

Ministry of Higher Education and Scientific Research

University of Oran 1, Ahmed Ben Bella

Institute of Applied Sciences and Techniques

End-of-study project

By

DIAF Amel

In preparation for obtaining the diploma of

Professional Master's degree

Specialty: Instrumentation and Metrology

Theme:

An IOT based in wastewater for irrigation using Deep Learning

Defended on July, 2024 before the jury

composed of:

Presidente
Supervisor
Examiner

Asma Benchkor
Amine Dahane
Tahri Ghrissi

ISTA, University of Oran 1
ISTA, University of Oran 1
ISTA, University of Oran 1

Academic Year 2023 / 2024

Abstract

Utilizing wastewater for irrigation offers significant benefits due to its high nutrient content, reducing the reliance on chemical fertilizers. Additionally, wastewater irrigation aids in the conservation of water resources. Treated wastewater has emerged as a dependable source for farmers, enabling year-round crop cultivation, particularly in arid regions. This study proposes a novel approach that integrates artificial intelligence and edge computing capabilities to offer instantaneous monitoring and prediction of water quality, adhering to parameter thresholds set by the World Health Organization (WHO). This master thesis encompasses both the front-end and back-end aspects, comprising software and hardware architectures. The hardware components include actuators, sensors, and controllers connected remotely via wireless networks. Meanwhile, the software incorporates applications linked with artificial intelligence (AI) models for intelligent predictions. The dataset comprises various features of five parameters essential for determining water quality: conductivity, turbidity, oxygen levels, pH, and temperature, all of which are monitored by sensors included in the sensor module.

Kay words: Water quality, Internet of Things (IoT), Artificial intelligence, Deep learning, Edge computing, Monitoring.

Acknowledgment

First and foremost, we thank God, the Almighty, for granting us health, willpower, His guidance, and protection.

Secondly, we would like to express our profound gratitude to Mr. Dahane Amine for their support, assistance, and for the trust they placed in us. We are grateful for their critiques and suggestions, which have been of great value to us.

Our heartfelt thanks also go to all the members of the jury, Mr. Tahri Ghrissi and Mme.Benchkour Assma, for the great honor they have bestowed upon us by agreeing to evaluate this work.

Table of contents

Abstract	1
Acknowledgment	2
Lists of abbreviations	8
Lists of tables	11
Lists of figures.....	12
Thesis Organization.....	14
Chapter 1	15
Introduction:	3
1. IOT (Internet of Things):.....	3
1.1 Definition:.....	4
1.2 Importance of IOT:	4
1.3 Characteristics of the Internet of Things:	4
1.3.1 Connectivity:	4
1.3.2 Identity of Things:.....	5
1.3.3 Data:	5
1.3.4 Intelligence:	5
1.3.5 Network:.....	5
1.3.6 Scalability:.....	6
1.3.7 Architecture:.....	6
1.3.8 Security:	6
1.4 The IoT reference model:	7
1.4.1 Application Layer:.....	7
1.4.2 Service and Application Support Layer:	7
1.4.3 Network Layer:	8
1.4.4 Device Layer:	8
1.5 Advantages and disadvantages:	8
1.5.1 Advantages:	8
1.5.2 Disadvantages:	9
2. The application domains of IoT:	9

2.1 Healthcare:	9
2.2 Manufacturing:	9
2.3 Agriculture:	10
2.4 Transportation:	10
3. IoT and precision agriculture:	10
3.1 Applications of IOT in agriculture:	10
3.1.1 Irrigation Management System:	11
3.1.2 Water Quality Monitoring:	12
3.1.3 Precision Farming Using UAV:	12
3.1.4 Precision Livestock Farming:	12
3.1.5 Green House Automation:	12
3.1.6 Crop Disease and Pest Management:	12
3.2 Precision agriculture:	13
3.3 The major components of an IoT-based precision agriculture system:	13
3.3.1 The physical structure:	14
3.3.2 Data acquisition:	14
3.3.3 Data processing:	14
3.3.4 Data analysis:	15
3.4 The benefits of IoT in precision agriculture:	15
3.4.1 Data Collection:	15
3.4.2 Better Yield Management:	15
3.4.3 Cost Management Improvement:	15
3.4.4 Increased Efficiency:	15
3.4.5 Product Quality and Volume Enhancement:	15
Conclusion:	15
Chapter 2	3
Introduction:	17
1. Integration of AI in Wastewater Treatment:	17
2. Related work:	18
Cost Effective Soft Sensing for Wastewater Treatment Facilities	24
Forecasting of Wastewater Treatment Plant Key Features Using Deep Learning-Based Models: A Case Study	24
New approach to monitoring a wastewater irrigation system controlled by the artificial neural network (ANN)	25
Times Series Prediction used in Treating Municipal Wastewater for Plant Irrigation	25

Application of machine learning at wastewater treatment facilities: a review of the science, challenges and barriers by level of implementation	25
Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment	26
Deep learning in wastewater treatment: a critical review	26
Wastewater treatment plant monitoring via a deep learning approach	26
Application of support vector machine in online monitoring of wastewater treatment based on combined kernel functions	27
Prediction of wastewater treatment system based on deep learning	27
Conclusion:.....	27
Chapter 3	17
Introduction:	28
1 Software Tools:	28
1.1 Definition of the Python programming language in computer science:.....	28
1.2 Definition of anaconda:	29
1.3 Definition of jupyter:.....	29
1.4 Definition of Arduino:.....	30
2 Hardware Tools:	30
2.1 The Arduino card:	30
2.2 PH sensor:	31
Features:	31
2.3 Temperature sensor:	31
Features:	32
2.4 Electrical conductivity sensor:	32
Features:	33
2.5 Turbidity sensor:	33
Features:	33
2.6 Dissolved oxygen sensor:.....	34
Features:	34
3 Development tools.....	34
3.1 Machine Learning:	34
3.2 Types of machine learning models:	35
3.2.1 Supervised Learning:.....	35
3.2.2 Unsupervised Learning:	36
3.2.3 Semi-supervised Learning:.....	36
3.2.4 Reinforcement Learning:.....	36

3.3	Regression:	36
3.3.1	Linear regression:	36
3.3.2	Polynomial regression:	36
3.3.3	Logistic regression:	37
3.4	Classification:	37
3.4.1	Binary classification:	37
3.4.2	multi-class classification:	38
3.4.3	multi-label classification:	39
3.4.4	Imbalanced classification:	40
3.5	supervised learning algorithms used:	40
3.5.1	K-Nearest Neighbors (KNN):	40
3.5.2	Decision Trees:	41
3.5.3	Random Forest:	42
3.5.4	Support Vector Machine (SVM):	42
3.5.5	Naïve Bayes:	43
3.5.6	Logistic Regression:	44
3.5.7	Gradient Boosting:	44
3.6	the machine learning process:	45
3.6.1	Data Collection:	45
3.6.2	Data cleaning:	46
3.6.3	Training a model:	46
3.6.4	Testing a model:	46
3.6.5	Enhance performance:	47
4	Result and discussion:	48
5	Predictions and Evaluation:	49
	Conclusion:	50
	Chapter 4	28
	Introduction:	51
1	General information on wastewater:	51
	Definition:	51
1.1	Wastewater Types:	52
1.1.1	Domestic wastewater:	52
1.1.2	Industrial wastewater:	52
1.1.3	Rainwater:	52
1.1.4	Agricultural water:	53

2.1	Characteristics of wastewater:	53
2.1.1	Physical Characteristics:	53
2.1.2	Chemical Characteristics:.....	54
2.1.3	Microbiological characteristics:	55
3.1	sanitation systems:	55
3.1.1	Types of sanitation:	55
2	presentation of the wastewater treatment plant:	56
2.1	Internship location:	56
2.2	The opening of the station:	57
2.3	BASIC DATA:	57
2.4	Justification of the adopted solution:	58
3	stages of treatment of waste water in the station:.....	59
3.1	Pre-treatment:	59
3.1.1	Grilling:	59
3.1.2	Sand Removal:	60
3.1.3	Grease and Oil Removal:	61
3.2	Primary treatments:	62
3.2.1	Natural sedimentation:	62
3.2.2	Physico-chemical sedimentation:	62
3.3	Secondary treatment (Biological):	63
3.3.1	Anaerobic treatments:	63
3.3.2	Aerobic treatments:	63
3.4	Tertiary treatments:	64
3.4.1	Nitrogen and phosphorus removal:	64
	Conclusion:.....	64
	General conclusion.....	66
	Bibliography.....	67

Lists of abbreviations:

IOT	Internet of Things
IIoT	Industrial Internet of Things
RGB	Red, Green, Blue
CWSI	Crop Water Stress Index
UAVs	Unmanned aerial vehicles
NDVI	Normalized Difference Vegetation Index
GIS	Geographic Information System
IGS	Intelligent Greenhouse System
MQTT	Message Queuing Telemetry Transport
AMQP	Advanced Message Queuing Protocol
CoAP	Constrained Application Protocol
DDS	Data Distribution Service
HTTP	Hypertext Transfer Protocol
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
HRAPs	High-Rate Algae Ponds
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
SVM	Support Vector Machine
WWTP	Wastewater Treatment Plant
ANNs	Artificial Neural Networks
UCI	University of California Irvine

IWQ	Irrigation Water Quality
EC	Electrical Conductivity
N	Nitrate
TN	Total Nitrogen
P	Phosphate
TP	Total Phosphorous
BOD	Biochemical Oxygen Demand
COD	Chemical Oxygen Demand
SS	Suspended Solid
VAE	Variational Autoencoder
RNN	Recurrent Neural Network
RBM	Restricted Boltzmann Machine
MBR	Membrane Bioreactor
RAS	Return Activated Sludge
DBNs	Deep Belief Networks
OCSVM	One-Class Support Vector Machines
IC	Ion Chromatography
CNN	Convolutional Neural Network
BiLSTM	Bidirectional Long Short-Term Memory
EWMA	Exponentially Weighted Moving Average
SVR	Support Vector Regression
k-NN	k-Nearest Neighbors
IWA	International Water Association
ASM	Activated Sludge Model
ADM1	Anaerobic Digestion Model Number 1
USA	United States of America
VFA	Volatile Fatty Acids
MySQL	My Structured Query Language
MSSQL	Microsoft SQL Server
NumPy	Numerical Python

SciPy	Scientific Python
Pandas	Panel Data
Scikit Learn	Sci-kit Learn
NLTK	Natural Language Toolkit
Jupyter	Julia, Python, and R
TDS	Total Dissolved Solids
SSS	Settleable Suspended Solids
T	Temperature
Cond	Conductivity
Turb	Turbidity
TSS	Total Suspended Solids
BOD	Biological Oxygen Demand
N-NH ₄ ⁺	Ammonium Nitrogen
N-NO ₃ ⁻	Nitrates
N-NO ₂ ⁻	Nitrites
PO ₄ ⁻³	Phosphates
O ₂	Dissolved Oxygen

Lists of tables

Table 1: Software Configuration	28
Table 2: Characteristics of PH sensor	31
Table 3: Characteristics of Temperature sensor	32
Table 4: Characteristics of Electrical conductivity sensor	33
Table 5: Characteristics of Turbidity sensor	33
Table 6: Characteristics of Dissolved oxygen sensor.....	34
Table 7: A comparative study between algorithms	48
Table 8: Values at processing entry	58
Table 9: Dismissal Limit Values	58

Lists of figures

Figure 1: Internet of thing	4
Figure 2: Characteristics of IOT	6
Figure 3: IoT reference model.....	7
Figure 4: IoT applications in precision agriculture	11
Figure 5: The major components of an IoT-based precision agriculture system.....	14
Figure 6: Logo Python	29
Figure 7: Logo Anaconda.....	29
Figure 8: Logo Jupyter	30
Figure 9: Logo Arduino	30
Figure 10: Arduino card	31
Figure 11: PH sensor.....	31
Figure 12: Temperature sensor.....	32
Figure 13: Electrical conductivity sensor.....	32
Figure 14: Turbidity sensor.....	33
Figure 15: Dissolved oxygen sensor	34
Figure 16: Types of machine learning models	35
Figure 17: Binary ranking	38
Figure 18: Multi-class classification	39
Figure 19: Multi-label classification	39
Figure 20: Imbalanced classification.....	40
Figure 21: K-Nearest Neighbors (KNN)	41
Figure 22: Decision Trees	41
Figure 23: Random Forest.....	42
Figure 24: Support Vector Machine (SVM).....	43
Figure 25: Naïve Bayes	43
Figure 26: Logistic Regression.....	44
Figure 27: Gradient Boosting	45
Figure 28: the machine learning process	45
Figure 29: dataset of water quality	46
Figure 30: dataset after data cleaning.....	46
Figure 31: location of the station with google maps	57
Figure 32: Rough and manual grid.....	60
Figure 33: Sand Classifier	60
Figure 34: Sewage Treatment Plant Sands.....	61
Figure 35: Grit and Grease Separator.....	61
Figure 36: Primary clarifier.....	62
Figure 37: Aeration tank.....	63

General introduction

The increasing global demand for water resources and the need for sustainable agricultural practices have propelled the exploration of alternative water sources for irrigation. Utilizing wastewater for irrigation presents a viable solution, offering significant benefits such as high nutrient content, which reduces the dependence on chemical fertilizers, and aiding in the conservation of water resources. Particularly in arid regions, treated wastewater has become a reliable resource for farmers, enabling continuous crop cultivation throughout the year.

In this context, the integration of advanced technologies into wastewater management systems offers a promising avenue for enhancing irrigation practices. This study introduces a novel approach that leverages artificial intelligence (AI) and edge computing capabilities to provide real-time monitoring and prediction of water quality. Adhering to the parameter thresholds established by the World Health Organization (WHO), this approach aims to ensure the safe and efficient use of wastewater for irrigation.

This master thesis explores both the front-end and back-end components of the proposed system, including software and hardware architectures. The hardware setup comprises actuators, sensors, and controllers connected via wireless networks, while the software integrates AI models for intelligent predictions. The dataset used in this study encompasses key parameters essential for assessing water quality—conductivity, turbidity, oxygen levels, pH, and temperature—all monitored by sensors within the sensor module.

Through this integrated system, the study aims to demonstrate the potential of AI and edge computing in revolutionizing wastewater irrigation practices, contributing to sustainable agriculture and water resource management.

Thesis Organization

The thesis is organized into several chapters, each detailing a specific aspect of the study:

General Introduction: An overview of the increasing global demand for water resources and the potential benefits of using wastewater for irrigation.

Chapter 1: IoT (Internet of Things): This chapter introduces the concept of IoT, its importance, characteristics, reference model, advantages, and disadvantages.

Chapter 2: AI and Machine Learning in Water Quality Monitoring: Explores various AI and machine learning models used for predicting water quality parameters.

Chapter 3: System Design and Implementation: Details the design stages, technological choices, and practical implementation of the IoT and AI-based system.

Chapter 4: Internship Report: Provides insights into the practical experience gained at the wastewater treatment plant, detailing the processes and technologies observed.

Results and Discussion: Analyzes the performance of different algorithms used for water quality prediction and their comparative effectiveness.

Predictions and Evaluation: Discusses the selection and retraining of the best-performing model for making predictions on new data instances.

Conclusion: Summarizes the development and implementation of the innovative system, emphasizing its contributions to optimizing wastewater use in irrigation.

Each chapter builds on the previous ones, culminating in a comprehensive study that bridges theoretical concepts with practical applications, ultimately proposing a sustainable solution for modern agricultural challenges.

Chapter 1

IoT in Precision Agriculture

Introduction:

Internet is currently experiencing a remarkable surge in growth with the increasing number of devices connecting to it. In the past, we only had personal computers and mobile phones connected to the internet, but now, with the Internet of Things (IoT) concept, millions of devices are connecting to it. This development in IoT leads to the idea of machine-to-machine communication, meaning that two machines can communicate with each other, and all the data that was previously on a private server is now available on the internet so that users can access it remotely. The application of IoT is feasible in almost every sector, especially where communication speed is not a concern.

