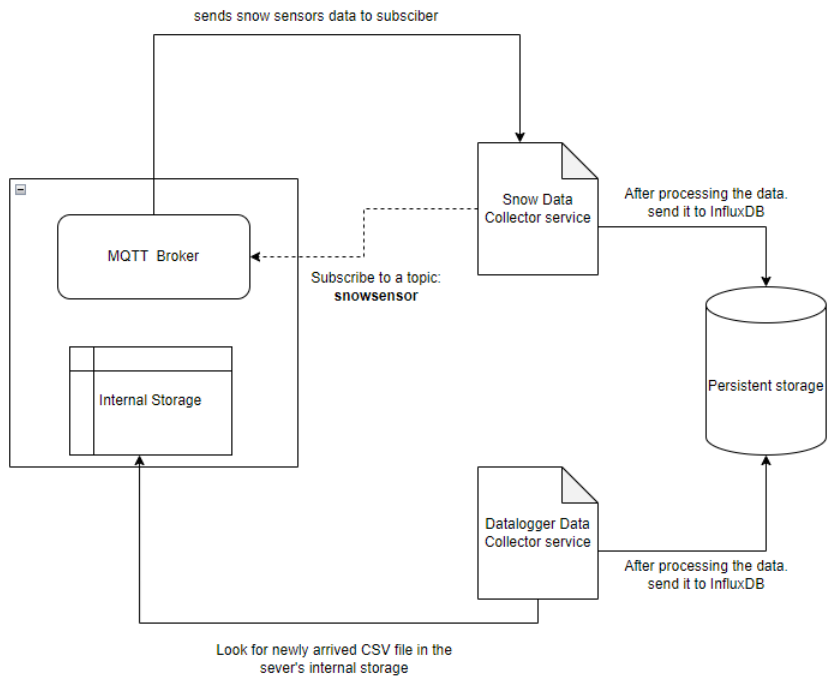


Figure 14: Data flow from Server to Persistent Storage(InfluxDB)

3.5 Technology Stack

The technology stack in the HYDROSIGHT system forms a crucial foundation for efficient and effective operation. Careful consideration was given to selecting the right mix of technologies, tools, and frameworks to ensure seamless data communication, real-time monitoring, and scalable data storage.

The MQTT (Message Queuing Telemetry Transport) protocol is at the heart of the system, which serves as the backbone for data transmission between the snow sensors and the central server. The lightweight nature of MQTT makes it ideal for low-power IoT devices, while its ability to support real-time data transmission ensures timely updates from the sensors. This protocol choice enables the system to capture and relay critical flood data in real-time, enhancing its responsiveness.

InfluxDB, a robust time-series database, was integrated into the HYDROSIGHT to efficiently handle data collection and storage. InfluxDB's high-performance capabilities and scalability allow it to handle large volumes of time-stamped sensor readings effectively. This ensures that historical flood data is stored and easily accessible for analysis and prediction, contributing to the system's accuracy and reliability.

Docker containers are utilized to encapsulate individual components to support the system's modularity and scalability. This containerization approach allows for easy deployment, management, and independent scaling of different services, enabling the system to adapt to varying demands and future expansion.

The HYDROSIGHT incorporates Nginx as a reverse proxy server to reinforce security and data privacy. Nginx acts as a protective intermediary, safeguarding data transmission and restricting unauthorized access to sensitive information. This aspect of the technology stack ensures that the system maintains high security and data integrity.

Python, renowned for its versatility and extensive libraries, is the primary programming language utilized in the HYDROSIGHT. Its robust ecosystem of libraries enables seamless integration with MQTT, data collection and data processing,

The thoughtfully selected technology stack collectively empowers the HYDROSIGHT system to fulfill its objectives effectively. By leveraging the strengths of each technology, the system becomes a powerful tool for real-time flood monitoring and data analysis. It can be used for informed decision-making for disaster management and flood resilience.

3.6 Scalability and Flexibility

The HYDROSIGHT system is designed with scalability and flexibility at its core, enabling it to adapt and grow in response to evolving needs and increasing data volumes. One critical architectural decision that facilitates scalability is the utilization of Docker [10] containers. Docker containers provide a lightweight and portable environment for deploying and managing the various components of the system. By encapsulating features like data storage, monitoring platform, and data collection within a container, the HYDROSIGHT system becomes highly modular, allowing for easy scaling of individual components based on demand.

Docker containers also promote flexibility by enabling rapid deployment and version control. New components or updates to existing ones can be seamlessly integrated into the system without disrupting its overall functionality. This adaptability is essential for accommodating future enhancements or deploying systems to different locations to cater to diverse monitoring requirements.

Furthermore, the architecture is built on microservices principles, allowing each functional module to operate independently. This decoupled nature enhances flexibility, enabling the addition or removal of specific services without impacting the rest of the system. As a result, the IoTPlatform can readily adapt to changing requirements and expand its capabilities to incorporate advanced features.

3.7 Evaluation and Testing

The system architecture underwent evaluations and performance testing at the Terrebonne, Lac-Superieur, and Ericsson facility deployment sites.

At Terrebonne and Lac-Superieur, the IoT environment was deployed with the Snow Sensor, Water Level Sensor, and Rain Metric Sensor. Real time data were collected to evaluate the system's ability to transmit, process, and visualize real-time sensor data for flood monitoring. The HYDROSIGHT system demonstrated efficient data reception, storage, and visualization, providing valuable insights into snow depth, water levels, and rainfall patterns.

Scalability and robustness were assessed through stress testing, where the system successfully handled varying sensor data loads without compromising performance or data integrity.

(a) (b) (c)

Figure 15: IoT setup at locations a. Ericsson Facility b. Lac-Superieur c. Terrebonne

3.8 Challenges

Throughout the project timeline, various challenges were encountered that required practical solutions for successfully implementing the HYDROSIGHT system. The process of selecting suitable IoT sensors posed complexities due to the need to align with project goals while considering factors like accuracy and compatibility. Creating an efficient data transmission process encountered the challenge of various IoT sensors having different modes of transmission. Striking a balance between real-time updates and data transmission limitations required careful protocol selection and data optimization techniques. The integration of microservices introduced challenges related to data source diversity and log management. Coordinating and standardizing logs from various microservices for comprehensive monitoring was achieved through centralized logging mechanisms and standardized log formats. Lastly, the deployment and testing phase revealed challenges in replicating real-world conditions and unforeseen system behavior. Rigorous field testing and simulation techniques were used to recreate diverse operational scenarios, enabling the identification and resolution of issues before full deployment.

