A Decision Support System for aquaponics prediction based on the Intelligent Internet of Things

by

Anila Mundackal

220635323

A thesis submitted in fulfilment of the requirements for the degree

Doctor of Information and Communications Technology

in the Faculty of Informatics and Design

at the Cape Peninsula University of Technology

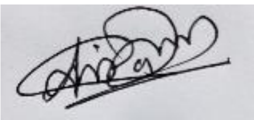
Supervisor: Prof Justine Olawande Daramola

Cape Town Campus

Date submitted: **October 2025**

DECLARATION

I, Anila Mundackal, declare that the contents of this thesis represent my own unaided work and that the thesis has not previously been submitted for academic examination towards any qualification. Furthermore, it represents my own opinions and not necessarily those of the Cape Peninsula University of Technology.



08-10-2025

Signed **Date**

ABSTRACT

Aquaponics is an emerging farming technique. Managing and optimising aquaponics systems is complex and requires expertise in aquaculture, hydroponics, and microbiology. Effective decision-making is crucial to maintaining optimal conditions for plants and fish so the system can thrive. Current research emphasises water quality monitoring but lacks the analysis of key parameters and their impact on plant growth and system productivity. There is a need for data-driven solutions to help users, especially beginners, optimise resource use and enhance performance.

The research aimed to develop a decision support system (DSS) for aquaponics that provides data-driven insights into plant growth and water quality using Explainable Artificial Intelligence (XAI). The following research objectives were used to achieve this: 1) Identify key parameters for monitoring plant growth and water quality. 2) Develop machine learning (ML) prediction models. 3) Evaluate the performance of different ML algorithms using regression metrics. 4) Design and develop a machine learning-based decision support system to facilitate decision-making in aquaponics. 5) Assess the decision support system's usability from the aquaponics stakeholders' perspective.

This study adopted an objectivist ontological stance to determine the feasibility of developing a DSS for aquaponics prediction. The epistemological stance was positivism. To meet the objectives, a deductive research approach was adopted with a quantitative methodological choice. The data parameters collected are plant height, plant diameter, Potential of Hydrogen (pH), Total Dissolved Solids (TDS), water temperature, ambient temperature and humidity. An experimental design was used to train and evaluate several supervised ML algorithms: linear regression, random forest, K-Nearest Neighbor (KNN), eXtreme Gradient Boosting (XGBoost), and Multi-Layer Perceptron (MLP). These models were assessed using the regression metrics Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R-squared (R^2), and Adjusted R-squared.

The results revealed that both random forest and XGBoost achieved the best performance for plant diameter prediction with MSE = 0.00, RMSE = 0.05, and MAE = 0.03 with R^2 and Adjusted R^2 scores of 94%. In plant height prediction, random forest performed well with MSE = 0.00, RMSE = 0.06, and MAE = 0.05, along with a high R^2 of 93% and Adjusted R^2 of 92%. XGBoost performed well in pH prediction with MSE = 0.02, RMSE = 0.13, and MAE = 0.09, along with high R^2 and Adjusted R^2 of 79%. In TDS prediction, linear regression performed well with MSE = 0.00, RMSE = 0.01, and MAE = 0.01, along with perfect R^2 and Adjusted R^2 scores of 100%.

A DSS was developed using the FLASK framework to predict plant height and diameter, water pH, and TDS. SHapley Additive exPlanations (SHAP) was used to enhance transparency by showing each feature's impact on predictions. The usability of DSS was evaluated by aquaponics stakeholders through the System Usability Scale (SUS) by. The DSS obtained a usability rating of 72%, which indicates an acceptable level of usability.

Theoretically, the study demonstrates applying ML and XAI to predict plant growth and water quality under South African conditions. Methodologically, it offers a structured approach to integrating ML, Internet of Things and AI in aquaponics. Practically, it delivers a DSS to help practitioners monitor and optimise key parameters, improving overall system performance and outcomes.

Keywords: Aquaponics, Machine Learning, Regression, Prediction, Decision Support System

ACKNOWLEDGEMENTS

I thank Almighty God for giving me this opportunity and walking alongside me throughout my journey thus far.