Precision agriculture is essentially a concept that emphasizes delivering the right number of resources at a specific moment and for an exact duration. These resources can include water, light, pesticides, and more [1].

IoT facilitates the implementation of precision agriculture by enabling more efficient resource management and data-driven decision-making.

In this chapter, we will first provide an overview of IoT, citing its areas of application (cities, healthcare, transportation, industry, agriculture), and then we will discuss precision agriculture.

1. IOT (Internet of Things):

The term "Internet of Things" (IoT) refers to the interconnection of various physical devices and objects worldwide via the internet. Coined by Kevin Ashton in 1999, IoT encompasses the networking of everyday objects to enable them to send and receive data. The subsequent section outlines the fundamental aspects of IoT, covering its different layers and key terminology. Essentially, IoT represents an expansion of the services offered by the internet. This section also delves into the architecture of IoT systems. For instance, when household devices in our daily lives are connected to the internet, the resulting system can be termed a Smart Home within an IoT environment. IoT is not merely a futuristic concept; it is already being implemented and is impacting more than just technological development [2].

1.1 Definition:

The Internet of Things (IoT) provides the ability for human and computers to learn and interact from billions of things that include vehicles, home appliances, and other items embedded with electronics, software, sensors, actuators, services, and other Internet connected objects [3].

The Internet of things is a platform where regular devices are connected to the internet so, they can interact, collaborate and exchange data with each other.

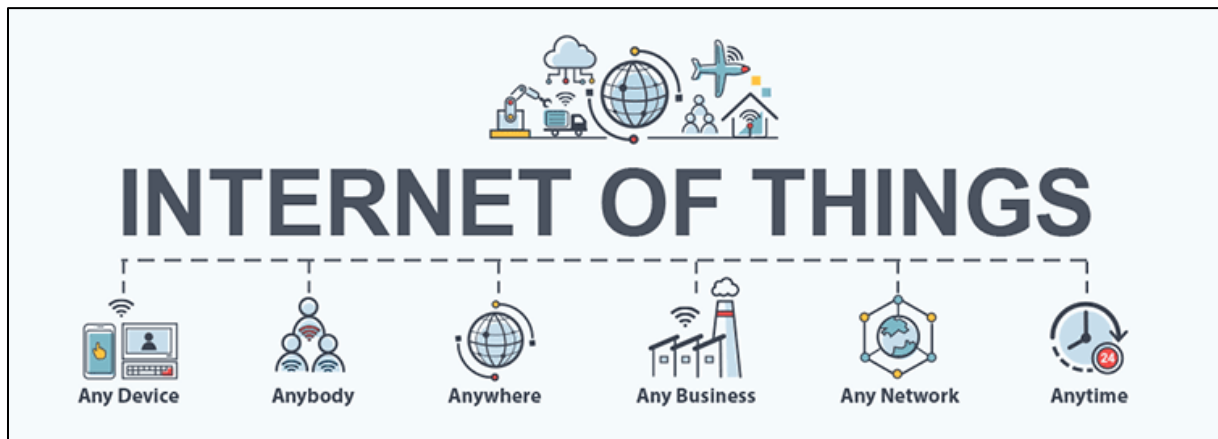


Figure 1: Internet of thing

1.2 Importance of IOT:

The Internet of Things helps people live and work smarter and more efficiently. Not only do these smart devices simplify the home, IoT is critical to businesses as well. IoT provides companies with real-time information on how operational processes are performing and provides insights into sales activities and equipment efficiency supply chains. IoT also enables companies to streamline processes and reduce personnel costs. It further improves service quality, reduces waste, lowers the cost of producing and distributing products, and ensures consumer purchasing responsibility. In plantation cultivation, IoT helps increase yields and efficiency. The Internet of Things is one of the most important innovations [4].

1.3 Characteristics of the Internet of Things:

The Internet of Things (IoT) is characterized by the following key features that are mentioned below:

1.3.1 Connectivity:

Connectivity is a key feature of the Internet of Things. With IoT, you can connect your cell phone, laptop, and other internet-connected devices. Anyone can learn about anything, anytime, anywhere.

The Internet of Things can be connected through a variety of wireless devices such as sensors, mobile phones, trackers, etc. This way, users do not need to wait for an Internet connection to operate the device.

1.3.2 Identity of Things:

The interaction of names and numbers provides identity to Internet devices. Assigning identities to devices is an important aspect of IoT. Identity helps distinguish different internet devices and select the one we want to send commands to.

Each device requires different control capabilities depending on the type of data provided. It is important to give each device a unique identity so that we can set passwords or other security measures. For example, fingerprints, facial recognition IP addresses, and face lock systems are among the various security measures that provide protection for various identified devices.

1.3.3 Data:

Analyse data collected by IoT devices to make future predictions. For example, a calorie meter. It helps regulate the number of calories per day. We also have fitness data, thermostats, and various devices that monitor our health. Therefore, we may use data collected through these devices.

1.3.4 Intelligence:

The intelligence of IoT devices depends on the intelligence of sensors. The sensor sends the data to the user for further analysis.

We need to regularly update IoT devices to do smart work. It extends their functionality and makes them smarter.

1.3.5 Network:

As the number of IoT devices on the network increases, maintaining communications to function properly becomes difficult. However, cloud services and gateways are some of the ways that can solve such problems.

Often, a device can leverage another device's connection to establish a network connection, even if the second device is not connected to the network. Because IoT devices can communicate with each other, it is more efficient and adaptable than other current technologies.

1.3.6 Scalability:

Scalability refers to the amount of data that can be processed efficiently. IoT creates a setup to process large amounts of data and generate useful analytics.

1.3.7 Architecture:

The architecture of IoT is designed to support a variety of devices, technologies, and protocols. Its main task is to check that each connected device does not interfere with other devices. In this way, the data security of each device is guaranteed.

1.3.8 Security:

As the number of IoT devices continues to grow, concerns regarding the security of personal data also emerge. With the collection, exchange, and generation of vast amounts of data, the potential for data breaches increases. The risk of personal data being transferred without consent is a significant concern.

To address this challenge, IoT establishes privacy-preserving networks, systems, and devices. Maintaining security poses a major challenge for IoT. Nonetheless, it manages to navigate through without major disruptions [5].



Figure 2: Characteristics of IOT

1.4 The IoT reference model:

Figure 3 presents the IoT reference model. This model comprises four layers associated with management and security capabilities. These four layers are as follows:

- Application Layer
- Service and Application Support Layer
- Network Layer
- Device Layer

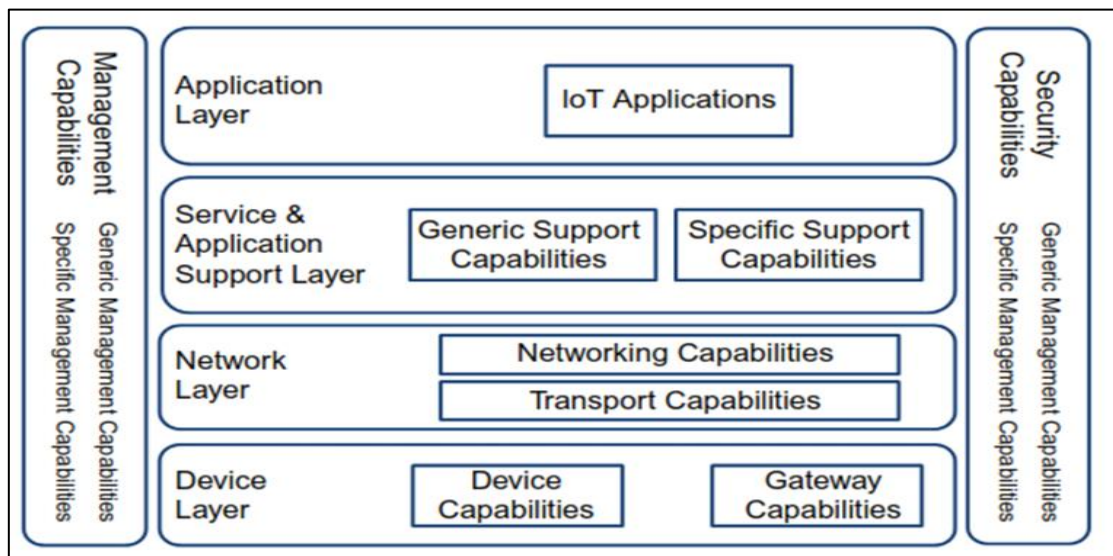


Figure 3: IoT reference model

1.4.1 Application Layer:

The application layer contains IoT applications.

1.4.2 Service and Application Support Layer:

The Service and Application Support Layer encompasses two categories of capabilities:

- **Generic Support Capabilities:** These are common functionalities utilized by multiple IoT applications, such as data processing and storage. Specific support capabilities may also leverage these generic functions to create tailored features.
- **Specific Support Capabilities:** Tailored functionalities designed to fulfil the requirements of various applications. Comprising distinct sets of capabilities, they cater to diverse support functions for different IoT applications.

1.4.3 Network Layer:

It is composed of two sets of capabilities:

- Network-related capabilities:

These ensure control over connectivity, such as resource access control, mobility management, and authentication.

- Transport-related capabilities:

These ensure the necessary connectivity for transporting information specific to each IoT service or application, as well as for transporting control and management information related to IoT.

1.4.4 Device Layer:

The capabilities of the device layer can be divided into two categories:

- Device capabilities:

Devices collect and then send data directly over the communication network (direct interaction with the communication network).

Devices form ad hoc networks in some cases, which require high modularity.

Devices may include sleep/wake mechanisms to conserve energy.

- Gateway capabilities:

Gateway capabilities are necessary for protocol conversion; for example, ZigBee at the device layer and 3G at the network layer [6].

1.5 Advantages and disadvantages:

The Internet of Things (IoT) is a technology that enables physical objects to communicate with each other and with the Internet.

Here are some advantages and disadvantages associated with this technology:

1.5.1 Advantages:

- Automation: A lot of jobs may be automated with the help of IoT, which can improve productivity and simplify daily life.

- **Monitoring:** The Internet of Things enables remote monitoring of environments and things. This can be helpful, for instance, in monitoring weather stations, water distribution networks, and industrial operations.
- **Accuracy:** IoT can help operations run more smoothly. IoT-based inventory tracking solutions, for instance, can offer precise real-time stock level information, reducing the risk of stockouts and waste.
- **Innovation:** The Internet of Things (IoT) presents chances for innovation in a number of sectors, including energy, transportation, healthcare, agriculture, and industry.

1.5.2 Disadvantages:

- **Security:** Since many objects are linked to the Internet, making them potentially vulnerable to cyberattacks, the Internet of Things may pose security risks.
- **Life:** The Internet of Things has the potential to gather vast amounts of personal data, which could provide privacy protection issues. The information that might be used to track user behaviour, which could be viewed as intrusive.
- **Complexity:** The Internet of Things is a sophisticated technological system that requires specialized skills to set up and maintain.
- **Costs:** The implementation of the Internet of Things may be required due to the need to purchase sensors and connected devices as well as the required infrastructure.

2. The application domains of IoT:

2.1 Healthcare:

Within the healthcare sector, IoT devices offer the capability to remotely monitor patients, gathering real-time data on vital signs like heart rate, blood pressure, and oxygen saturation. Analysing this sensor data enables the detection of patterns and early identification of potential health concerns. Furthermore, IoT devices facilitate the tracking of medical equipment, inventory management, and monitoring medication adherence.

2.2 Manufacturing:

In manufacturing, Industrial IoT (IIoT) devices play a vital role in monitoring machine performance, identifying equipment failures, and streamlining production processes. For instance, sensors can monitor temperature and humidity levels in manufacturing facilities to maintain ideal conditions for sensitive product production. Additionally, IIoT devices enable

inventory tracking, supply chain management, and quality control monitoring for finished products.

2.3 Agriculture:

In agriculture, IoT devices find utility in monitoring soil conditions, weather changes, and crop development. For instance, sensors can gauge soil moisture levels, facilitating timely irrigation for crops. Additionally, IoT devices aid in livestock health monitoring, equipment tracking, and supply chain management. Furthermore, low-power or solar-powered devices are often deployed in remote areas with minimal supervision.

2.4 Transportation:

Within the transportation sector, IoT devices serve to monitor vehicle performance, streamline routes, and oversee shipments. For instance, sensors can assess the fuel efficiency of connected vehicles, thereby cutting fuel expenses and enhancing environmental sustainability. Moreover, IoT devices enable the monitoring of cargo condition to guarantee its arrival at the destination in prime condition [7].

3. IoT and precision agriculture:

Due to population growth, agricultural demand has significantly increased. With the introduction of technology, this century has witnessed a shift from traditional methods to more advanced ones. The Internet of Things (IoT), today and tomorrow, will impact every aspect of everyone's life by making it smart. It's a network of various computers that can configure themselves. The Internet of Things (IoT) is changing the efficiency and volume of agriculture. The integration of IoT, Big Data analytics, wireless sensor networks, and cloud computing offers a wide range of possibilities for predicting, processing, and analysing environments and improving operations in real time. The objective is to provide farmers with the necessary technology to record real-time data on crops, soil moisture, temperature, and lighting so that they can take necessary steps towards smart farming. This also enables farmers to use limited resources to increase yields. This review highlights the importance of IoT in agriculture, the different components used in agricultural IoT [8].

3.1 Applications of IOT in agriculture:

The Internet of Things (IoT) will change lifestyles worldwide. By utilizing the latest wireless sensors and IoT technologies in agriculture, every aspect of traditional agricultural methods can be fundamentally altered. In order to meet the ongoing population growth, the agricultural

industry ideally should adopt IoT technology in the future. For instance, to address the increasing demand for food production, farmers must confront challenges such as climate change due to global warming, highly variable weather conditions, and the environmental impacts of pollution and other agricultural practices. Smart agriculture through IoT technology can assist farmers in minimizing waste and increasing productivity. Smart agriculture through IoT technology can help farmers minimize waste and increase productivity. Smart agriculture is fundamentally a high-tech, clean, and sustainable food production facility on a large scale.[8]