Chapter 4

Online Machine Learning for Flood Prediction

Machine learning has been widely used in the modern world with abundant data and applications in artificial intelligence and data analytics. Conventional machine learning paradigms have often worked through batch learning or offline learning capabilities where a model is trained by some learning algorithm from the entire training dataset at once (over multiple epochs) and then deployed for inference without (or seldom) performing any update afterward. In contrast, online learning is an area of research under machine learning where the model learns incrementally from data sequentially.

Floods significantly threaten human life, infrastructure, and the environment. Flood prediction is critical to mitigating floods' impact, minimizing human casualties, preventing infrastructure loss, and preserving ecosystem imbalance. Flood prediction techniques used by various forecasting agencies involve usage of physically based deterministic hydrological models like HYDROTEL [14] at the *Direction de l'Expertise Hydrique et de l'Atmosphere (in English; Center for Water and Atmosphere Expertise)*(DEHA) of Quebec (Canada); Distributed Hydrological Soil Vegetation Model [56] at the Advances Hydrologic Prediction Services, associated with the National Oceanographic and Atmospheric Administration in the USA. *Pagano et al.* [40] highlights that these models utilize high computational resources and have difficulty accurately modeling fine spatial resolutions. This limitation was further resolved by data-driven flood prediction models, which used machine learning techniques [11, 29]. Implementing batch training in such flood forecasting models requires substantial datasets comprising historical records of multiple hydrological variables. However, for remote Internet of Things (IoT)based flood forecasting systems, storage capacity constraints and connectivity uncertainty pose significant challenges in accessing such extensive datasets. Consequently, online machine learning methodologies are deemed more suitable for remote IoT flood prediction, enabling real-time processing of newly acquired data without depending on local storage.

Numerous studies [1, 4, 35, 53] from different geographical locations across the globe have documented the fact that the frequency, severity, and duration of flood disasters are experiencing an increasing trend due to climate change, land use dynamics, and increasing human activities. The costs related to flooding have quadrupled in the past 40 years [26], and with the temperature rise by 2-6°C expected in Canada, the hydrological cycle is anticipated to be impacted [11]. With the increasing frequency and intensity of extreme weather events in recent years, the importance of flood prediction models to evolve with these changes becomes extremely important. Traditional batch learning methods become expensive when re-training the model on new data and may not adapt well to changing patterns or evolving data distributions. However, online machine learning overcomes the drawbacks of batch learning since the model is updated regularly with new incoming data. In the case of flood prediction, when data from IoT weather sensors are continuously streamed, an online learning model can adapt to the changing weather patterns and provide up-to-date predictions without the need for costly re-training processes. This enables more accurate and timely predictions in dynamic environments where the data distribution evolves.

This chapter proposes an online learning approach for flood forecasting using IoT to forecast water-level. We conduct preliminary tests on the performance of this approach for the next timestep and next 1 hour predictions. We also compare the results to the traditional batch learning paradigm for next-timestep prediction and observe that online machine learning improves flood forecasting capabilities with continual training.

4.1 Online Machine Learning

Online learning represents a distinct paradigm in machine learning, differing from the traditional batch learning approach. In the batch learning method, the learning process relies on having the entire training data available beforehand, and the training occurs offline due to its resource-intensive nature [20]. However, this approach has inherent drawbacks, such as inefficiency in terms of time and space costs and limited scalability for large-scale applications, as the model needs to be retrained entirely when new data arrives.

In contrast, online learning is a dynamic and sequential method that handles data in a streaming fashion. As new data instances arrive individually, the learner continuously updates its predictive model to adapt to the latest information. This real-time adaptation makes online learning highly efficient and scalable, particularly for big data applications with high data velocity. By instantly updating the model with each new data instance, online learning overcomes the limitations of batch learning, providing a more agile and adaptive approach to machine learning tasks. This makes online learning a valuable choice for real-world data analytics applications where data arrive continuously and in large volumes.