I would like to express my sincere gratitude to the following individuals:

- My respected Prof Justine Olawande Daramola, my supervisor, for all the guidance, support and direction he gave me ever since my first interaction with him. Thank you for allowing me to explore my academic freedom, providing opportunities to grow and most of all, the immense encouragement all the way.
- Colleagues and management of the Belgium Campus and the University of Johannesburg, who have supported me through my study journey.
- I remember my loved ones in heaven, who are looking down on me and giving me the strength and belief I need to pursue my dreams.
- During my journey, my friends and family have been a pillar of strength, especially when I seemed lost in the volume of work in front of me. You have stood by me and held my hand, reassuring me that I needed to forge ahead.

DEDICATION

I dedicate this study to my late parents, Joy Nadakkalan and Rosemol Joy, whom I miss dearly, and to my loving and immensely supportive husband, Jiby Mundackal, my lovely children, Gavin and Millan and my sister Anitha Joyson.

PUBLICATIONS FROM THE THESIS

Published paper

Anila, M. & Daramola, O. 2024. Applications, technologies, and evaluation methods in smart aquaponics: a systematic literature review. *Artificial Intelligence Review*, 58(1): 25. doi: https://doi.org/10.1007/s10462-024-11003-x.

Anila Mundackal and Olawande Daramola (2025). Smart aquaponics prediction using explainable artificial intelligence, 14th Computing Conference, London, Lecture Notes in Networks and Systems (Under review).

GLOSSARY

WHO	World Health Organisation
R^2	R-squared
Wi-Fi	Wireless Fidelity
4IR	4th industrial revolution
Adjusted R^2	Adjusted R-squared
AI	Artificial Intelligence
B	Boron
C	Carbon
Ca	Calcium
Cl	Chlorine
CO_2	Carbon Dioxide
Cu	Copper
DL	Deep Learning
DeepLIFT	Deep Learning Important FeaTures
DNN	Deep Neural Network
DO	Dissolved Oxygen
DSS	Decision Support System
DT	Decision Tree
DWC	Deep Water Culture
EC	Electric Conductivity
ES	Expert System
Fe	Iron
H	Hydrogen
H_2O	Water
IoT	Internet of Things
IIoT	Intelligent Internet of Things
K	Potassium
KNN	k-Nearest Neighbors
LDA	Linear Discriminant Analysis
LIME	Local Interpretable Model-agnostic Explanations
LoRa	Long Range
LR	Linear regression
MAE	Mean Absolute Error
MBS	Media-based systems
Mg	Magnesium.
ML	Machine Learning
MLP	Multi-Layer Perceptron
Mn	Manganese
Mo	Molybdenum
MQTT	Message Queuing Telemetry Transport
MSE	Mean Squared Error
N	Nitrogen
NFT	Nutrient Film Technique
NH3	ammonia
NH4+	Ionized ammonia
Ni	Nickel

O	Oxygen
OLS	Ordinary Least Squares
P	Phosphorus
pH	Potential of Hydrogen
RF	Random Forest
RMSE	Root Mean Squared Error
S	Sulfur
SHAP	SHapley Additive exPlanations
SUS	System Usability Scale
SVM	Support Vector Machines
TAN	Total Ammonia Nitrogen
TDS	Total Dissolved Solids
WSNs	Wireless Sensor Networks
XAI	Explainable AI
XGBoost	Extreme Gradient Boosting
Zn	Zinc

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE
INTRODUCTION AND BACKGROUND

1.1 Motivation for the study

As the global human population grows, the demand for food will proportionally increase (Hsiao & Sung, 2020). One key global challenge is producing food for a population that is constantly growing, using the limited resources available (Chandramenon et al., 2024). Farming outputs and produce suffer from reduced soil quality, insufficient water, and the impact of climate change (Singh et al., 2021; Nair et al., 2025). There is uncertainty in weather patterns due to constant fluctuations and extreme weather conditions. According to Statistics South Africa (Stats SA), poverty is rapidly growing in South Africa (Statistics South Africa, 2017). South Africa is already a water-scarce country and one of the 30 driest countries in the world (Bwapwa, 2019).