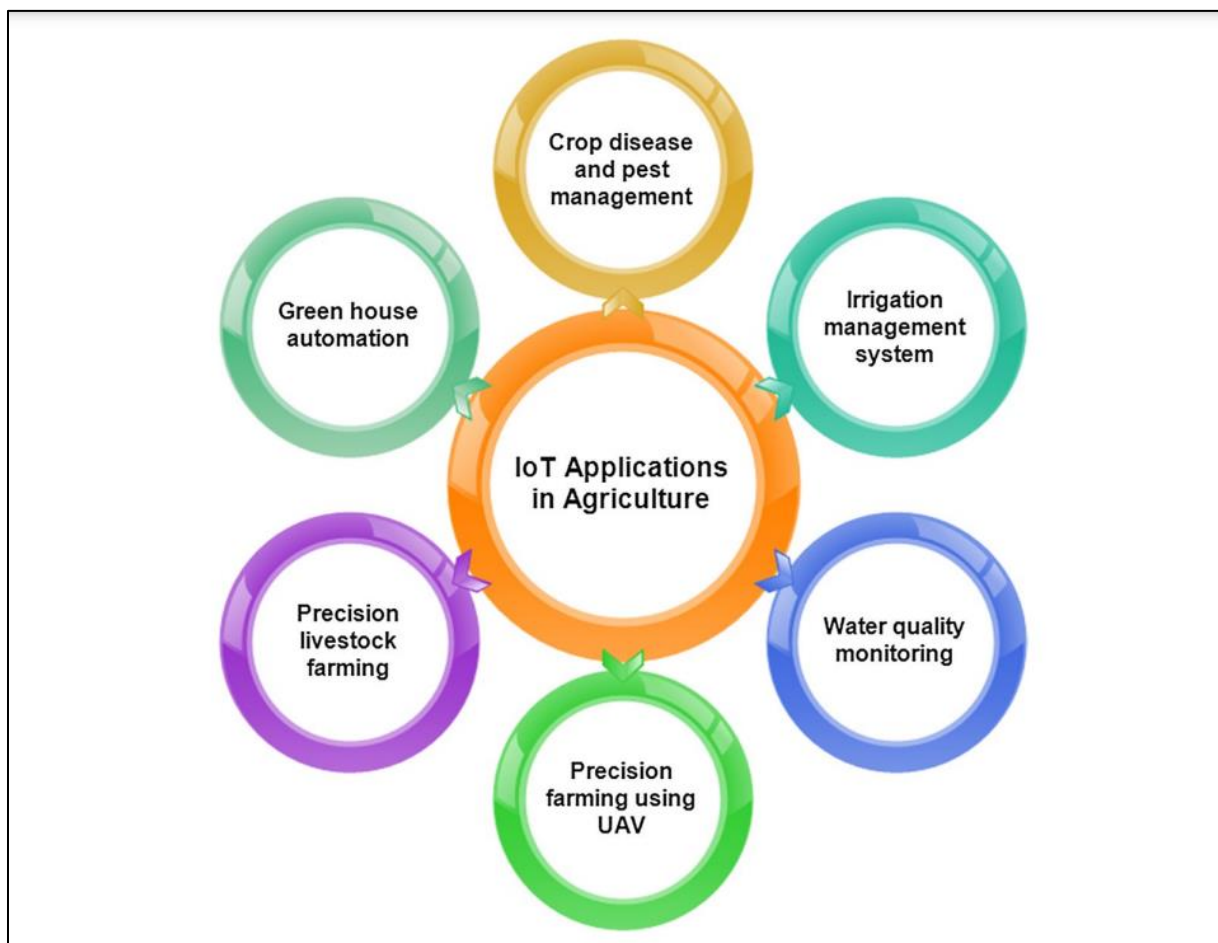


Figure 4: IoT applications in precision agriculture

3.1.1 Irrigation Management System:

Data collection through sensors enables precise irrigation cycles determination. Thus, IoT applications can facilitate swift water management. As a result, water wastage decreases day by day. Research in the field of irrigation systems aims to streamline water usage for various crops, from the most basic to the most advanced. To achieve water conservation, a framework for irrigation systems employing various technologies has been proposed, such as thermal

imaging, RGB (Red, Green, Blue) imaging, Crop Water Stress Index (CWSI), direct measurement of soil moisture, etc.

3.1.2 Water Quality Monitoring:

Another issue that can affect crop health and limit harvests is water quality. Temperature, pH, conductivity, dissolved oxygen, and other variables all impact water quality. IoT-based water quality monitoring systems are emerging as a solution to this challenge. These devices can be used to remotely monitor and regulate the physical and chemical properties of water.

3.1.3 Precision Farming Using UAV:

Unmanned aerial vehicles (UAVs) and drones are also making their way into the agricultural sector in the modern era. Agricultural drones are an intriguing innovation in IoT-based agriculture and are utilized in practice in two types: ground drones and aerial drones. They can be programmed to detect details such as Normalized Difference Vegetation Index (NDVI), water stress, or specific nutrient deficiencies in crops. This provides advantages in terms of ease of use, time savings, crop health imaging, integrated Geographic Information System (GIS) mapping, and increased yields.

3.1.4 Precision Livestock Farming:

In livestock monitoring and management systems, IoT applications have a significant impact. Here, location trackers are implanted in the livestock to track them while they graze. As a result, the health status of all animals is recorded simultaneously (with the support of IoT technology).

3.1.5 Green House Automation:

Rapid climate changes have an impact not only on agriculture but also on the agricultural system in place. In this context, IoT applications play a crucial role in improving agriculture by installing sensors and actuators both outside and inside the agricultural sector. Moreover, within the framework of the intelligent greenhouse system (IGS), IoT applications adjust the climate conditions according to specific pre-established instructions. This can be achieved using sensors capable of collecting real-time data to monitor the automatic irrigation system in the IGS.

3.1.6 Crop Disease and Pest Management:

Crop diseases and pests have caused significant losses to farmers since the dawn of agriculture. The Internet of Things (IoT) provides a solid foundation for developing effective

agricultural disease and pest management strategies. This involves three elements: detection, evaluation, and treatment, which are complex tasks in traditional farming practices [9].

3.2 Precision agriculture:

Precision agriculture is an agricultural management term that focuses on observing, estimating, and responding to variability within fields and between crops. The aim of precision agricultural research is to develop a comprehensive decision support system in agricultural management to maximize yields while preserving capital. Farmers are able to pinpoint their exact location within a field, enabling them to map spatial variations of numerous factors such as crop yields, soil nutrient and moisture levels, topography, nitrogen levels, pH, magnesium, potassium, and other essential plant elements. In recent years, precision agriculture has become one of the most common applications of IoT in agriculture, with numerous organizations worldwide beginning to implement this technology [8].

3.3 The major components of an IoT-based precision agriculture system:

IoT based smart farming consist of four major components as shown in Figure 5. These four major components are physical structure, data acquisition, data processing, and data analytics.

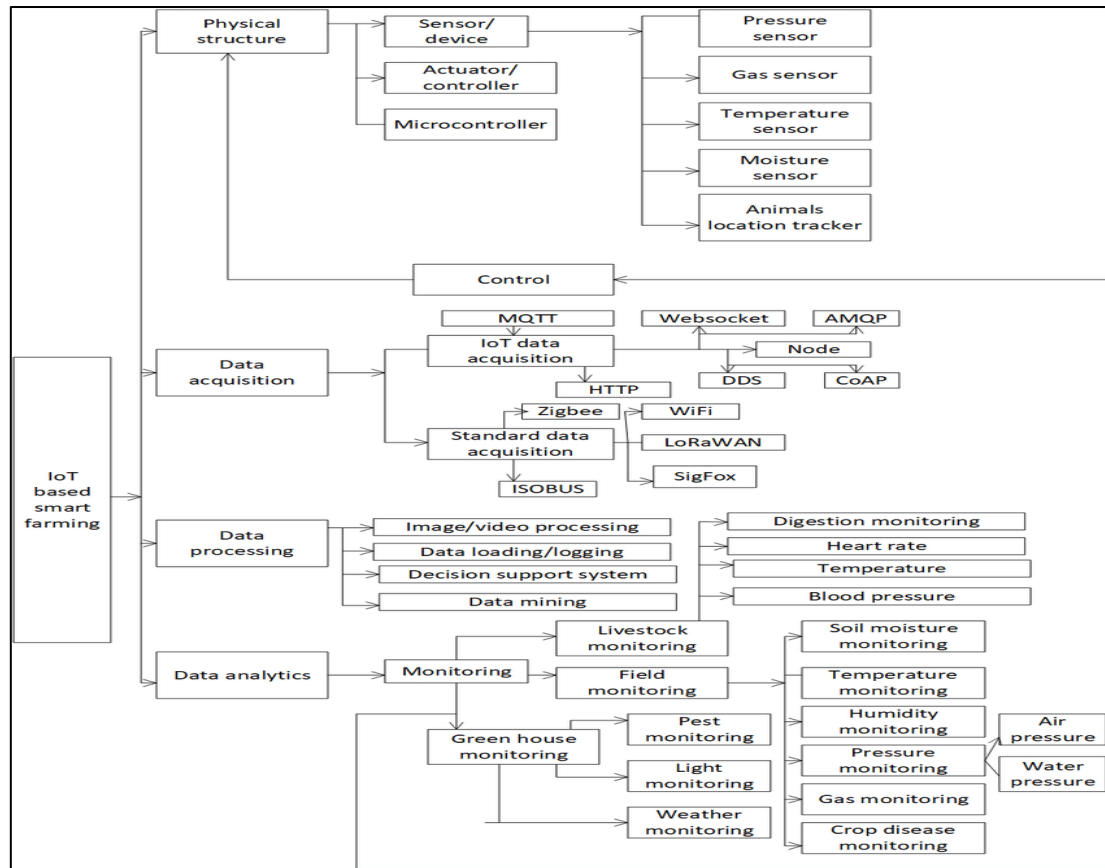


Figure 5: The major components of an IoT-based precision agriculture system

3.3.1 The physical structure:

is the most critical aspect of a precision agriculture system, as it helps to prevent any undesirable events. The system is designed to oversee sensors, actuators, and devices. A sensor performs multiple tasks such as measuring soil moisture and temperature. It can also carry out various control functions such as node discovery and device identification.

3.3.2 Data acquisition:

is divided into two sub-components:

- IoT data acquisition consists of seven protocols: MQTT, WEB-SOCKET, AMQP, Node, CoAP, DDS, and HTTP.
- Standard data acquisition encompasses protocols such as ZigBee, WIFI, LoraWan, and SigFox.

3.3.3 Data processing:

comprises several functionalities, including image or video processing and decision support system, tailored to the system's requirements.

3.3.4 Data analysis:

consists of two main functionalities: monitoring and control. Monitoring includes three primary applications in precision agriculture: animal monitoring (temperature, heart rate, and digestion), field monitoring, and greenhouse monitoring [10].

3.4 The benefits of IoT in precision agriculture:

Utilizing IoT in agriculture presents a multitude of benefits, a selection of which is outlined below [11]:

3.4.1 Data Collection:

Smart agricultural sensors can gather data on weather conditions, soil quality, crop growth progress, or livestock health. These data points can be utilized to monitor your farm's overall status as well as personnel performance, equipment efficiency, etc.

3.4.2 Better Yield Management:

Being able to forecast your production yield allows for better product distribution planning. Knowing exactly how much harvest you can expect ensures you avoid unsold inventory.

3.4.3 Cost Management Improvement:

By spotting anomalies in crop growth or livestock health, you'll be able to anticipate yield loss risks, contributing to better cost control.

3.4.4 Increased Efficiency:

By employing smart devices, you can automate several processes throughout your production cycle, such as irrigation, fertilization, or pest control.

3.4.5 Product Quality and Volume Enhancement:

Process automation allows for better production control, maintaining crop quality and growth capacity [11].

Conclusion:

Throughout this chapter, we have been able to define the impact of IoT on the development of various fields, with a particular focus on the realm of precision agriculture.

Our focus in this thesis is wastewater monitoring for irrigation purposes //////////////

Chapter 2

AI in wastewater treatment

Introduction:

The application of artificial intelligence (AI) and the Internet of Things (IoT) to wastewater treatment is a recent development that has enabled significant progress and milestones. These advancements include the use of AI algorithms for data analysis and predictive modeling, as well as the integration of IoT devices to monitor and control wastewater treatment systems in real time.

In this chapter, I will talk about a comprehensive overview of the uses of artificial intelligence and the Internet of things in wastewater treatment, and also conduct a comparative study on various studies on this topic, by clarifying the objective and defining the limitations of each study.

1. Integration of AI in Wastewater Treatment:

Artificial intelligence (AI) can optimize wastewater treatment processes by analyzing large amounts of data and creating predictive models. This enables operators to make informed decisions and take proactive measures to improve system efficiency. For example, AI algorithms can analyze sensor data to predict equipment failures or optimize chemical dosing for better treatment performance. Additionally, AI has the potential to enhance the energy efficiency of wastewater treatment. By analyzing data from various sensors and integrating weather forecasts, AI algorithms can optimize energy consumption, leading to significant cost savings and a reduced environmental impact.

2. Related work:

Study	Objective	Dataset	Sensors	Low cost	M L	DL	Anomaly detection	Limitations
[12]	Research aims to create a cost-effective soft sensor for wastewater treatment by developing a data division algorithm and innovative method for measuring ammonium content using inexpensive sensors. An annotated dataset will also be released to support reproducible studies.	the secondary wastewater treatment plant at Luggage Point, Queensland, Australia. The data was collected from the High-Rate Algae Ponds (HRAPs) as part of the wastewater treatment process. (real time data)	pH; Dissolved Oxygen; Turbidity; Total Suspended Solids; Ammonium and Nitrate; temperature; rain.	Yes	no	yes	involves identifying days with poor predictions across different models, particularly during extreme wet weather events.	This work faces limitations due to reliance on high-quality data for deep learning models and the complexity of these models, making implementation and interpretation challenging for non-specialized users.
[13]	This article focuses on using deep learning models like LSTM and GRU as soft-sensors to predict key features of a wastewater treatment plant. The goal is to	coastal municipal wastewater treatment plant located in Saudi Arabia. (real time data)	No sensors used.	Yes	no	yes	The project utilized a one-class support vector machine (SVM) algorithm combined with deep learning methods for	Limitations of the project include dependence on specific facility data, lack of transparency in soft sensors, and need for larger, higher

	improve forecasting accuracy and speed in dealing with complex univariate WWTP processes.						anomaly detection in wastewater treatment plants, such as seawater intrusion or water supply shutdown.	frequency datasets for improved performance.
[14]	The project aims to use Artificial Neural Networks (ANNs) to monitor and control wastewater quality for irrigation, optimizing resource management and environmental sustainability. It integrates a wireless sensor network system for enhanced accuracy.	The dataset used in the project is collected from sensor nodes deployed in an actual wastewater treatment station in Morocco. (real time data)	physicochemical sensors to convert water descriptors into electrical signals for analysis.	Yes	yes	No	The project uses Artificial Neural Networks (ANNs) to detect anomalies in wastewater quality parameters, enabling real-time intervention and corrective actions for potential contamination issues.	The project faces limitations due to potential accuracy issues in extreme environmental conditions and potential interference in the wireless sensor network, necessitating improvement to enhance robustness in challenging operational conditions.
[15]	The project aims to develop a low-cost IoT smart device to monitor water	University of California Irvine (UCI) water treatment	electrical conductivity.	yes	yes	no	Anomaly detection in this project involves predicting electrical	The project's limitations include the need for future calculations of

	quality in real-time, automating data collection and analysis using Artificial Neural Networks for improved treatment processes.	plant (real time data)					conductivity of wastewater using machine learning algorithms to identify any unusual or unexpected variations in the data that may indicate potential issues or abnormalities in the wastewater treatment process.	controller parameters and risk assessments for different plant types to better manage wastewater irrigation risks.
[16]	The project aims to improve wastewater treatment facilities' machine learning applications by providing a structured framework, guiding water managers, emphasizing stakeholder engagement, ongoing monitoring,	The data is obtained from online sensors, on-site laboratory analyses, off-site laboratory analyses, as well as visual observations conducted by operators in wastewater treatment facilities.	pH, dissolved oxygen, turbidity, temperature, flow rates.	yes	yes	no	This project uses machine learning algorithms to detect anomalies in wastewater treatment data, alerting operators to potential issues, enabling proactive maintenance.	The project faces challenges in data availability, quality, and cost in wastewater treatment facilities, and requires expertise in data science and domain knowledge for meaningful machine learning applications.