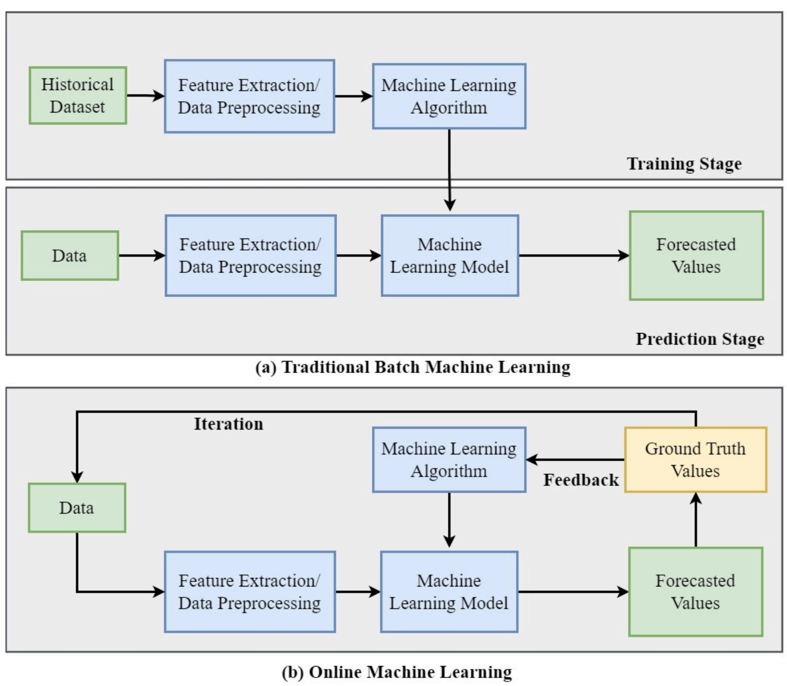


Figure 16: Batch Machine Learning vs Online Machine Learning

4.1.1 Data Collection

The data collection process for online machine learning involves retrieving real-time sensor data from the Water Level Sensor (MX2001-01-SS-S) of the HYDROSIGHT system deployed at various physical locations. The Water Level Sensor communicates through the Water Level Sensor Module (RXMOD-W1) via a 4G network, enabling seamless data transmission. Water-level data is transmitted in regular intervals of 5 minutes via the 4G network to the Data Logger Station (RX3004) which plays a crucial role in this process. The data logger station accumulates the water level data daily and exports it to the backend server for further processing and analysis. A scheduled cron job is executed on the server to collect and pre-process the water level data. The pre-processing process involves quality checks like checking null and negative values, verifying the correctness of data based on parameters like sensors installation height, and defining minimum and maximum values for each measurement. The pre-processed data is then stored in a time-series database to enable efficient retrieval and management of the sensor readings. The time-series database stores the water level data separated by 5-minute intervals, capturing detailed and frequent measurements. This data is utilized for training and testing the online machine-learning model. The model continuously updates and adapts its predictions based on the real-time data from the Water Level Sensor, allowing for accurate and timely forecasting of water level heights.

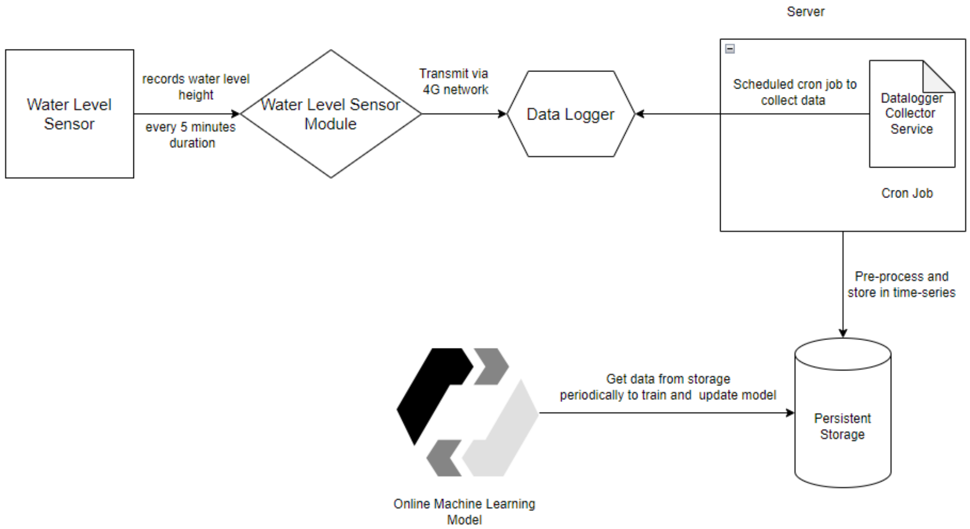


Figure 17: Data Collection for Online Machine Learning Model

4.1.2 Proposed Online Machine Learning Model

In this section, we mention the Online machine learning model used to achieve accurate water-level forecasting using time series data collected from the HYDROSIGHT system. The data is gathered at five-minute intervals and stored in a dedicated database on a server. The methodology involves training a machine learning model to forecast the water level for the next time step based on the information from the last ten time steps in the database.

Before inputting the data into the model, several pre-processing techniques are applied to ensure data quality. These techniques include verifying values and correcting any missing entries. The data is also scaled using MinMaxScaler [42] to improve the model's performance.

Algorithm 1: Online Machine Learning Algorithm

Initialize: $w_1 = 0$; // Initialization of model weight w_1

for `t = 1, 2, .., T` **do**

  Learner receives incoming sensor data x_t;

  Learner predicts next value: $\hat{y}_t = f(x_t; w_t)$;

  Output: $\hat{y}_t$

  True value y_t received from sensor ;

  Learner calculates the loss: $l = MSE(y_t, \hat{y}_t)$;

  if $loss(\varepsilon) > 0$ **then**

    Update the learner;

    $w_{t+1} = w_t + \delta(w_t, (x_t, y_t))$;

Once the data is processed and scaled, the model utilizes its prior training to predict the water level for the next step. At the beginning of the training process, the model's weights are random, but as it progresses, it continuously learns and refines its predictions through online learning. The forecast value, denoted as $\hat{y}_t$, is then compared to the actual water-level value, y_t, to calculate the prediction error. However, due to the 5-minute sensor reading delay, the error calculations occur after this duration.

Following error calculation, the online training of the model begins. The model is trained on a single training sample of the last ten time steps, allowing it to continuously improve and enhance its forecasting capabilities for the next step. After training, the model is stored in a model registry, which efficiently manages and versions the models.

For subsequent water-level forecasting tasks, instead of deploying the model directly, the latest version of the model is retrieved from the model registry. This ensures that the most up-to-date and refined model is always used for making predictions. Due to resource constraints, only one model is stored in this system. However, if storage for multiple models is available, the model registry allows for tracing back to a model with better accuracy in case of very high error in the current model.

This cyclic online training, evaluation, and deployment process enables the proposed approach to adapt continuously to dynamically changing river water-level data, ensuring accurate forecasts over time. As a result, this approach provides an efficient and practical solution for real-time water-level forecasting, which is crucial for various water resource management applications.