	and data quality.	(real time data)					nce and intervention to improve system reliability and performance.	
[17]	The objective of this project is to develop and evaluate machine learning models to predict 10 Irrigation Water Quality (IWQ) parameters in the Bouregreg watershed, focusing on improving the monitoring and assessment of water suitability for irrigation purposes.	Bouregreg watershed in Morocco, specifically from surface water quality recorded at 9 monitoring stations in the region. (real time data).	Electrical Conductivity (EC), pH	yes	yes	No	The project does not specifically mention anomaly detection. It primarily focuses on using machine learning models to predict irrigation water quality parameters accurately based on the input variables measured by automated sensors.	The project's limitations include models' inability to predict extreme values of parameters like Managed Aquifer Recharge and Permeability Index, decreased prediction accuracy in different locations, and challenges in predicting parameters accurately.
[18]	The objective of the project is to explore the opportunities and challenges presented	The dataset used in the project is derived from diverse types of wastewater	NO ₃ -N(Nitrate), TN (Total Nitrogen), NH ₄ -N(Ammonium), PO ₄ -	yes	no	yes	This project uses deep learning techniques like transformer-based models,	The project faces limitations due to the lack of large public datasets in the wastewater

	by deep learning methods in wastewater modelling. The review aims to discuss the value of different modelling approaches in wastewater processes and identify open research problems in the field.	r treatment processes. It mentions that some public datasets have been introduced on data hubs such as Arcgis, but many existing public datasets suffer from low frequency and lack important metadata such as climate information.	P(Phosphate), TP (Total Phosphorous), BOD (Biochemical Oxygen Demand), COD (Chemical Oxygen Demand), SS (Suspended Solid).				ResNet-based VAE, and RNN-RBM to improve fault detection, diagnosis, and missing data imputation in wastewater treatment plants.	industry, challenges in interpreting results, and the need for explainable AI to enhance trust and address biases.
[19]	The project presents a fault detection method based on an unsupervised deep learning to monitor operating conditions of wastewater treatment plants (WWTPs)	Mines Park Water Reclamation Test Site, a decentralized wastewater treatment plant located at the Colorado School of Mines campus in Golden, Colorado, USA. (real time data)	membrane bioreactor (MBR) permeate pressure, MBR dissolved oxygen, permeate turbidity, return activated sludge (RAS) dissolved oxygen content, RAS pH, RAS total suspended solids, and	yes	yes	yes	The project uses Deep Belief Networks (DBNs) and One-Class Support Vector Machines (OCSVM) to detect anomalies in wastewater treatment plant data, separating normal	The project's limitations include relying on historical data for machine learning models, potentially excluding future scenarios and issues, and the effectiveness of the anomaly detection approach being influenced by data

			permeate tank conductivity.				from abnormal cases and enabling early system fault identification.	quality and operational conditions.
[20]	The project aims to create a high-precision, fast, and robust soft measurement model using Support Vector Machine (SVM) for real-time monitoring and prediction of wastewater treatment processes, specifically for acid fuchsine dye wastewater.	decolorization reaction test of dyeing wastewater. (real time data)	No sensors	—	yes	No	The project focuses on developing a Support Vector Machine (SVM) model for real-time monitoring and prediction of wastewater treatment processes, without explicitly mentioning anomaly detection.	The project focuses on developing a Support Vector Machine (SVM) model for real-time monitoring and prediction of wastewater treatment processes, without specifying its limitations.
[21]	The project aims to predict COD value of effluent from an IC reactor in a papermaking wastewater treatment plant using	A papermaking wastewater treatment plant. (real time data)	No sensors	—	no	yes	The project uses anomaly detection techniques to identify outliers in the sample set,	The project lacks detailed sensor information, affecting data accuracy and reliability. It focuses on IC

	an artificial neural network model, improving the model's prediction ability and understanding of input factors.						analysing the influence degree of each input variable and using MATLAB programming to visualize data distribution.	reactor effluent COD value prediction but doesn't address other factors affecting efficiency. Lack of comprehensive cost analysis hinders implementation feasibility.
--	--	--	--	--	--	--	--	---

Cost Effective Soft Sensing for Wastewater Treatment Facilities:

The article introduces an innovative soft sensing approach for wastewater treatment, focusing on predicting ammonium concentration to enhance treatment efficiency. Data collected from a secondary wastewater treatment system (HRAP) and a weather station are utilized for analysis. The GRUconv model, which combines GRU and CNN networks, is proposed to provide superior performance in predicting ammonium levels while balancing inference quality and training cost. This approach aims to improve the monitoring of process variables in algae ponds to optimize wastewater treatment performance [12].

Forecasting of Wastewater Treatment Plant Key Features Using Deep Learning-Based Models: A Case Study:

The article focuses on forecasting key characteristics of wastewater treatment plants using deep learning-based models. LSTM, GRU, BiLSTM, and EWMA models were employed to predict key variables of a municipal coastal wastewater treatment plant in Saudi Arabia. Results showed that the GRU-EWMA model exhibited the highest accuracy, with faster convergence than LSTM. Physical variables were found to be more predictable than chemical/biochemical variables. Recommendations include integrating the models into optimization scenarios, exploring multivariate deep learning models, and considering sensor networks for larger datasets [13].

New approach to monitoring a wastewater irrigation system controlled by the artificial neural network (ANN):

The research article on sustainable groundwater development presents a cost-effective WSN system for wastewater monitoring in irrigation, utilizing ANN for accurate classification of wastewater quality with a high recognition rate of over 95% and an F1 score of 96.81%. The integration of Kohonen self-organizing map and multilayer perceptrons enabled robust performance. The system demonstrated effectiveness in monitoring and controlling wastewater quality, offering valuable insights for algorithmic improvements and optimized resource management, thereby contributing to environmental sustainability in agriculture [14].

Times Series Prediction used in Treating Municipal Wastewater for Plant Irrigation:

The article presents a method for predicting the electrical conductivity of wastewater for plant irrigation. By combining decision tree and particle swarm optimization, the algorithm achieved a precision of 0.95 in R-squared for prediction. The data used are from the UCI Water Treatment Plant dataset.

The proposed control system aims to enhance the quality of irrigation water based on the electrical conductivity of wastewater. Future work could include plant-specific parameters and risk assessment. In conclusion, this method accurately predicts wastewater quality for plant irrigation, providing a promising solution for wastewater treatment [15].

Application of machine learning at wastewater treatment facilities: a review of the science, challenges and barriers by level of implementation:

This article explores the application of machine learning (ML) in wastewater treatment facilities, highlighting ML application types such as prediction, optimization, and control, with a focus on prediction. Level 3 applications may be hindered by challenges related to boundary definition, while Level 4 applications encompass entire facilities and require extensive sensor networks. Data requirements vary across ML application levels, influencing their feasibility and posing challenges and opportunities for effective implementation [16].

Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment:

This article focuses on predicting irrigation water quality parameters in semi-arid regions using various machine learning models. Models such as Adaptive Boosting have shown better performance compared to others like SVR and k-NN. The studied parameters include sodium, carbonate, bicarbonate, magnesium, salinity risks, and chloride toxicity. The study areas include the Bouregreg watershed in Morocco and the Nfikh basin, with diverse climatic and agricultural conditions. The article also references other studies and models used for predicting and assessing water quality[17].

Deep learning in wastewater treatment: a critical review:

Modeling wastewater treatment processes is essential for optimizing complex systems. Traditional mechanical models, such as the IWA ASM series and ADM1, while effective, have limited efficiency for complex systems. The increasing use of data-driven modeling, particularly deep learning (DL), offers new opportunities to optimize wastewater treatment processes. DL applications, such as performance prediction, process control, and fault detection, are becoming more common in this field. Key deep learning frameworks, such as recurrent neural networks (RNN) and long short-term memory (LSTM) networks, along with model interpretation, play a crucial role in predictive accuracy and operational excellence of wastewater treatment systems [18].

Wastewater treatment plant monitoring via a deep learning approach:

This article proposes an innovative approach to monitoring wastewater treatment plants by combining deep belief networks (DBN) and one-class support vector machines (OCSVM) for fault detection. The approach uses the DBN to capture complex aspects of wastewater treatment plant data and the OCSVM to distinguish between normal and abnormal characteristics. The results show that the DBN-OCSVM method is effective for monitoring wastewater treatment plants and even outperforms the DBN-K-Means algorithm in fault detection. Tests on real data from a wastewater treatment plant in Golden, Colorado, USA, demonstrate the potential of this approach to enhance the monitoring of wastewater treatment facilities [19].

Application of support vector machine in online monitoring of wastewater treatment based on combined kernel functions:

This article studies the application of a combined kernel function support vector machine (SVM) in real-time monitoring of printing and dyeing wastewater treatment. The authors developed an SVM model with a combination of adjustable linear kernel functions for smooth, high-precision measurements. The results show that the combined kernel function model offers high accuracy, fast speed, and strong generalization ability, proving its effectiveness in real-time monitoring of wastewater treatment [20].

Prediction of wastewater treatment system based on deep learning:

This article presents a study on the use of the BP neural network model to predict the COD value of effluents from the IC reactor. The selected input factors were inflow rate, COD, pH, and VFA. The BP neural network model established in MATLAB accurately predicted the COD value, with 75% of the data samples having a prediction error of less than 10%. This approach provides a data-driven method to optimize wastewater treatment processes [21].

Conclusion:

Artificial intelligence (AI) and the Internet of Things (IoT) have enormous potential to transform wastewater treatment processes. By integrating AI algorithms and IoT devices, operators can optimize processes, improve energy efficiency, reduce costs, and enhance the overall performance of the system.

Chapter 3

Conception and Realization

Introduction:

In this chapter, we discuss the software and hardware that contributes to design and implementation.

Chapter 3 of this research focuses primarily on the design and execution of the proposed system. After outlining the prerequisites and specifications in the previous chapters, this chapter delves into how these concepts are transformed into a practical and effective solution. The design and execution phases of the development process are crucial as they determine the feasibility and effectiveness of the proposed solution.

1 Software Tools:

All tests were conducted using a laptop with the software configuration details shown in the table:

Table 1: Software Configuration

Computer brand	DESKTOP-SDSDHAI
System type	64-bit operating system, x64 processor
Operating system	Windows 10 Professionnel
Processor type	Intel(R) Core (TM) i5-7300U CPU @
Processor speed	2.60GHz 2.71 GHz
Installed memory (RAM)	8 GB
Software Tools	Anaconda, Arduino

1.1 Definition of the Python programming language in computer science:

Python is a general-purpose, high-level programming language. It is free, open source, and cross-platform. Python also supports data types such as numeric values, strings, lists, tuples, and dictionaries. Additionally, it is an interpreted language. The interpreter reads the source code line by line. Consequently, it is slower compared to compiler-based languages like C or C++. The syntax of this language is simple and easy to learn. Moreover, Python supports databases such as MySQL and MSSQL. Overall, Python is a versatile language that allows the creation of various applications. It is popular among beginners and developers [22].

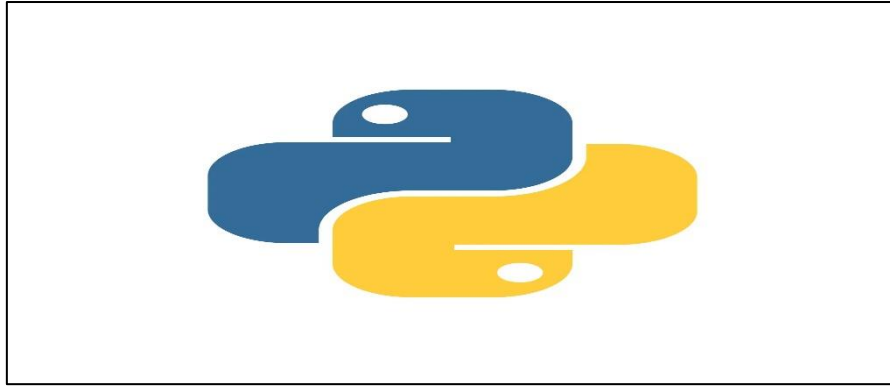


Figure 6: Logo Python

1.2 Definition of anaconda:

Anaconda is a free computer platform. It can be installed on Windows, Linux, MacOS operating systems. It consists of Python and R distributions and a package manager called conda. Anaconda provides a set of pre-installed libraries and packages. Some of them are NumPy, SciPy, Pandas, Scikit Learn, NLTK, and Jupyter [22].



Figure 7: Logo Anaconda

1.3 Definition of jupyter:

Jupyter is a web application used for programming in over 40 programming languages, including Python, Julia, Ruby, R, and Scala. It is a community-driven project aimed at developing open-source software, open formats, and services for interactive computing. Jupyter is an evolution of the IPython project. It enables the creation of notebooks, which are documents containing both Markdown text and executable code in languages like Julia, Python, or R. These notebooks are commonly used in data science for data exploration and analysis [22].



Figure 8: Logo Jupyter

1.4 Definition of Arduino:

Arduino is an open-source electronic platform consisting of a microcontroller card and an integrated development environment. It makes it easy to create interactive objects or devices by connecting various electronic components (sensors, actuators, displays, etc.) to the map and programming it using a simplified programming language based on C/C++. Arduino is widely used for rapid prototyping of electronic projects, in the fields of the Internet of Things, robotics and home technology, thanks to its low cost, flexibility and ease of use [23].

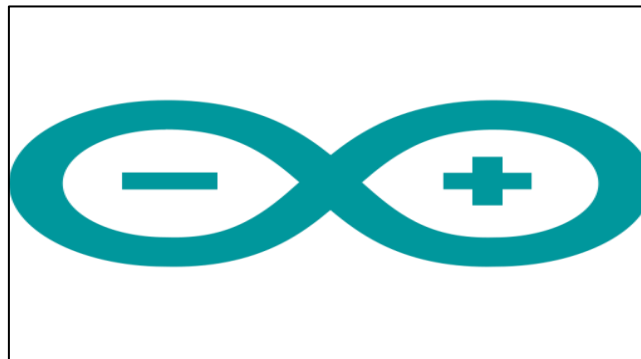


Figure 9: Logo Arduino

2 Hardware Tools:

2.1 The Arduino card:

An Arduino card is a small electronic card equipped with a micro-controller. The micro-controller allows, from events detected by sensors, to program and control actuators; the Arduino card is therefore a programmable interface [24].



Figure 10: Arduino card

2.2 PH sensor:

The pH of water is an element that helps to assess the amount of hydrogen ions that are formed in a certain amount of water. An electric current can be produced by an alkaline solution that contains fewer positively charged ions than an acidic solution. The function of the pH-meter is similar to that of a voltmeter which evaluates the electrical potential generated by the acid solution. The pH difference is measured by comparing the electrical potential generated with the known value, and then the value is deducted [24].



Figure 11: PH sensor

Features:

Table 2:Characteristics of PH sensor

Range of measurement	0-14 Ph
Measuring temperature	0-60°C
Response time	less than 2min
Working voltage	5v
Board dimensions(mm)	58.8(L)x20.0(W)x27.0(H)

2.3 Temperature sensor:

DS18B20 is a Maxim IC 1 wire digital temperature sensor. It indicates degrees in Celsius with a precision of 9 to 12 bits, from -55 to 125 °C (+/-0,5) [24].



Figure 12: Temperature sensor

Features:

Table 3: Characteristics of Temperature sensor

Power supply	3 to 5 Vcc
Accuracy	0.5% from -10 to +85 °C
Measurement range	-55 to +125 °C
Connects to a digital input	
Sonde dimensions	Ø6 x 35 mm
Cord length	90 M

2.4 Electrical conductivity sensor:

The conductivity of the solution refers to the ability of a solution to transmit or transport an electric current. It is crucial to measure the specific conductivity in order to detect impurities in the water. The number of salts and minerals present in water is determined by the total soluble particles (TDS) present in the water [24].



Figure 13: Electrical conductivity sensor

Features:

Table 4: Characteristics of Electrical conductivity sensor

Wide voltage operation	3.3 ~ 5.5V
Analog signal output 0 ~ 2.3V, compatible with 5V, 3.3V two control systems	
The excitation source is a ca signal, effectively preventing polarization of the probe	
Waterproof probe for long-term immersion in water	
Compatible, easy to connect, ready to use, no welding required	

2.5 Turbidity sensor:

A turbidity sensor is used to evaluate the level of water, which is also used to detect water quality through measurements. It can detect suspended particles in the water. By measuring the rate of transmission and diffusion of light, variations in the number of suspended solids (SSS) present in water can be assessed [24].