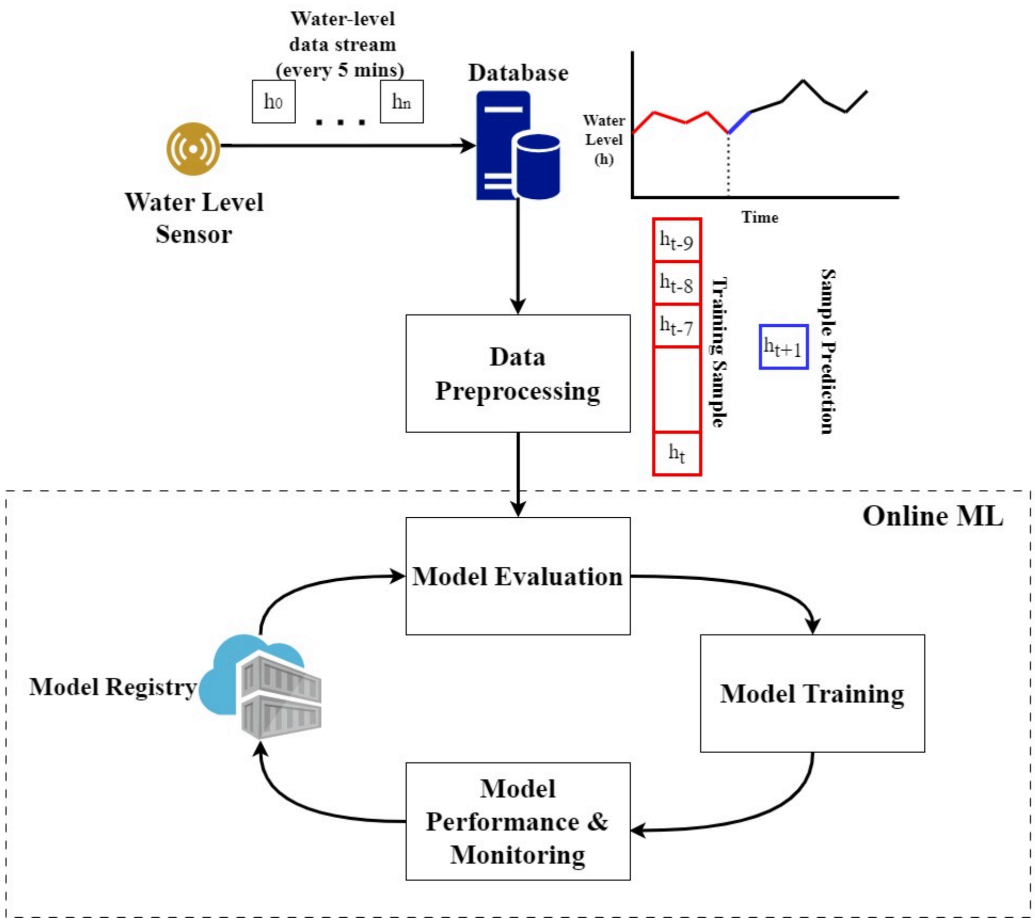


Figure 18: Online Machine Learning Architecture for Water Level Forecasting

4.2 Results

Our approach was evaluated using time series data from Lac-Supérieur, one of the test sites where the HYDROSIGHT system was deployed for testing. Unfortunately, the system deployed at Terrebonne and Ericsson facility was not fully functional during testing and was not utilized for this experiment. To assess the performance of the online machine learning model, we divided the experiment into two phases, Phase-1 and Phase-2.

In Phase-1, water level measurements from 5th June 2023 to 25th June 2023 were used as training datasets. For Phase-2, the model was further trained with data up to 30th June 2023 to incorporate more varied water level measurements. The online machine learning model's performance was evaluated for two forecasting duration: 5 minutes and 1 hour, using a testing dataset spanning from 30th June 2023 to 4th July 2023.

Figure 19 illustrates the rationale behind the specific dataset distribution. It was crucial to account for the rapidly changing environment and the continuous evolution of data distributions in our evaluation. The time gap of more than five days between the evaluation data and the testing phase significantly impacted the Mean Squared Error (MSE) results. This highlighted the necessity of regularly updating the model to accommodate the transient changes in river water levels.

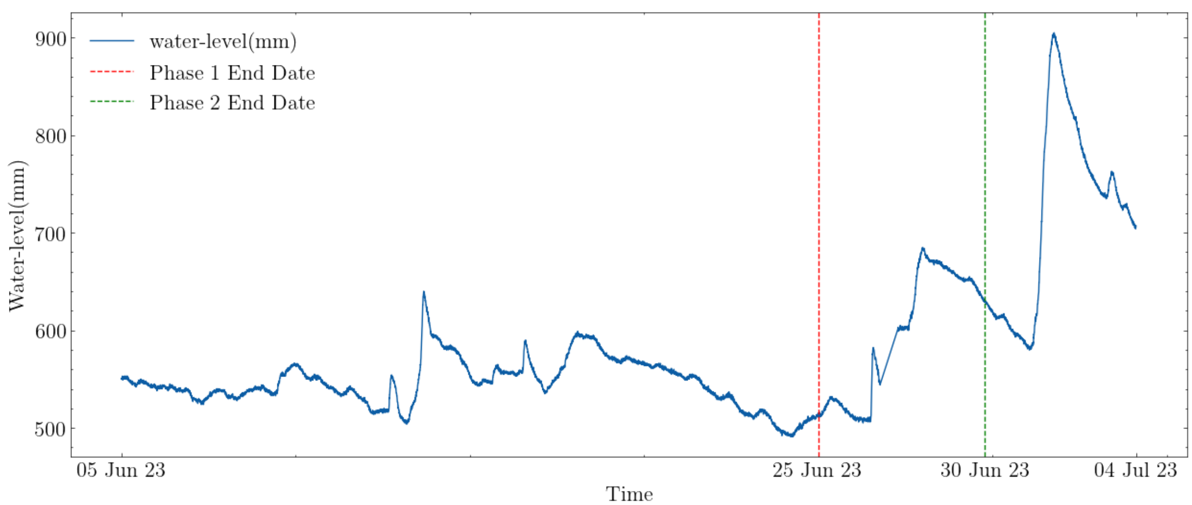


Figure 19: Visualization of water-level collected from Lac Superior and corresponding training phase dates

The results presented in Table 1 show a notable difference between the MSE of the testing phase. However, with Phase-2, where the model was continuously updated with incoming data until 30th June 2023, we observed a significant enhancement in the test error. The MSE decreased from 0.047 to 0.012, representing an impressive 74% decrease compared to the scenario where the model was not continuously updated. The same trend was observed in the results for forecasting water levels with a 1-hour duration. The continuous updating of the model resulted in improved accuracy and performance.