Figure 14: Turbidity sensor

Features:

Table 5: Characteristics of Turbidity sensor

Usage voltage	5VCC
Consumption	40mA max
Response time	<500ms
Insulation resistance	100M min
Output modes	• Analog Output: 0-4.5V • Digital Output: High/Low Level Signal (Adjustable Resistance Stand)
Usage temperature	5°C~90°C
Weight	30g
Adapter dimensions	38 28 x 10 mm

2.6 Dissolved oxygen sensor:

The dissolved oxygen sensor is used to measure the amount of dissolve oxygen in the water, which allows measurement of water quality. It is commonly used in various areas related to water quality, such as aquaculture, environmental monitoring, natural sciences, etc [25].



Figure 15: Dissolved oxygen sensor

Features:

Table 6: Characteristics of Dissolved oxygen sensor

Power supply	3.3 to 5 Vcc
Measurement range	0 to 20 mg/L
Response time	90 seconds on average
Operating pressure range	0 to 3.44 bar
Electrode life	1 year in normal use
Life of the membrane hood	• dirty water: 1 to 2 months. • clean water: 4 to 5 months
Replacement of filling solution	1 x per month
Sonde interface	BNC
Card interface	analog on connector Gravity (0 to 3 Vcc)
Dimensions	• Converter : 42 x 32 x 7 mm (sans connecteur BNC). • Probe: Ø16,5 x 14.2 mm. • Cord length: 2 meters.

3 Development tools

3.1 Machine Learning:

In terms of usage and domain, artificial intelligence is undoubtedly one of the most popular words and exists in almost every detail of our lives. Artificial intelligence is a branch of computer science that aims to create intelligent machines that "oppose" the natural intelligence of living beings. The term has changed significantly over time and now encompasses the entire idea of enabling machines to imitate and surpass human cognitive abilities.

Machine learning is at the heart of artificial intelligence, and its applications range from search engines and character recognition to online searches.

Machine learning is an application of artificial intelligence that allows systems to learn and improve from experience without being explicitly programmed.

Machine learning focuses on developing computer programs capable of accessing data and using it to learn autonomously.

Machine learning can be used to analyze the vast amounts of information readily available in the world to aid in decision-making [26].

3.2 Types of machine learning models:

Traditional machine learning is often classified based on how an algorithm learns to become more accurate in its predictions. There are four fundamental approaches: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. The type of algorithm data scientists choose to use depends on the type of data they aim to predict [26].

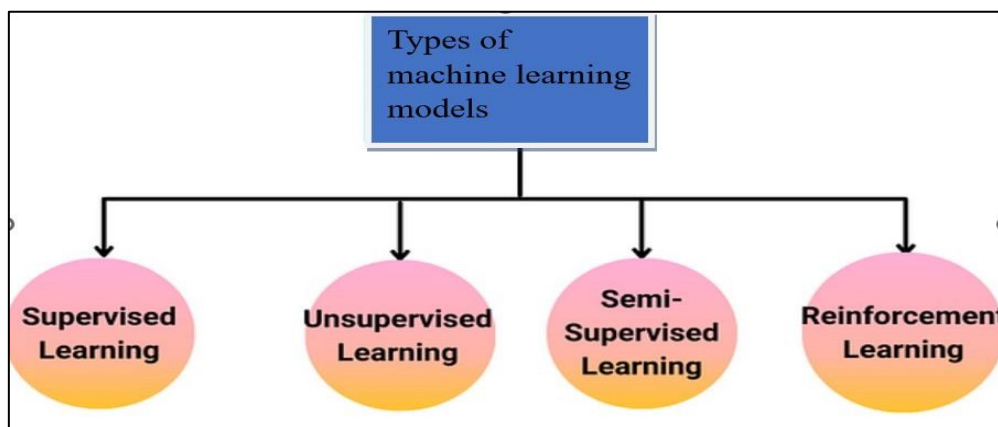


Figure 16: Types of machine learning models

3.2.1 Supervised Learning:

Supervised learning algorithms are suited to a variety of tasks. They enable the classification of data into two distinct categories in the context of binary classification, as well as selecting from multiple classes in the case of multi-class classification. Additionally, they are used for regression modeling, where the goal is to predict continuous values rather than categories. Finally, ensemble techniques utilize multiple machine learning models to combine their predictions and produce more accurate results. These capabilities make supervised learning algorithms essential tools in solving many complex problems in data science and artificial intelligence [26].

3.2.2 Unsupervised Learning:

Unsupervised learning algorithms are suitable for several key tasks. They are used for clustering, which involves dividing a dataset into groups based on the similarity of elements. Additionally, they enable anomaly detection by identifying unusual data points within a dataset. They also facilitate association exploration, finding sets of items that frequently occur together in the data. Finally, they are employed for dimensionality reduction, which involves decreasing the number of variables in a dataset while retaining as much relevant information as possible [26].

3.2.3 Semi-supervised Learning:

Semi-supervised learning is used in various domains such as machine translation, where algorithms are trained to translate a language with an incomplete dictionary of words, fraud detection, where it helps identify fraud cases based on a limited number of positive examples, and data labeling, where algorithms trained on small datasets learn to automatically apply labels to larger datasets [26].

3.2.4 Reinforcement Learning:

Reinforcement learning is widely applied in various fields such as robotics, where it enables robots to learn to perform tasks in complex physical environments, in video games, where it is used to train agents to play various games effectively, and in resource management, where it helps businesses optimize the allocation of limited resources based on specific objectives [26].

3.3 Regression:

There are various regression models. Each of these models can be used to predict a dependent variable based on one or more independent variables [22].

3.3.1 Linear regression:

The purpose of multiple linear regression is to describe changes in a dependent variable (y) compared to changes in several independent variables. In machine learning, it is used to estimate a linear function between Output (with continuous numerical values) and Input [22].

3.3.2 Polynomial regression:

Polynomial regression is a particular case of linear regression. New features can be created during the data preparation step by multiplying the values of the old features [22].

3.3.3 Logistic regression:

Logistic regression is a statistical model that studies the relationship between a set of qualitative variables X_i and qualitative variables Y . It is a generalized linear model using a logistical function as a link function.

Logistical regression models can also predict the probability of an event occurring (a value of 1) or not (a value of 0) based on the optimization of regression coefficients. The result always varies between 0 and 1. When the predicted value is above the threshold, the event may occur, but when the value is below the same threshold, it will not occur [22].

3.4 Classification:

In machine learning, classification is the concept of supervised learning, which basically consists of classifying a data set. It can be a binary classification problem or a multi-class problem.

There are many types of classification tasks that can be encountered in machine learning, each of which can use specialized modeling methods. The most common ones are voice recognition, face detection, handwritten writing, document classification, etc [22].

- There may be four main types of classification tasks:

3.4.1 Binary classification:

Binary classification refers to a classification task with two class labels. Generally, binary classification tasks involve one category being a normal state and the other category being an abnormal state. The class in the normal state is labeled as class 0, and the class in the abnormal state is labeled as class 1. It is common to model binary classification tasks using a model that predicts a Bernoulli probability distribution for each example [22].

Popular models used for binary classification include:

- Logistic regression.
- k-Nearest Neighbors.
- Decision trees.
- Support Vector Machines.
- Naive Bayes.

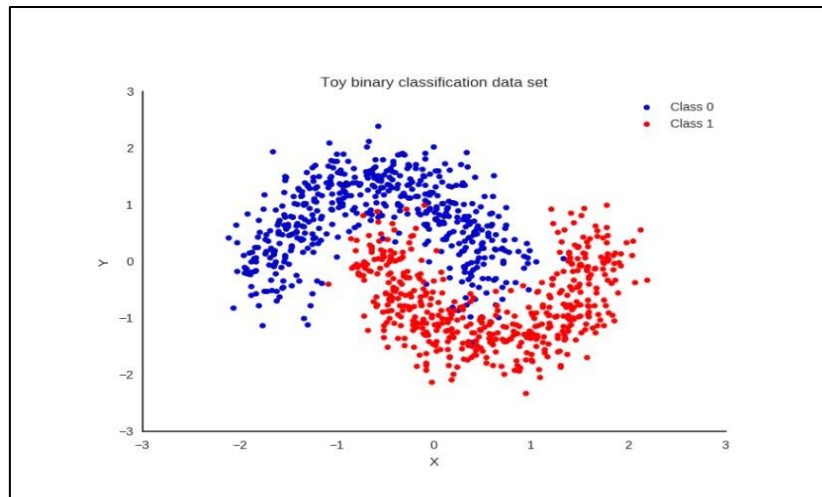


Figure 17: Binary ranking

3.4.2 multi-class classification:

Multi-class classification refers to classification tasks that involve more than two class labels. Unlike binary classification, there is no distinction between normal and abnormal results. Instead, each example is assigned to one of the several known classes. Typically, a multi-class classification model predicts a multinomial probability distribution for each example [22].

Among the models commonly used for multi-class classification are:

- k-Nearest Neighbors (k-voisins les plus proches)
- Decision trees
- Naive Bayes
- Random Forest (random forest)

Binary classification models that can be adapted to a multi-class classification include:

- Logistic regression
- SVM (Support Vector Machines)

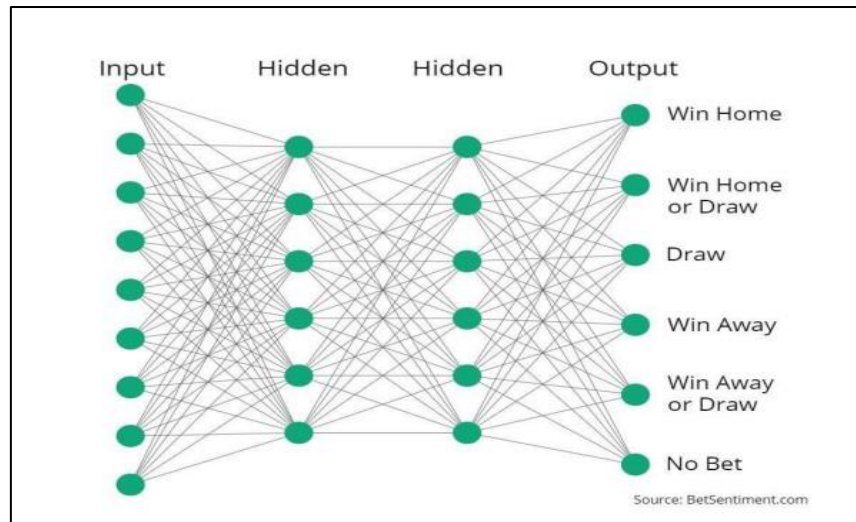


Figure 18: Multi-class classification

3.4.3 multi-label classification:

Multi-label classification refers to classification tasks where each example can be associated with two or more class labels, allowing for the prediction of one or multiple class labels for each example.

Models used for binary or multi-class classification cannot be directly applied to multi-label classification. Instead, specialized versions of standard classification algorithms, known as multi-label versions of algorithms, are used. These include multi-label decision trees, multi-label random forests, and other methods tailored to predicting multiple labels for each example [22].

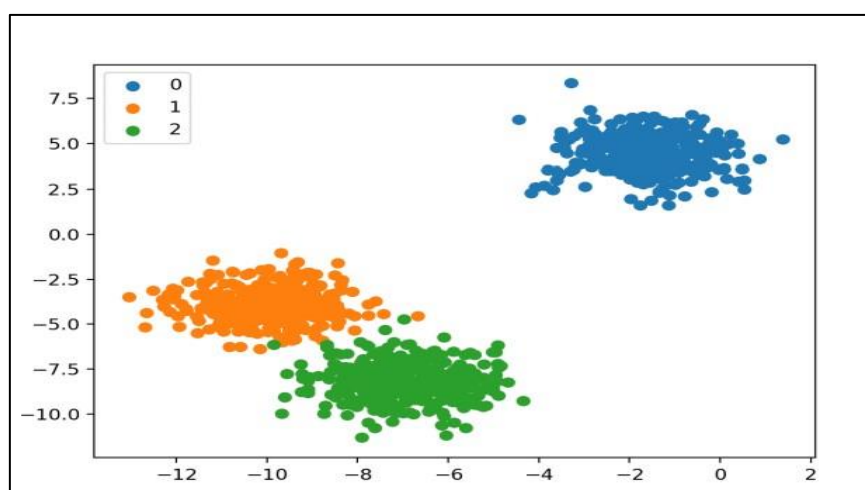


Figure 19: Multi-label classification

3.4.4 Imbalanced classification:

Generally, in unbalanced classification tasks, binary classification problems are found where most examples in the learning data set belong to the normal class, while a minority of examples belong in the abnormal class [22].

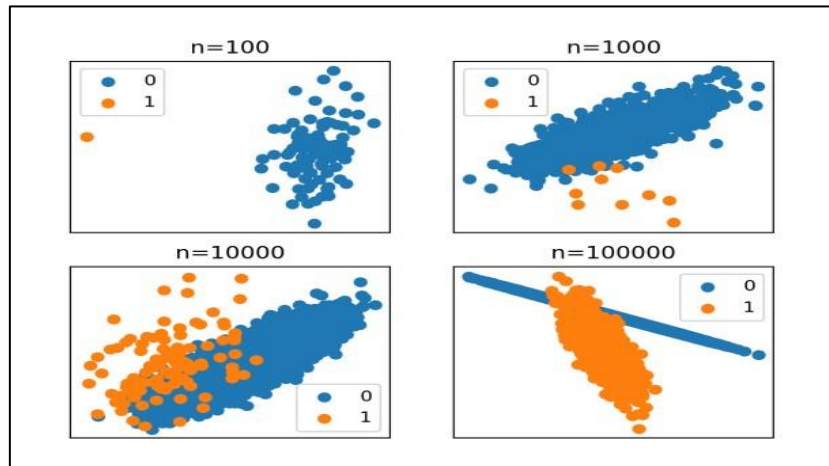


Figure 20: Imbalanced classification

3.5 supervised learning algorithms used:

3.5.1 K-Nearest Neighbors (KNN):

K-Nearest Neighbors (KNN) is one of the simplest supervised learning methods, used to solve classification and regression problems. Its principle of operation involves classifying new data points based on their similarity to neighboring data points.

- KNN is an algorithm that makes no assumptions about the structure or distribution of data, making it a non-parametric.
- It is also called a lazy learner's algorithm, because it does not immediately practice with the learning set, but instead stores the set of data and performs an action on these data at the time of classification.
- KNN works by classifying or predicting on the basis of a fixed number (K) of data points closest to the input point. This means that for a chosen K value, an input point will be classified or assigned to the same class as the majority of the nearest adjacent K points[27].

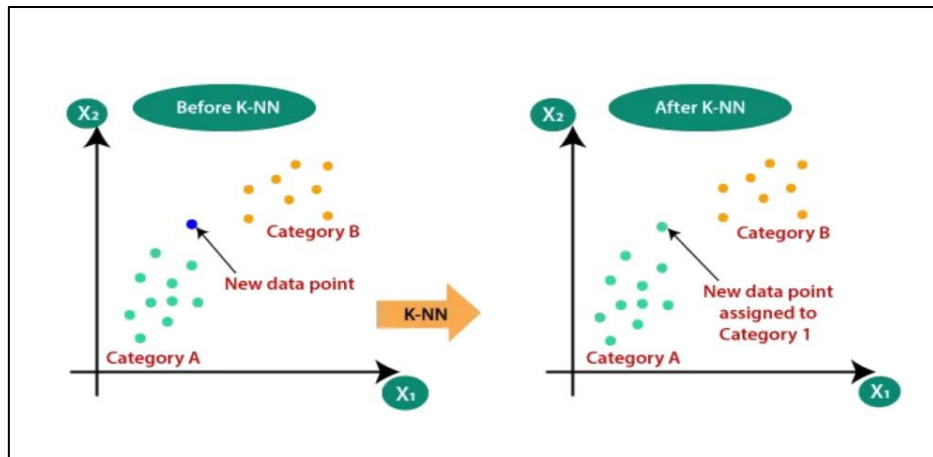


Figure 21: K-Nearest Neighbors (KNN)

3.5.2 Decision Trees:

Decision Trees are one of the most widely used and practical supervised learning algorithms, suitable for solving various types of problems including classification and regression:

- A decision tree is a tree-like structure resembling an organizational chart where each internal node represents a feature (or attribute), each branch represents a decision rule, and each leaf node represents an outcome. This structure aids in decision-making.
- It is a non-parametric algorithm, which means there are no underlying assumptions about the distribution of the data [27].

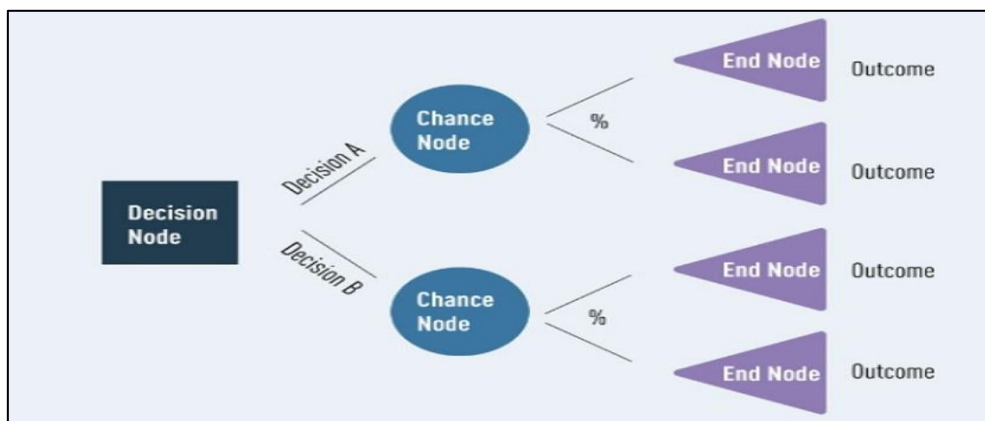


Figure 22: Decision Trees

3.5.3 Random Forest:

Random Forest is a very popular supervised learning algorithm, widely used for regression and classification. It uses a set of learning algorithms, a process that combines multiple algorithms to solve a complex problem and improve the performance of the model. It is an algorithm that creates many decision trees (hence the name "forest") from various sub-sets of the data set. It aggregates the predictions of each tree and, based on the majority votes of the forecasts, produces the final result [27].

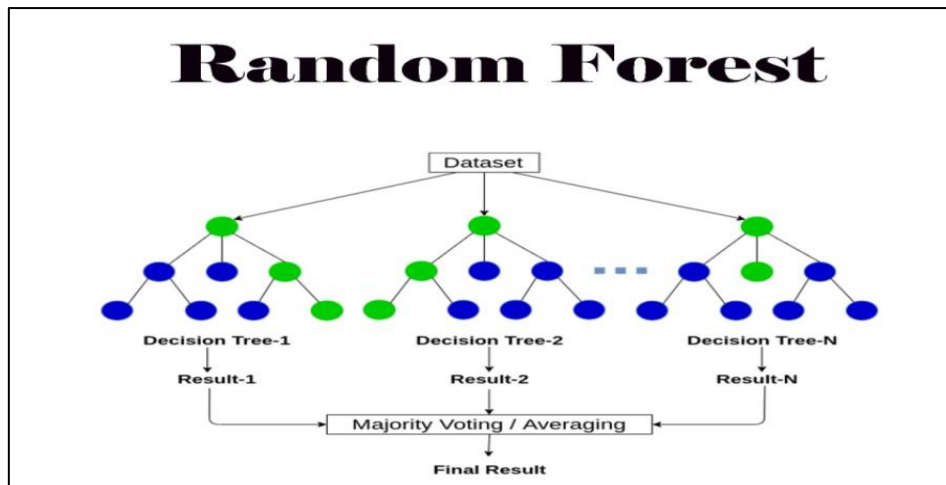


Figure 23: Random Forest

3.5.4 Support Vector Machine (SVM):

The Support Vector Machine, or SVM, is one of the most popular supervised learning algorithms, primarily used for classification in machine learning. Its goal is to create the best line or decision boundary, known as a hyperplane, that can effectively separate n-dimensional space into different classes. This optimal decision boundary allows for accurate classification of new data points in the future.

SVM selects extreme points or vectors that help define the hyperplane. These extreme cases are called support vectors; hence the algorithm is named Support Vector Machine[27].

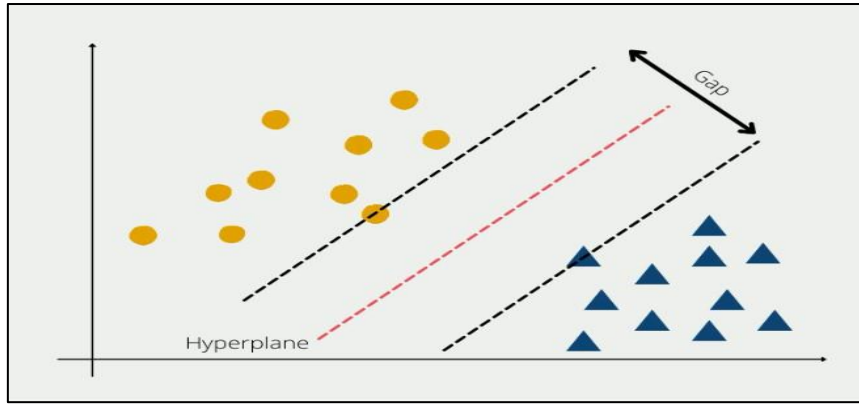


Figure 24: Support Vector Machine (SVM)

3.5.5 Naïve Bayes:

The Naive Bayes is a supervised learning algorithm commonly used for classification. It is a simple probabilistic classifier based on the Bayes theorem, enabling you to build quick models and make effective predictions. The algorithm is based on a simplifying hypothesis, called "naive", which assumes that all characteristics are independent of each other. This hypothesis reduces the complexity of the model and accelerates the calculation of probability. The Naive Bayes is appreciated for its simplicity and speed, although it may be less efficient when the characteristics are correlated. Despite this limitation, it remains very effective in many contexts where speed and simplicity are priority[27].

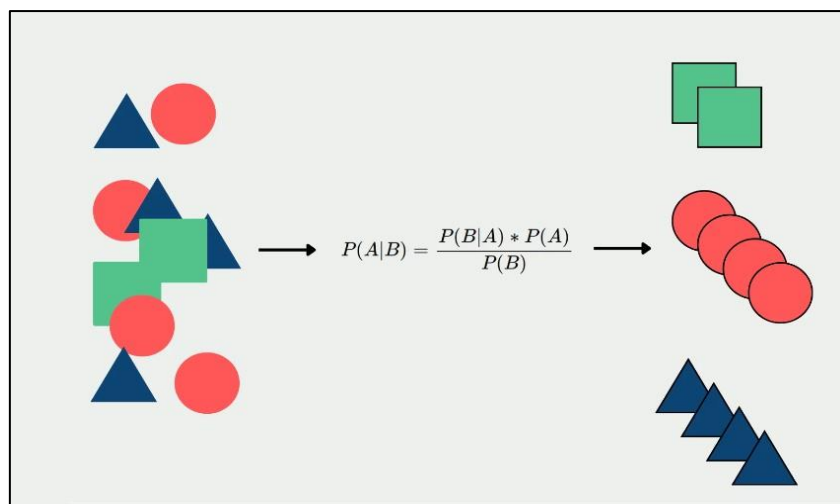


Figure 25: Naïve Bayes

3.5.6 Logistic Regression:

The logistic model is a statistical model that uses a logistic function and is also used in classification problems. Logistic regression does not require a linear relationship between dependent and independent variables, but it does require large sample sizes to improve accuracy in likelihood estimation.

Logistic regression can be binary or multinomial. Binary logistic regression applies to situations where the observed outcome for a dependent variable can have only two possible types, labeled as "0" and "1" [28].

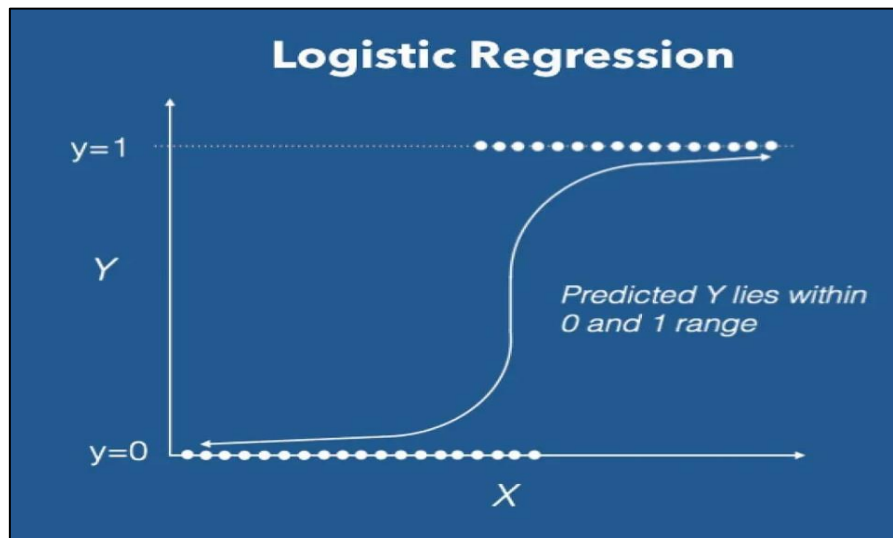


Figure 26: Logistic Regression

3.5.7 Gradient Boosting:

Gradient Boosting is a widely used boosting algorithm in machine learning for classification and regression tasks. It works by sequentially combining multiple weak models to form a strong model. At each step, the model aims to correct the errors of the previous model by minimizing a loss function, such as mean squared error for regressions or cross-entropy for classifications, using gradient descent [29].

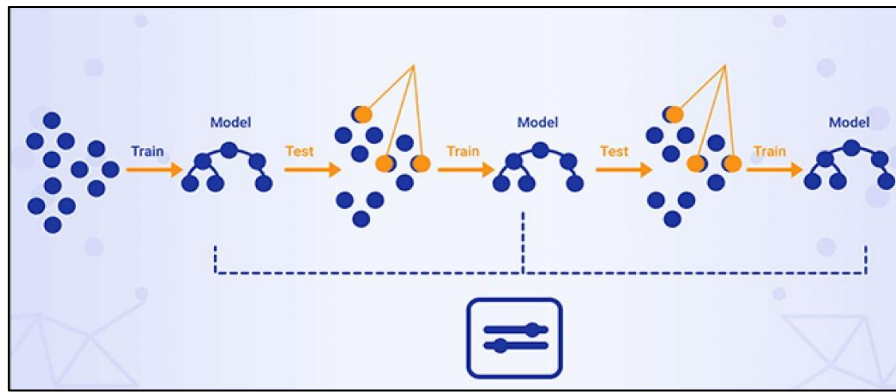


Figure 27: Gradient Boosting

3.6 the machine learning process:

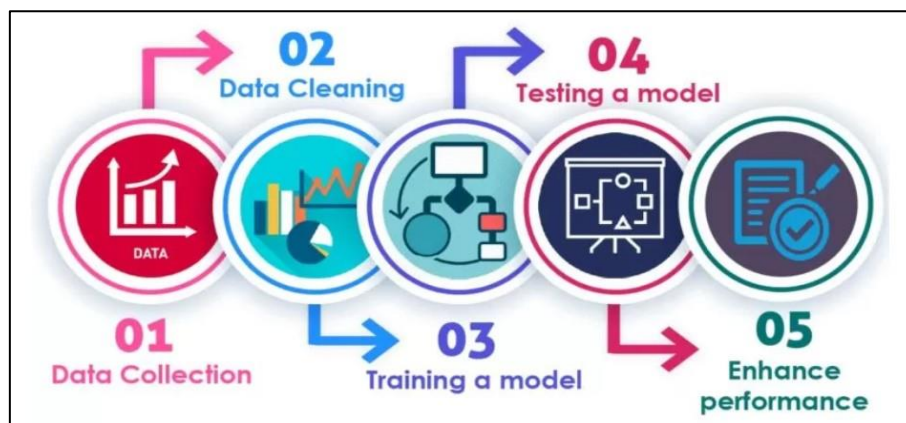


Figure 28: the machine learning process

3.6.1 Data Collection:

Data collection for a water quality dataset involves obtaining precise and consistent measurements of various physical, chemical, and biological parameters of water from different sampling sources. These parameters include air temperature ($T(\text{air})$ °C), water temperature ($T(\text{water})$ °C), pH, conductivity (Cond $\mu\text{s}/\text{cm}$), turbidity (turb FTU), suspended solids (MES mg/l), chemical oxygen demand (COD mg/l), biological oxygen demand (BOD mg/l), ammonium nitrogen (N-NH_4^+ mg/l), nitrates (N-NO_3^- mg/l), nitrites (N-NO_2^- mg/l), phosphates (PO_4^{3-} mg/l), and dissolved oxygen (O_2). Each row in the dataset represents a unique observation of the water conditions at a specific time and location, totaling 678 observations.

	Dates	T(air) c°	T(eau) c°	ph	cond µs/cm	turb(FTU)	MES(mg/l)	DCO(mg/l)	DBO(mg/l)	N-NH4+ (mg/l)	N-NO3- (mg/l)	N- NO2(mg/l)	PO4-3 (mg/l)	O2	EAU
673	Mars 2024 2eme mesure	NaN	18.0	7.83	2100.0	30.0	47.0	79.0	28.0	5.50	6.700	0.170	15.60	6.24	1
674	Avril 2024 1er mesure	20.0	18.0	7.88	2043.0	216.0	468.0	472.0	220.0	47.21	0.325	0.255	25.60	0.51	0
675	Avril 2024 1er mesure	NaN	18.0	7.82	2143.0	30.0	49.0	95.0	30.0	6.65	5.670	0.077	16.70	4.34	1
676	Avril 2024 2eme mesure	23.0	21.0	8.13	1400.0	233.0	432.0	352.0	150.0	47.21	0.421	0.219	22.33	0.85	0
677	Avril 2024 2eme mesure	NaN	20.0	8.32	2220.0	11.0	19.0	80.0	13.0	6.39	4.650	0.090	10.76	3.28	1

Figure 29: dataset of water quality

3.6.2 Data cleaning:

Data cleaning for this water quality dataset involves preparing and correcting the data to ensure its accuracy, consistency, and usability for analysis. This process includes managing missing values (NaN) through imputation or deletion, detecting and treating outliers to minimize their impact, and verifying consistency to ensure values fall within plausible ranges and expected data types. It also encompasses removing duplicates to ensure each observation is unique, correcting errors and inconsistencies, and normalizing or standardizing data to make parameters comparable. By cleaning the data in this way, we obtain a precise and reliable dataset ready for analysis or machine learning tasks.

	T(air) c°	T(eau) c°	ph	cond µs/cm	turb(FTU)	MES(mg/l)	DCO(mg/l)	DBO(mg/l)	N-NH4+(mg/l)	N-NO3-(mg/l)	N-NO2(mg/l)	PO4-3(mg/l)	O2	EAU
80	11.0	10.0	7.04	189.0	156.0	180.0	456.0	220.0	32.25	1.07	0.31	3.09	0.83	0
84	9.0	8.0	7.48	217.0	147.0	180.0	450.0	220.0	42.00	0.72	0.10	2.83	3.46	0
88	12.0	8.0	7.00	254.0	176.0	214.0	340.0	190.0	24.25	0.65	0.22	2.79	1.31	0
92	16.0	14.0	7.75	203.0	119.0	190.0	344.0	260.0	47.85	0.58	0.13	6.40	0.81	0
96	16.0	7.0	7.02	202.0	243.0	238.0	350.0	240.0	41.02	0.68	0.14	3.82	1.69	0

Figure 30: dataset after data cleaning

3.6.3 Training a model:

Next, we trained various machine learning models such as Decision Trees, Random Forests, Logistic Regression, Support Vector Machine (SVM), K Nearest Neighbors (KNN), Gradient Boosting Classifier, and Gaussian Naïve Bayes on the training dataset. This involved feeding the training data into each individual classifier and allowing them to learn the patterns and relationships between the features and the target variable.