Forecasting Duration	Training	Testing Mean Squared Error(MSE)
Next timestep (5 minutes)	Phase-1	0.04796
	Phase-2	0.01235
1 hour	Phase-1	0.05049
	Phase-2	0.01331

Table 1: Results of water level forecasting using online learning approach

To gain initial insights into the applicability of Online ML for water level forecasting in contrast to Traditional Batch learning, we conducted an experiment focusing on forecasting water level heights for a 5-minute duration. We compared the results obtained from the Online ML approach with those derived from the Traditional Batch learning approach.

As shown in Table 2, the experiment's outcomes clearly demonstrate the benefit of the Online ML paradigm, particularly in Phase-2. The Online ML model exhibited more effective forecasting capabilities than the Traditional Batch learning model. This improvement can be attributed to the Online ML model's ability to adapt and continuously update with incoming data, thus making it better suited to handle the rapidly changing environment and evolving data distributions in water level measurements.

These preliminary findings indicate the promising potential of Online ML in the context of water level forecasting and underscore the advantages it holds over the more conventional Traditional Batch learning approach.

Machine Learning Approach	Training	Testing Mean Squared Error(MSE)
Batch ML	Phase-1	0.043801
	Phase-2	0.028732
Online ML	Phase-1	0.047962
	Phase-2	0.012355

Table 2: Comparison of Batch ML and Online ML approaches for water level forecasting

Chapter 5

Conclusions and Future Work

5.1 Conclusion

The development and implementation of HYDROSIGHT have provided valuable insights and lessons for flood monitoring and disaster management. One crucial aspect is the careful selection of sensors and communication protocols tailored to specific environmental conditions, considering factors like accuracy, operating range, scalability, and cost-effectiveness. Adopting MQTT communication protocol and InfluxDB time series database has proven highly efficient for real-time data transmission and storage, optimizing data flow and processing.

The rigorous testing of HYDROSIGHT in diverse environmental conditions highlighted the importance of system robustness and scalability, achieved through containerization techniques. The successful integration of Grafana and Loki for real-time data visualization and logging emphasized the significance of user-friendly interfaces and data presentation, enabling stakeholders to access critical information promptly during flood scenarios.

Furthermore, an online learning approach for flood forecasting, leveraging IoT and machine learning, showcased significant improvements compared to traditional batch learning methods. The adaptability of the online learning model allows continuous updates based on dynamically changing environmental conditions.

In conclusion, this research contributes to the advancement of flood forecasting methodologies, offering an innovative online learning approach that integrates IoT and machine learning. The promising results pave the way for future research and development in the field, with the ultimate goal of mitigating the impacts of floods and enhancing resilience in vulnerable communities.

5.2 Future Work

The successful implementation of the HYDROSIGHT flood monitoring system paves the way for continuous improvement and expansion of its capabilities.

For the IoT platform part, one aspect that can be enhanced is the incorporation of health checks for the sensors and their connectivity. By implementing health checks, the system can ensure the sensors' correct functioning and uninterrupted data transmission to the central monitoring platform, enabling real-time flood monitoring. Furthermore, integrating a battery usage monitoring feature for the sensors will offer valuable insights into power consumption. Optimizing sensor battery usage is crucial, particularly in remote or off-grid locations, to maintain continuous data collection and monitoring.

Exploring cloud services for system architecture can significantly enhance scalability and performance. Leveraging cloud services can optimize server loads, handle increased user requests, and boost overall system reliability and responsiveness. To enhance user engagement and system responsiveness, a notification system is essential. Timely notifications to users about critical events detected by the sensors will enable prompt disaster response and mitigation. Moreover, the potential of the system can be further unlocked by integrating advanced machine learning models for water level forecasting. Utilizing historical and real-time data, these models can significantly improve accuracy and precision in flood predictions, empowering effective disaster management strategies. To support disaster management efforts, forecast water level data can be seamlessly integrated into the monitoring platform, offering stakeholders a comprehensive view of flood conditions and guiding informed decision-making.

In the context of online learning for water level forecasting, there is room for future research and improvements. Fine-tuning the online learning approach by controlling learning rates and optimizing hyperparameters can enhance its performance. Dynamic model sizing based on data availability and input characteristics can lead to more efficient and flexible models, optimizing computational resources while maintaining forecasting accuracy. Incorporating additional meteorological parameters as model inputs, such as rain, snow, pressure, etc., can further enhance the predictive capabilities of the flood forecasting system, providing a deeper consideration of the factors affecting water levels.

The HYDROSIGHT flood monitoring system and the online learning approach for water level forecasting presented in this thesis lay the foundation for continuous development and advancements in flood monitoring technology. By addressing the outlined future works, the system can achieve even greater efficiency, accuracy, and real-time responsiveness, contributing to more effective flood management and resilience strategies.

Appendix A

Additional Description of IoT sensors

A.1 Ultrasonic Snow Depth Sensor

The sensor operates at a frequency of 49.4 kilohertz. The sensor performs two consecutive measurements during each cycle and compares them to enhance measurement accuracy. The last measurement is considered valid and utilized if the difference between the measurements is within one centimeter. However, in cases where the difference exceeds one centimeter, the sensor discards the older reading and initiates a new measurement and comparison. This retry algorithm can be repeated up to ten times if necessary. If the sensor cannot obtain a valid measurement or fails to receive an echo, it returns a zero value, ensuring data integrity and reliability.

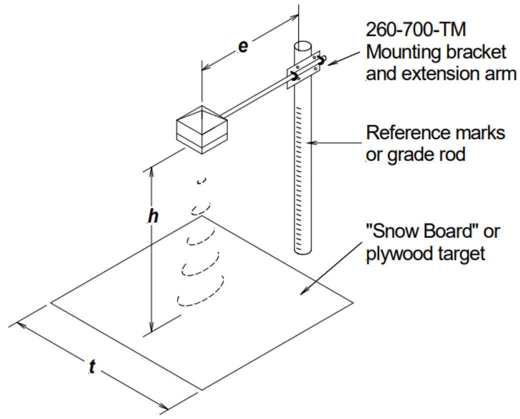


Figure A.1: Source: Snow Sensor Installation [39]

The sensor completes the measurement process within 1.5 to 3.5 seconds, depending on the number of retries required to achieve accurate measurements. To account for the impact of air temperature on sound speed, the sensor incorporates a built-in temperature sensor. By measuring the air temperature, the sensor calculates the speed of sound and applies a correction factor to the time-of-flight data, resulting in accurate distance calculations. This temperature compensation feature enhances the precision of the snow depth, particularly in varying environmental conditions.