3.6.4 Testing a model:

Testing a model for this water quality dataset involves evaluating how machine learning algorithms generalize on unseen new data. This includes using a test dataset to measure the

model's ability to accurately predict the values of the target variable (e.g., water quality) based on the features provided in the dataset. Common evaluation metrics include accuracy, precision, recall, and F1 score for classification models.

3.6.5 Enhance performance:

Enhancing performance in the context of machine learning refers to optimizing and increasing a model's capacity to perform a specific task with improved accuracy, greater robustness, and increased efficiency.

- **Accuracy:** The overall accuracy score on the test set indicates the proportion of correct predictions made by the model.

(1)

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

- **Precision:** This metric measures the proportion of true positives among the positive predictions made by the model for each class. And it's determined by dividing the number of true positive samples by the total number of positive samples predicted by the model:

(2)

$$Precision = \frac{(TP)}{(TP + FP)}$$

- **Recall:** this metric measure the proportion of actual positives that the model correctly identified for each class and it is calculated by dividing the number of correctly recommended or predicted crops or fertilizers by the total number of crops or fertilizers. It can be defined as:

(3)

$$Recall = \frac{(TP)}{(TP + FN)}$$

- **F1-score:** The F1-score is the harmonic mean of precision and recall, providing a balanced measure of the model's performance for each class.

(4)

$$F1 - score = \frac{(2 \times TP)}{(2 \times TP + FP + FN)}$$

True Positive (TP): represents the number of abnormal samples correctly classified as exacerbation (correct detection)

False Positive (FP): represents the number of normal samples wrongly classified as attacks (incorrect detection)

True Negative (TN): represents the number of normal samples wrongly classified (correct detection)

False Negative (FN): represents the number of abnormal samples wrongly classified as attacks (incorrect detection)

4 Result and discussion:

Table 7: A comparative study between algorithms

Algorithms	Precision	Recall	F1-score
Random Forest	0.994950	0.994898	0.994898
Decision tree	0.994949	0.994898	0.994898
KNN	0.979788	0.979592	0.979588
Gradient Boosting	0.994950	0.994898	0.994898
Logistic regression	0.989796	0.989796	0.989796
SVM	0.962235	0.959184	0.959098
Gaussian naïve Bayes	0.975740	0.974490	0.974480

After analyzing the results, it is evident that the Random Forests have shown exceptional performance with precision, recall, and F-score all around 99.5%. This model uses an ensemble of decision trees to improve robustness and reduce overfitting, resulting in high generalization ability and precise prediction of water quality based on measured parameters.

Logistic Regression has demonstrated strong performance with precision, recall, and F-score all close to 98.98%. This linear model is well-suited for data where relationships between

features and the target are linear, providing clear interpretation of coefficients while maintaining good prediction performance.

On the other hand, SVM achieved precision of 96.22%, recall of 95.92%, and an F-score of 95.91%. This model is effective in finding an optimal hyperplane to separate classes in high-dimensional space, though it requires careful hyperparameter tuning to avoid overfitting.

KNN showed solid performance with precision, recall, and F-score around 97.98%. This model relies on similarity between neighboring instances to classify new examples, offering a flexible non-parametric approach that can be sensitive to dimensionality and choice of K value.

However, Decision Trees achieved precision, recall, and F-score all around 99.49%, demonstrating their effectiveness in dividing data into homogeneous subgroups based on the most discriminative features. However, they can be prone to overfitting if tree depth is not regulated.

Gradient Boosting also achieved high performance with precision, recall, and F-score all close to 99.5%. This model sequentially builds trees, correcting errors of previous models, which allows it to progressively improve predictions and achieve competitive results.

Gaussian Naive Bayes showed respectable performance with precision of 97.57%, recall of 97.45%, and an F-score of 97.45%. This model assumes conditional independence between features, making it quick to train and suitable for data where this assumption holds reasonably well.

In conclusion, each algorithm exhibits strengths and nuances in its ability to model and predict water quality based on measured parameters. The choice of the optimal model depends on specific project goals, data characteristics, and trade-offs between precision, interpretability, and available computational resources.

5 Predictions and Evaluation:

In this final step, we selected the best-performing model (Random Forest) and retrained it on the entire dataset, combining both the training and testing sets. We then used this model to make predictions on new data instances.

Conclusion:

In conclusion, the design and implementation phase of this project allowed for the development of an innovative system combining the Internet of Things (IoT) and machine learning to optimize the use of wastewater in irrigation. This chapter detailed the various design stages, technological choices, and practical implementation of the system.

Chapter 4

Internship Report

Introduction:

As part of the requirements for my Master's degree in Instrumentation and Metrology at institute of applied science and technology Ahmed Ben Bella University Oran 1, I had the opportunity to complete a month internship at the wastewater treatment plant d'Ain Temouchent. This internship took place from 3rd of March to April 3 and aimed primarily to provide me with practical experience in instrumentation and metrology applied to the management and treatment of wastewater.

The objective of this work is to understand the functioning principle of depuration unit. Indeed, I did an internship to get the parameters needed for my project.

This work is divided into three parts, the first part of which is devoted to general information on wastewater and sanitation systems. The second part is his summary at the station presentation of Ain Temouchent. In the third part, I explained the stages of the wastewater treatment of this station. At the end of this work, I made a general conclusion that summarizes most of my work.

1 General information on wastewater:

Wastewater is water that has been affected by various human activities, including solid, liquid, or gaseous substances. It primarily originates from domestic, industrial, agricultural, and urban uses.

Definition:

Wastewater is a heterogeneous mixture of organic matter and toxic minerals. Wastewater is an unclean liquid mixture that needs to undergo a series of treatments before it can be purified. To

this end, a sanitation system is in place so that this water can be collected before it is discharged into the environment.

Pollution of water consists of adverse or harmful changes in physical, chemical and biological properties caused directly or indirectly by human activities, making it unsuitable for normal use and requiring treatment [30].

1.1 Wastewater Types:

Depending on the origin of wastewater, four major categories are distinguished:

- Domestic wastewater.
- Industrial wastewater.
- Rainwater.
- Agricultural water.

1.1.1 Domestic wastewater:

Domestic wastewater consists of water from the discharge valves of: gray water in toilets, kitchens and bathrooms.

The waste present in these waters is contaminated with degradable organic matter and minerals. These substances are either dissolved or in suspension. The sanitation network leads to treatment plants where the water is treated [30].

1.1.2 Industrial wastewater:

All discharges other than those from domestic water are considered industrial discharges. Industrial wastewater generally has a more specific composition and is directly related to the type of industry involved. Regardless of the load of organic or mineral contamination, and whether or not they have putrefactive properties, they present their own toxicity profile associated with the chemicals transported [30].

1.1.3 Rainwater:

These include drainage waters (rainwater, water for watering public roads, water to rinse the taps, markets and courts). The water that flows from roofs, courts, gardens, green spaces, public roads and markets carries a variety of mineral and organic waste: Soils, lemons, plant waste... and various micro-pollutants (hydrocarbons, pesticides, detergents...) [30].

1.1.4 Agricultural water:

These are waters contaminated with substances used in the agricultural sector. In the context of efficient and intensive agriculture, the farmer is forced to use various products of industrial or agricultural origin, some of which present or may present hazards to the environment, including water quality.

These mainly include:

- Mineral fertilizers (commercial products or animal feces produced or not produced on the farm);
- Phytosanitary products (herbicides, fungicides, insecticides,...) [30].

2.1 Characteristics of wastewater:

2.1.1 Physical Characteristics:

- **Temperature :**

Temperature is an important ecological factor for physico-chemical and biological reactions, often overlooked in urban collectors, but should be most measured, for industrial discharges into the grid, and also for the metabolism of growth of microorganisms in water[30].

- **Conductivity :**

Conductivity is a parameter that measures the ability of water to transmit current between two electrodes. In water, the majority of dissolved substances are electrically charged ions, which makes it possible to determine the number of salts solved in water [30].

- **Turbidity :**

This criterion defines the transparency of a liquid due to the presence of unsolved substances such as clay and microorganisms. The more or less significant presence of insoluble suspension material in waste water, whether mineral or organic, is responsible for the turbidity [30].

- **Suspended and decantable Materials :**

The amount of organic and mineral pollution in water is represented by suspended matter (MES), which are suspended and colloidal substances. These substances are non-reactive for two hours and are measured in cm³/l. Impurities in water come from minerals, plants, or animals. These suspended substances are significant, moving without physical or chemical treatment [30].

2.1.2 Chemical Characteristics:

- **Hydrogen potential (pH):**

The pH is an indicator that evaluates the acidity and alkalinity of wastewater. It influences physico-chemical characteristics (acidity, aggressiveness) and biological processes, some of which require very strict pH values ranging from 6.5 to 8.5 [30].

- **Dissolved oxygen :**

The concentration of dissolved oxygen is an important parameter for organic matter degradation and photosynthesis. Heavily airy water is usually saturated with oxygen, while water containing organic matter that can be degraded by micro-organisms is sub-saturated with oxygen [30].

- **Biological oxygen demand (DBO5) :**

Indicates the amount of oxygen required by environmental microorganisms to destroy or degrade the organic material present in wastewater. This parameter is measured by the oxygen consumption for 5 days in the dark at 20°C, which is the time required to ensure the biological oxidation of carbonated organic matter, expressed in mg O₂/l [30].

- **Chemical oxygen demand (DCO) :**

The amount of oxygen consumed by the oxidable substances present in a liter of water is called the chemical demand for Oxygen, expressed in mg of O₂/l [30].

- **Nitrogen :**

Nitrogen in household sewage exists in the form of organic matter and ammonia and is measured by measuring N-NTK (Total Nitrogen Kjeldahl) and N-NH₄.

Organic nitrogen is the main component of proteins and is constantly recycled by animals and plants. Ammonia nitrogen is found in two forms: ammonia NH₃ and ammonium NH₄⁺, whose relative proportions depend on pH and temperature [30].

- **Nitrites (NO₂):**

Nitrite (NO₂⁻) ions are an intermediate step between ammonium (NH₄⁺) and nitrate ions (NO₃⁻). Nitrifying bacteria (Nitrosomonas) convert ammonium to nitrite. This nitrification operation requires the consumption of a large amount of oxygen. Nitrite is derived from the bacterial reduction of nitrate, called denitrification. Nitrite is a dangerous poison for aquatic life even at very low concentrations, the toxicity increases with temperature [30].

- **Nitrates (NO₃-) :**

Nitrates are the final stage of the oxidation of organic nitrogen in water. Nitrate bacteria (nitrifying bacteria) convert nitrite to nitrate. Nitrates are not toxic, but high concentrations of nitrates can cause the proliferation of algae, leading to eutrophization of the environment [30].

2.1.3 Microbiological characteristics:

Determination of total mesophilic aerobic bacteria, total coliformes, fecal coliforms, Staphylococcus, Streptococcus and Salmonella and Shigella, as well as certain pathogens, may indicate the risks associated with the use of certain types of water [30].

3.1 sanitation systems:

Urban sanitation aims to ensure that waste water is collected, transported and purified before discharging into the receiving environment. These are collective sanitation plants located in areas connected to the sewage system and equipped with municipal waste treatment stations. It was clarified that in areas not connected to the sewerage system, sewage systems are neither collective nor autonomous.

In the nineteenth century, according to Charles de Freycinet, sanitation was defined as "the restitution to the soil of the fertilizing elements that they contain and the return to the rivers of the liquids deprived of their corrupting elements."

As cities grew, expanded, and disconnected from their countryside, these challenges gradually disappeared. There have long been methods of ecological sanitation that focus on the preservation of aquatic ecosystems and the enhancement of soil nutrients, but traditional sanitation remains the most widespread[31].

3.1.1 Types of sanitation:

There are three types of sanitation for individuals: collective sanitation (sewer system), individual sanitation, also known as non-collective sanitation, and semi-collective sanitation.

Collective sanitation:

Collective sanitation consists of large pipelines that collect wastewater and transport it by gravity to a treatment plant, sometimes, lifting stations are necessary. It is referred to as collective sanitation when the wastewater from multiple houses or buildings is collected by the public sanitation network and then transported to a treatment plant[31].

non-collective sanitation:

Today, autonomous sanitation, also known as non-collective sanitation (NCS), is carried out with a full-water pit in municipalities that do not have a common sanitation system[31].

Semi-collective sanitation:

Semi-collective sanitation systems, also called small diameter systems, are a technology for the evacuation of feces.

Domestic water first enters a waterproof septic pit, where organic compounds undergo anaerobic fermentation, turning them into mud, and where solid particles are deposited under the action of gravity.

The resulting filter contains no solid waste and is thrown into an occasional sewer network to reach the sewer.

However, due to the reduction in the diameter of the pipelines, the accidental introduction of solid particles into the pipeline network can lead to the blockage of the Pipelines and the failure of the entire installed system[31].

2 presentation of the wastewater treatment plant:

Effective wastewater management is essential in order to preserve the environment and public health. In this process, treatment plants play a key role in treating waste water to remove pollutants before they are discharged into the natural environment. The Ain Temouchent wastewater treatment plant is presented in these lines, highlighting its different treatment stages and its positive impact on the community and the surrounding environment.

2.1 Internship location:

The internship is held at the level of the wastewater treatment station of the wilaya of Ain Temouchent, is an establishment opened its doors. The head of this laboratory is Mr. Ghermaoui Fethi.



Figure 31: location of the station with google maps

2.2 The opening of the station:

The HYDRO-TREATMENT/COMSA Group is building a wastewater treatment plant in the town of Ain Temouchent. The tender was made on the basis of a tender launched by the Hydraulic Office of the Wilaya of Ain Temouchent. The treatment plant will treat all the sewage in the town of Ain Temouchent. The chosen solution for wastewater treatment is a low load biological solution using activated sludge. Treatment of phosphorus removal is also implemented in three levels. The design of the treatment plant is based on 2015 data, and an extension is planned for 2030.

2.3 BASIC DATA:

The treatment station will be completed by the horizon of 2015, it will support the sewage of 72,800 inhabitants, for an input flow of 10.920 m³/d which corresponds to 82,000 equivalent inhabitants, with the possibility of extension for the horizon of 2030 to: 119,000 equivalents inhabitants and a flow of: 13.500 m³ /d.

- Values at processing entry :

Table 8: Values at processing entry

PARAMETRES	UNITES	2015	2030
Number of inhabitants		72.800	90.000
Number of equivalent inhabitants	E.H	82.000	119.000
Average wastewater flow	M3 /d	10.920	13.500
Rainy time peak flow	M 3 /h	1.365	1.687,5
Daily D.B.O.5	Kg/d	3.640	4.500
Daily D.C.O.	Kg/d	6.552	8.100
M.E.S Daily	Kg/d	873	1080
Daily Phosphores	Kg/d	436	540

- Dismissal Limit Values :

Table 9: Dismissal Limit Values

PARAMETRES	UNITES	Value
MES	Mg/l	30
DBO5	Mg/l	30
DCO	Mg/l	90
NTK	Mg/l	10
Phosphore	Mg/l	2

2.4 Justification of the adopted solution:

According to the specifications outlined in the tender document for the construction of the wastewater treatment plant in the city of Ain Temouchent, the following treatment facility has been developed:

- Water Line :

The pre-treatment of water consists of a grove well, a rescue grilling, a rough grinding, a fine grilling and a dew-de-oiling. The removal of carbon and nitrogen pollution is carried out by biological means. Phosphorus is removed physico-chemically in a clarifloculator through coagulation with ferric chloride (FeCl_3) and flocculation with an anionic polyelectrolyte.

The oils and fats as well as the refusal of grilling will be evacuated by tank outside the station.

- Sludge Line :

The biological sludge in the clarifier output is in a stable state with regard to the age of the clarifier (sludge age approximately 19 days) and is then sent to the thickener.

After concentration, the dosing of the polyelectrolyte is carried out before the dehydration which is ensured by two pressure carpets. A series of emergency drying beds is provided when the two rolling carpets are closed.

3 stages of treatment of waste water in the station:

Wastewater treatment is an essential step aimed at purifying contaminated water before returning it to nature. As human activity increasingly impacts water resources, sewage treatment has become essential to sustainable water resources management. This complex process requires a series of carefully planned steps, each of which plays a crucial role in eliminating pollutants and creating clean and safe water. In this situation, it is crucial to understand the stages of wastewater treatment, from initial pre-treatment to final disinfection, in order to fully grasp the importance of this practice for public health and the preservation of the environment.

I have examined in depth each step of wastewater treatment in the following paragraphs, highlighting the processes and technologies used to ensure the effectiveness and sustainability of this crucial enterprise.

3.1 Pre-treatment:

The purpose of pre-treatment is to separate the rough materials and elements that may interfere with and alter the subsequent processing processes, and by mechanical operations and Physical methods: grilling, Sand Removal and Grease and Oil Removal.

3.1.1Grilling:

It consists of passing the sewage through a grid whose grids are more or less spaced, retaining the roughest elements. After cleaning the grid by mechanical, manual or automatic means, the waste is disposed of with the household garbage.



Figure 32: Rough and manual grid

3.1.2 Sand Removal:

Sand removal eliminates sand from wastewater to prevent its subsequent sedimentation. When water flows at a reduced speed in a basin known as a "grit chamber," the sand particles settle at the bottom. These particles are then suctioned by a pump. The recovered sand is drained and washed before being either sent to a landfill or reused. Sand removal specifically targets mineral particles with a diameter greater than 0.2 mm.



Figure 33: Sand Classifier

Sand removal is a crucial process in water and wastewater treatment plants to ensure the protection of equipment and to improve treatment efficiency. This process is a fundamental step that helps reduce costs and increase the overall effectiveness of the treatment system.



Figure 34: Sewage Treatment Plant Sands

3.1.3 Grease and Oil Removal:

Grease and oil removal aims to eliminate fats and oils from wastewater, which can hinder the effectiveness of subsequent biological treatments. This operation is conducted through flotation: air is injected at the bottom of the structure to bring fats to the surface. The grease and oils are then skimmed off the surface, temporarily stored, and subsequently disposed of (either by landfilling or incineration).



Figure 35: Grit and Grease Separator

3.2 Primary treatments:

Despite pretreatment, the effluents still contain dissolved pollutants and suspended solids (TSS), so primary treatment is necessary, which involves physical or physicochemical processes.

3.2.1 Natural sedimentation:

The suspended solids settle to the bottom of the basin and are mechanically concentrated in pits located under the baffles. The material resulting from primary sedimentation forms primary sludge. It is extracted by pumping into the sludge treatment system. Clarified water is collected at the surface through troughs and directed to biological treatment.



Figure 36: Primary clarifier

3.2.2 Physico-chemical sedimentation:

Physico-chemical sedimentation is used to reduce water turbidity and color, primarily caused by very small colloidal particles that can remain suspended for long periods, even passing through very fine filters. This process involves two main stages: coagulation and flocculation. Coagulation aims to destabilize suspended particles by promoting their aggregation using cationic mineral salts such as iron or aluminum salts, which are rapidly dispersed in the water. Subsequently, flocculation involves promoting interactions between destabilized particles through slow mixing, allowing larger aggregates (flocs) to form that settle out easily during the sedimentation process.

3.3 Secondary treatment (Biological):

Wastewater biological treatment is a process that utilizes bacteria and microorganisms to break down organic waste and pollutants present in wastewater. The goal of biological treatment is to establish a system where the decomposition products are easily collected for proper disposal.

3.3.1 Anaerobic treatments:

Anaerobic treatments involve bacteria that do not utilize oxygen, particularly methane-producing bacteria which, as their name suggests, generate methane from organic matter, along with some CO₂. This process is delicate and requires careful monitoring. Maintaining a stable and sufficiently high temperature is crucial. It's also important to avoid sudden pH fluctuations and substances that inhibit bacterial growth, such as cyanides, heavy metal salts, and phenols.

3.3.2 Aerobic treatments:

Aerobic treatment is a crucial step in the biological treatment of wastewater, aiming to degrade organic matter through the activity of aerobic microorganisms in the presence of oxygen. This process typically occurs in facilities where wastewater is mixed with activated sludge or passed through a trickling filter. In both cases, oxygen is introduced to stimulate the growth of microorganisms that break down organic pollutants into carbon dioxide, water, and biomass. These treatments effectively reduce the pollutant load in wastewater, thereby improving its quality before discharge into the environment. Benefits include not only a reduction in biochemical oxygen demand and odors but also compliance with environmental standards.



Figure 37: Aeration tank

3.4 Tertiary treatments:

Tertiary treatments, also known as complementary treatments, aim to eliminate nitrogen, phosphorus, and biological pollution from domestic wastewater. These treatments refine or enhance the secondary treatment to improve the overall water quality for potential reuse in industry or irrigation. They are implemented when required by the receiving environment. Typically, these treatments include operations such as biological or combined (biological and physico-chemical) nitrification-denitrification and phosphorus removal, as well as bacteriological and viral disinfection.

3.4.1 Nitrogen and phosphorus removal:

- **Nitrogen removal:**

Nitrogen can be eliminated through three possible methods:

- ✓ **Physical:** by sedimentation when it is associated with suspended solids.
- ✓ **Physico-chemical:** a gaseous solute is driven out of the water by the action of another gas, known as a scrubbing gas.
- ✓ **Biological:** Nitrogen removal is most often achieved through biological treatments, such as "nitrification-denitrification".

- **Phosphorus removal :**

Phosphorus can be removed or "dephosphorated" using two methods:

- ✓ **Chemical method:** (For smaller units) This process involves precipitating phosphate using iron salts, aluminum salts, or lime, thereby forming an insoluble precipitate of metallic phosphate.
- ✓ **Biological methods:** Biological phosphorus removal involves promoting the accumulation of phosphorus in bacterial cultures within sludge. The biological mechanism for phosphorus removal is relatively complex and its efficacy varies. In large wastewater treatment plants, this method is often combined with physical and chemical phosphorus removal to meet specified discharge standards.

Conclusion:

This internship at the wastewater treatment plant in Ain Temouchent was an extremely enriching and formative experience, crucial for obtaining my Master's degree. By bringing and working on my dataset containing the water parameters of this station, I was able to not only

apply the theoretical knowledge acquired during my studies but also gain new practical skills essential for my professional development.

During this period, I had the opportunity to closely observe the wastewater treatment processes, understand the associated technical and environmental challenges, and actively contribute to improving water quality control methods. Working with my dataset allowed for in-depth analysis of the water parameters, providing valuable insights to optimize treatment processes and ensure better water resource management.

This internship also highlighted the importance of teamwork, interdisciplinary collaboration, and the rigorous application of standards and regulations. The skills acquired and lessons learned from this experience will undoubtedly be invaluable in my future career, preparing me to tackle technical and environmental challenges with confidence and competence.

In conclusion, this internship at the Ain Temouchent wastewater treatment plant has been a pivotal step in my academic and professional journey, allowing me to apply my knowledge and make a significant contribution to an environmentally important project. This experience was essential for obtaining my Master's degree and solidified my commitment to environmental protection and sustainable water resource management.

General conclusion

In conclusion, this master thesis highlights the potential and practicality of using wastewater for irrigation, integrating advanced technologies to enhance the efficiency and safety of this practice. The study has demonstrated that treated wastewater, with its high nutrient content, can significantly reduce the reliance on chemical fertilizers, promoting sustainable agricultural practices. The innovative system developed combines the Internet of Things (IoT) and artificial intelligence (AI) to offer real-time monitoring and prediction of water quality, ensuring it meets the standards set by the World Health Organization (WHO). The comprehensive approach includes both hardware and software components, providing a robust solution for continuous crop cultivation, particularly in arid regions. The successful implementation of this system signifies a major step towards sustainable water resource management and agricultural productivity.

Bibliography

- [1] M. Dholu et K. A. Ghodinde, « Internet of Things (IoT) for Precision Agriculture Application », in *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli: IEEE, mai 2018, p. 339-342. doi: 10.1109/ICOEI.2018.8553720.
- [2] P. Gokhale, O. Bhat, et S. Bhat, « Introduction to IOT », vol. 5, p. 41-44, janv. 2018, doi: 10.17148/IARJSET.2018.517.
- [3] A. Ngu, M. Gutierrez, V. Metsis, S. Nepal, et Q. Sheng, « IoT Middleware: A Survey on Issues and Enabling Technologies », *IEEE Internet Things J.*, vol. PP, p. 1-1, oct. 2016, doi: 10.1109/JIOT.2016.2615180.
- [4] P. Ranjan, R. Shringar Rao, K. Kumar, et P. Sharma, *Wireless Communication: Advancements and Challenges*, 1^{re} éd. New York: CRC Press, 2022. doi: 10.1201/9781003181699.
- [5] « What are the characteristics of Internet of Things(IoT) ». Consulté le: 26 juin 2024. [En ligne]. Disponible sur: <https://www.tutorialspoint.com/what-are-the-characteristics-of-internet-of-things-iot>
- [6] *IoT Architecture | Internet Of Things Architecture For Beginners | IoT Tutorial | Simplilearn*, (11 février 2020). Consulté le: 26 juin 2024. [En ligne Vidéo]. Disponible sur: <https://www.youtube.com/watch?v=KeauUcw02Q>
- [7] « What is the Internet of Things (IoT)? | IBM ». Consulté le: 26 juin 2024. [En ligne]. Disponible sur: <https://www.ibm.com/topics/internet-of-things>
- [8] N. A. Hussain, D. S. S. M.Sc, M.Phil, et Ph.D, « Precision Agriculture using IoT Sensor Network System – A Review », *Turk. Online J. Qual. Inq.*, vol. 12, n° 7, Art. n° 7, août 2021.
- [9] A. Dahane, R. Benameur, et B. Kechar, « An IoT Low-Cost Smart Farming for Enhancing Irrigation Efficiency of Smallholders Farmers », *Wirel. Pers. Commun.*, vol. 127, n° 4, p. 3173-3210, déc. 2022, doi: 10.1007/s11277-022-09915-4.
- [10] « (PDF) A Survey on the Role of IoT in Agriculture for the Implementation of Smart Farming », *ResearchGate*, doi: 10.1109/ACCESS.2019.2949703.
- [11] « L'Internet des Objets pour une agriculture plus intelligente | Mutualia ». Consulté le: 26 juin 2024. [En ligne]. Disponible sur: <https://www.mutualia.fr/agriculteur/infos/economie-et-societe/news/internet-des-objets-iot-pour-une-agriculture-plus-intelligente>
- [12] M. Alvi, T. French, R. Cardell-Oliver, P. Keymer, et A. Ward, « Cost Effective Soft Sensing for Wastewater Treatment Facilities », *IEEE Access*, vol. 10, p. 1-1, janv. 2022, doi: 10.1109/ACCESS.2022.3177201.

- [13] T. Cheng, F. Harrou, F. Kadri, Y. Sun, et T. Leiknes, « Forecasting of Wastewater Treatment Plant Key Features Using Deep Learning-Based Models: A Case Study », *IEEE Access*, vol. 8, p. 1-11, oct. 2020, doi: 10.1109/ACCESS.2020.3030820.
- [14] F. Essahlaoui, N. Elhajrat, M. Halimi, et A. Abbassi, « New approach to monitoring a wastewater irrigation system controlled by the artificial neural network (ANN) », *Groundw. Sustain. Dev.*, vol. 23, août 2023, doi: 10.1016/j.gsd.2023.100999.
- [15] M. Mohebbian, S. S. Vedaiei, A. N. Bahar, K. A. Wahid, et A. Dinh, « Times Series Prediction used in Treating Municipal Wastewater for Plant Irrigation », in *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, Edmonton, AB, Canada: IEEE, mai 2019, p. 1-4. doi: 10.1109/CCECE.2019.8861819.
- [16] S. Imen, H. Croll, N. McLellan, M. Bartlett, G. Lehman, et J. Jacangelo, « Application of machine learning at wastewater treatment facilities: a review of the science, challenges and barriers by level of implementation », *Environ. Technol. Rev.*, vol. 12, p. 493-516, août 2023, doi: 10.1080/21622515.2023.2242015.
- [17] A. El Bilali et T. Abdeslam, « Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment », *J. Saudi Soc. Agric. Sci.*, vol. 19, août 2020, doi: 10.1016/j.jssas.2020.08.001.
- [18] « Deep learning in wastewater treatment: a critical review », *Water Res.*, vol. 245, p. 120518, oct. 2023, doi: 10.1016/j.watres.2023.120518.
- [19] F. Harrou, A. Dairi, Y. Sun, et S. Mohamed, « Wastewater treatment plant monitoring via a deep learning approach », févr. 2018, p. 1544-1548. doi: 10.1109/ICIT.2018.8352410.
- [20] L. Liang, Q. Wang, et Y. Chen, « Application of support vector machine in online monitoring of wastewater treatment based on combined kernel functions », sept. 2011, doi: 10.1109/ICECENG.2011.6056825.
- [21] W. Lin, Y. Hanyue, et L. Bin, « Prediction of wastewater treatment system based on deep learning », *Front. Ecol. Evol.*, vol. 10, p. 1064555, nov. 2022, doi: 10.3389/fevo.2022.1064555.
- [22] « TH.M.INF.FR.2021.29.pdf ». Consulté le: 1 juillet 2024. [En ligne]. Disponible sur: <http://dspace.univ-tiaret.dz/bitstream/123456789/5509/1/TH.M.INF.FR.2021.29.pdf>
- [23] « Cours4.pdf ». Consulté le: 1 juillet 2024. [En ligne]. Disponible sur: https://www.unilim.fr/pages_perso/deneuille/docs/Info2PC/Cours4.pdf
- [24] « Memoire_PFE_M2.pdf ». Consulté le: 1 juillet 2024. [En ligne]. Disponible sur: http://dspace.univ-bouira.dz:8080/jspui/bitstream/123456789/14071/1/Memoire_PFE_M2.pdf
- [25] « Measuring Dissolved Oxygen », Environmental Measurement Systems. Consulté le: 1 juillet 2024. [En ligne]. Disponible sur: <https://www.fondriest.com/environmental-measurements/measurements/measuring-water-quality/dissolved-oxygen-sensors-and-methods/>
- [26] C.-A. Azencott, « Introduction au Machine Learning ».

- [27] « memoire de master.pdf.pdf ». Consulté le: 1 juillet 2024. [En ligne]. Disponible sur: <https://dspace.univ-ouargla.dz/jspui/bitstream/123456789/29107/1/memoire%20de%20master.pdf.pdf>
- [28] M. A. Merad et M. L. Bettadj, « Prédiction des propriétés des matériaux par apprentissage automatique ».
- [29] « Gradient Boosting in ML », GeeksforGeeks. Consulté le: 1 juillet 2024. [En ligne]. Disponible sur: <https://www.geeksforgeeks.org/ml-gradient-boosting/>
- [30] B. N. Eddine, « OPTIMISATION DE QUELQUES PARAMETERES DE POLLUTION DES EAUX USEES EN UTILISATION UN BIOMATERIAU ».
- [31] « Memoire final corrigé 08.11.2020.doc2.pdf ». Consulté le: 1 juillet 2024. [En ligne]. Disponible sur: <https://di.univ-blida.dz/jspui/bitstream/123456789/10826/1/Memoire%20final%20corrig%C3%A9%2008.11.2020.doc2.pdf>