A.2 Water Level Sensor and Water Level Sensor Module

This sensor can measure water levels up to approximately 0 to 9 m (0 to 30 ft) of water depth at sea level and is durable and made with ceramic which can withstand freezing environments, making it suitable for use in harsh environments.

The MX2001-01-SS-S water level sensor utilizes hydrostatic pressure to calculate water level accurately. As the sensor is submerged deeper into the water, the pressure exerted by the water increases proportionally. This is due to the weight of the water column above the sensor, which adds to the hydrostatic pressure.

The RXMOD-W1 Water Level Sensor Module fulfills critical selection criteria, making it an ideal choice for an IoT platform. It prioritizes user-friendliness with its plug-and-play functionality, allowing for easy installation and setup. Moreover, its low power consumption ensures energy efficiency and extends battery life, making it well-suited for remote or battery-operated arrangements.

A.3 Rain Metric Sensor and Wireless Manager Module

The RXW-RGF-900 Rain Metric Sensor uses a mechanism to measure rainwater accurately. When rain falls, it enters the sensor's collector cone and passes through a debris-filtering screen, ensuring that any unwanted particles or debris. The rainwater then collects in the chamber of the tipping mechanism.

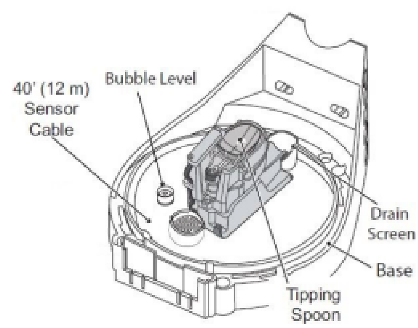


Figure A.2: RXW-RGF-900 Rain Metric Sensor Internal Component [48]

A special tipping spoon plays a crucial role in the measurement process inside the sensor. The spoon is designed to tip over when it accumulates a specific amount of water, equal to the increment in which the sensor measures rainfall. For instance, if the sensor is set to measure rainfall in increments of 0.2 millimeters or 0.01 inches, the spoon will tip when it has collected this precise amount of rainwater.

As the spoon tips over, it causes a switch to close temporarily, indicating that a tipping event has occurred. After tipping, the spoon springs back to its original position, ready to collect more rainwater for the next measurement.

The rainwater accumulates in the chamber and drains out through screened drains at the collector's base. This draining process ensures the sensor is ready to measure the next rainfall event accurately.

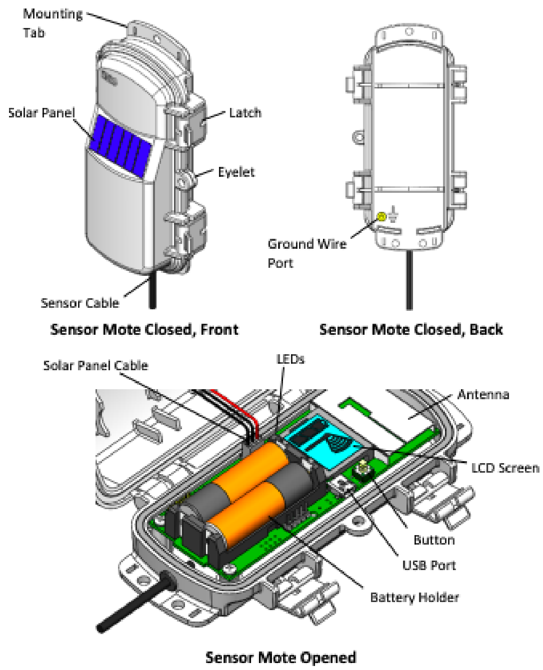


Figure A.3: RXMOD-RXW-900 Wireless Manager Module Components [48]

The RXMOD-RXW-900 Wireless Manager Module has several essential components that enhance its operational capabilities. It features mounting tabs for secure installation, and a solar panel positioned towards the sun charges the mote batteries efficiently. The sensor cable enables seamless data transmission between the mote and the sensor, while an eyelet provides additional security with the option to attach a padlock.

The module's latches facilitate easy access to the mote door, simplifying battery replacement and maintenance. A ground wire port ensures added safety during deployment. The module includes a built-in antenna and LED indicators for radio communications that display the mote's network status. The LCD screen provides comprehensive information about the module's current status.

The RXMOD-RXW-900 module's solar panel cable ensures an effective power supply, and the battery holder offers a convenient location for installing batteries. The USB port allows connection to a computer for firmware updates, enhancing the module's versatility. Additionally, a button on the module enables LCD screen illumination or the initiation of a search for an RX Wireless Sensor

Network to join.

References

[1] L. Alfieri, B. Bisselink, F. Dottori, G. Naumann, A. de Roo, P. Salamon, K. Wyser, and L. Feyen. Global projections of river flood risk in a warmer world. *Earth's Future*, 5(2):171–182, 2017.

[2] M. Ancona, A. Dellacasa, G. Delzanno, A. Camera, and I. Rellini. An Internet of Things vision of the flood monitoring problem. In *The International Conference on Ambient Computing, Applications, Services and Technologies*, volume 3, pages 26–29, 2015.

[3] S. Bande and V. V. Shete. Smart flood disaster prediction system using IoT & neural networks. In *International Conference On Smart Technologies For Smart Nation (SmartTechCon)*, pages 189–194. IEEE, 2017.

[4] G. Blöschl, J. Hall, A. Viglione, R. A. Perdigão, J. Parajka, B. Merz, D. Lun, B. Arheimer, G. T. Aronica, A. Bilibashi, et al. Changing climate both increases and decreases european river floods. *Nature*, 573(7772):108–111, 2019.

[5] Communauté métropolitaine de Montréal(CMM). https://cmm.qc.ca/. Accessed: July 30, 2023.

[6] Grues grand montreal. https://www.cruesgrandmontreal.ca/, 2022. Accessed: August 21, 2023.

[7] Onset RX3004 4G Data Logger Station. https://www.onsetcomp.com/products/data-loggers/rx3000. Accessed: July 30, 2023.

[8] N. De Caro, W. Colitti, K. Steenhaut, G. Mangino, and G. Reali. Comparison of two lightweight protocols for smartphone-based sensing. In *IEEE Symposium on Communications and Vehicular Technology in the Benelux (SCVT)*, pages 1–6, 2013.

[9] T. G. Dietterich. Machine learning for sequential data: A review. In *Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshops SSPR and SPR Windsor, Ontario, Canada, Proceedings*, pages 15–30. Springer, 2002.

[10] I. Docker. Docker. *lınea].Disponible en: https://www. docker. com/what-docker*, 2020.

[11] I. Ebtehaj and H. Bonakdari. A reliable hybrid outlier robust non-tuned rapid machine learning model for multi-step ahead flood forecasting in quebec, canada. *Journal of Hydrology*, 614:128592, 2022.

[12] S. Elhadi, A. Marzak, N. Sael, and S. Merzouk. Comparative study of IoT protocols. *Smart Application and Data Analysis for Smart Cities (SADASC)*, 2018.

[13] P. T. Eugster, P. A. Felber, R. Guerraoui, and A.-M. Kermarrec. The many faces of publish/subscribe. *ACM computing surveys (CSUR)*, 35(2):114–131, 2003.

[14] J.-P. Fortin, R. Turcotte, S. Massicotte, R. Moussa, J. Fitzback, and J.-P. Villeneuve. Distributed watershed model compatible with remote sensing and GIS data. I: Description of model. *Journal of Hydrologic Engineering*, 6(2):91–99, 2001.

[15] K. Fysarakis, I. Askoxylakis, O. Soultatos, I. Papaefstathiou, C. Manifavas, and V. Katos. Which IoT Protocol? Comparing Standardized Approaches over a Common M2M Application. In *IEEE Global Communications Conference (GLOBECOM)*, pages 1–7, 2016.

[16] C. Gemirter, C. Senturca, and S. Baydere. A comparative evaluation of AMQP, MQTT and HTTP protocols using real-time public smart city data. In *International Conference on Computer Science and Engineering (UBMK)*, pages 542–547, 2021.

[17] Geosapiens. `https://www.hydrometeo.net/index.php/services/instrumentation`, 2022. Accessed: August 21, 2023.

[18] M. Golnaraghi, S. Surminski, S. Mehryar, C. Kousky, and V. Roezer. Building flood resilience in a changing climate. *The Geneva Association*, 2020.

[19] Grafana. https://grafana.com/. Accessed: July 26, 2023.

[20] S. C. Hoi, D. Sahoo, J. Lu, and P. Zhao. Online learning: A comprehensive survey. *Neurocomputing*, 459:249–289, 2021.

[21] U. Hunkeler, H. L. Truong, and A. Stanford-Clark. MQTT-S — A publish/subscribe protocol for Wireless Sensor Networks. In *International Conference on Communication Systems Software and Middleware and Workshops (COMSWARE)*, pages 791–798, 2008.

[22] H. C. Hwang, J. Park, and J. G. Shon. Design and implementation of a reliable message transmission system based on MQTT protocol in IoT. *Wireless Personal Communications*, 91:1765–1777, 2016.

[23] Hydro météo. https://www.hydrometeo.net/. Accessed: July 30, 2023.

[24] O. D. Incel and S. O. Bursa. On-device deep learning for mobile and wearable sensing applications: A review. *IEEE Sensors Journal*, 2023.

[25] InfluxDB. https://www.influxdata.com/. Accessed: July 26, 2023.

[26] International Bureau of Canada. https://www.ibc.ca/news-insights/news/severe-weather-in-2022-caused-3-1-billion-in-insured-damage-making-it-the-3rd-worst 2022. Accessed: July 26, 2023.

[27] F. A. Kraemer, H. A. Asad, K. Bach, and B.-C. Renner. Online machine learning for 1-day-ahead prediction of indoor photovoltaic energy. *IEEE Access*, 11:38417–38425, 2023.

[28] S. Kraijak and P. Tuwanut. A survey on internet of things architecture, protocols, possible applications, security, privacy, real-world implementation and future trends. In *IEEE International Conference on Communication Technology (ICCT)*, pages 26–31. IEEE, 2015.

[29] X.-H. Le, H. V. Ho, G. Lee, and S. Jung. Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. *Water*, 11, 2019.

[30] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller. Efficient backprop. In *Neural networks: Tricks of the trade*, pages 9–50. Springer, 2002.

[31] S.-W. Lee, C.-Y. Lee, D.-H. Kwak, J. Kim, J. Kim, and B.-T. Zhang. Dual-memory deep learning architectures for lifelong learning of everyday human behaviors. In *IJCAI*, pages 1669–1675, 2016.

[32] L. Melgar-García, D. Gutiérrez-Avilés, C. Rubio-Escudero, and A. Troncoso. A novel distributed forecasting method based on information fusion and incremental learning for streaming time series. *Information Fusion*, 95:163–173, 2023.

[33] B. Mishra and A. Kertesz. The Use of MQTT in M2M and IoT Systems: A Survey. *IEEE Access*, 8:201071–201086, 2020.

[34] Onset MX2001-01-SS-S Water level sensor. `https://www.onsetcomp.com/products/sensors/mx2001-s`. Accessed: July 26, 2023.

[35] N. Najibi and N. Devineni. Recent trends in the frequency and duration of global floods. *Earth System Dynamics*, 9(2):757–783, 2018.

[36] D. Nallaperuma, R. Nawaratne, T. Bandaragoda, A. Adikari, S. Nguyen, T. Kempitiya, D. De Silva, D. Alahakoon, and D. Pothuhera. Online incremental machine learning platform for big data-driven smart traffic management. *IEEE Transactions on Intelligent Transportation Systems*, 20(12):4679–4690, 2019.

[37] Nginx. `https://www.nginx.com/`. Accessed: July 26, 2023.

[38] C. Nicolas, B. Naila, and R.-C. Amar. Energy efficient Firmware Over The Air Update for TinyML models in LoRaWAN agricultural networks. In *International Telecommunication Networks and Applications Conference (ITNAC)*, pages 21–27. IEEE, 2022.

[39] Novalynx 260-700-D Snow Depth Sensor. https://novalynx.com/manuals/260-700-d-manual.pdf. Accessed: July 23, 2023.

[40] T. C. Pagano, A. W. Wood, M.-H. Ramos, H. L. Cloke, F. Pappenberger, M. P. Clark, M. Cranston, D. Kavetski, T. Mathevet, S. Sorooshian, et al. Challenges of operational river forecasting. *Journal of Hydrometeorology*, 15(4):1692–1707, 2014.

[41] M. Pohl, J. Kubela, S. Bosse, and K. Turowski. Performance evaluation of application layer protocols for the Internet-of-Things. In *International Conference on Enterprise Systems (ES)*, pages 180–187, 2018.

[42] V. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma. Study the influence of normalization/transformation process on the accuracy of supervised classification. In *Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pages 729–735. IEEE, 2020.

[43] D. S. Rani, G. N. Jayalakshmi, and V. P. Baligar. Low Cost IoT based Flood Monitoring System Using Machine Learning and Neural Networks: Flood Alerting and Rainfall Prediction. In *International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, pages 261–267, 2020.

[44] J. Read, A. Bifet, B. Pfahringer, and G. Holmes. Batch-incremental versus instance-incremental learning in dynamic and evolving data. In *Advances in Intelligent Data Analysis XI: International Symposium, IDA, Helsinki, Finland. Proceedings 11*, pages 313–323. Springer, 2012.

[45] H. Ren, D. Anicic, and T. A. Runkler. TinyOL: TinyML with Online-Learning on Microcontrollers. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2021.

[46] Rut955 industrial router. https://teltonika-networks.com/products/routers/rut955/. Accessed: July 30, 2023.

[47] Onset RXMOD-W1 Water Level Sensor Module. https://www.onsetcomp.com/products/accessories/rxmod-w1. Accessed: July 26, 2023.

[48] Onset RXW-RGF-900 Rain Metric Sensor. https://www.onsetcomp.com/products/sensors/rxw-rgf-xxx. Accessed: July 26, 2023.

[49] D. Sahoo, Q. Pham, J. Lu, and S. C. H. Hoi. Online Deep Learning: Learning Deep Neural Networks on the Fly. In *Proceedings of the International Joint Conference on Artificial Intelligence*, IJCAI, page 2660–2666. AAAI Press, 2018.

[50] D. Satria, S. Yana, R. Munadi, and S. Syahreza. Design of information monitoring system flood based internet of things (IoT). In *Proceedings of MICoMS*, pages 337–342. Emerald Publishing Limited, 2018.

[51] T. Singh, R. Kalra, S. Mishra, Satakshi, and M. Kumar. An efficient real-time stock prediction exploiting incremental learning and deep learning. *Evolving Systems*, pages 1–19, 2022.

[52] D. Soni and A. Makwana. A survey on MQTT: a protocol of internet of things (IoT). In *International conference on telecommunication, power analysis and computing techniques (ICTPACT)*, volume 20, pages 173–177, 2017.

[53] B. Tellman, J. A. Sullivan, C. Kuhn, A. J. Kettner, C. S. Doyle, G. R. Brakenridge, T. A. Erickson, and D. A. Slayback. Satellite imaging reveals increased proportion of population exposed to floods. *Nature*, 596(7870):80–86, 2021.

[54] A. H. Thiessen. *Weather glossary*. US Department of Commerce, Weather Bureau, 1946.

[55] P. H. Whitfield. Floods in future climates: a review. *Journal of Flood Risk Management*, 5(4):336–365, 2012.

[56] M. S. Wigmosta, L. W. Vail, and D. P. Lettenmaier. A distributed hydrology-vegetation model for complex terrain. *Water resources research*, 30(6):1665–1679, 1994.

[57] R. J. Williams and D. Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural computation*, 1(2):270–280, 1989.

[58] H. C. Winsemius, J. C. Aerts, L. P. Van Beek, M. F. Bierkens, A. Bouwman, B. Jongman, J. C. Kwadijk, W. Ligtvoet, P. L. Lucas, D. P. Van Vuuren, et al. Global drivers of future river flood risk. *Nature Climate Change*, 6(4):381–385, 2016.

[59] Y. Y. Chen and T. Kunz. Performance evaluation of IoT protocols under a constrained wireless access network. In *International Conference on Selected Topics in Mobile & Wireless Networking (MoWNeT)*, pages 1–7, 2016.

[60] L. Yan, J. Feng, Y. Wu, and T. Hang. Data-driven fast real-time flood forecasting model for processing concept drift. In *Cloud Computing, Smart Grid and Innovative Frontiers in Telecommunications: EAI International Conference, CloudComp, and EAI International Conference, SmartGIFT , Beijing, China*, pages 363–374. Springer, 2020.

[61] L. Yu, S. K. Tan, and L. H. Chua. Online ensemble modeling for real time water level forecasts. *Water resources management*, 31:1105–1119, 2017.

[62] M. I. Zakaria, W. A. Jabbar, and N. Sulaiman. Development of a smart sensing unit for LoRaWAN-based IoT flood monitoring and warning system in catchment areas. *Internet of Things and Cyber-Physical Systems*, 3:249–261, 2023.

[63] G. Zhou, K. Sohn, and H. Lee. Online incremental feature learning with denoising autoencoders. In M. G. ENeil D. Lawrence, editor, *15th International Conference on Artificial Intelligence and Statistics (AISTATS)*, Proceedings of Machine Learning Research (PMLR), pages 1453–1461, La Palma, Canary Islands, 2012.