Access to and affordability of organic healthy food are big challenges these days. According to the World Health Organisation (WHO), Africa had the sharpest rise in hunger (World Health Organization, 2021). According to Jerry (2020), there is a growing demand for fresh, organic, healthy produce to feed a growing global population. Traister (2018) observed that some chemically treated food with pesticides has been scientifically proven to be harmful in causing various types of diseases, organ damage, and may even lead to death. To feed the world's increasing masses sustainably, a rethinking or shift is required from the existing method of growing crops, with more innovative approaches that need to be introduced faster. It is in this light that new farming methods such as Hydroponics, Aeroponics and Aquaponics have come to fruition (Kok et al., 2024; Nair et al., 2025).

Aquaponics is a combination of aquaculture and hydroponics. Hydroponics is a soilless farming method in a nutrient-rich water solution with or without a medium (Kumar & Savaridassan, 2023). However, aquaculture is the process of cultivating fish in water (Kathuria et al., 2024). Aquaponics is a complex system which requires continuous monitoring and water quality management, fish health, and plant growth, making it tedious manual labour. This necessity has driven the development of smart aquaponics systems that integrate advanced technologies such as the Internet of Things (IoT), Machine Learning (ML), Artificial Intelligence (AI), and more (Jiang & Liu, 2024; Liu & Jiang, 2024; Perumal et al., 2024). These systems not only bridge the gap between technology and agriculture but also enhance the efficiency, sustainability and productivity of aquaponics farming (Liu & Jiang, 2024; Sridevi et al., 2024). Intelligent Internet of Things (IIoT) is a technology that combines IoT, machine learning (ML), and artificial intelligence (AI). IoT enables data collection while ML and AI process data to extract valuable insights (Zhang, 2021; Aouedi et al., 2024). The implementation of IIoT technology in aquaponics systems enhances the time taken to gather data. It further provides more accurate data, avoiding manual labour interventions. IIoT technology thus assists with timely decision-making and optimising system performance through the intelligent analysis of the aquaponics system. A Decision Support System (DSS) derives much benefit from using an IIoT technology. The benefit is that it can pick up exceptions and anomalies. Furthermore, once these slight variations are picked up, changes can be made more proactively instead of reactively, which prevents crop losses and operational inefficiencies. The Intelligent Internet of Things boosts the power and value of DSS.

Various studies have explored the predictive models for key aspects of aquaponics, such as leaf disease identification, biomass prediction, pH prediction, plant growth, and fish length and weight (Ghandar et al., 2021; Mori et al., 2021; Amano et al., 2022; Debroy & Seban, 2022b; Debroy & Seban, 2022a; Khandakar et al., 2024). Khandakar et al. (2024) focused on fish weight and fish length prediction for fish farming optimisation. In this study, the team integrated local interpretable model-agnostic explanations (LIME) for model transparency.

The application of Explainable AI (XAI) in smart aquaponics remains limited, with few studies focusing on the explainability and interpretability of machine learning (ML) and deep learning (DL) models used in this domain. While predictive analytics methods have been widely applied in aquaponics research addressing key areas like plant health, fish growth, and environmental conditions, these models are often not communicated effectively to end-users for practical, real-world decision-making (Liu & Jiang, 2024; Anila & Daramola, 2024). Using IIoT in aquaponics more intelligently can lead to smart aquaponics systems, thus contributing to a better decision support system.

The development of decision support systems (DSS) for aquaponics using IIoT would provide stakeholders with actionable insights derived from predictive models. Most studies focus on generating predictions but fail to integrate these outputs into a user-friendly system that can guide daily operations or assist in making strategic decisions (Mori et al., 2021; Debroy & Seban, 2022b; Liu & Jiang, 2024). This limits the practical utility of these models in real-world aquaponics applications (Anila & Daramola, 2024).

1.2 Background

Aquaponics has been an emerging revolution in the farming world (König et al., 2018; Turnsek et al., 2020). Aquaponics is a very beneficial food production technique as people can cultivate organic vegetables, fruits, and fish simultaneously and efficiently on a small scale using minimal resources (Rakocy et al., 2006; Nair et al., 2025). Aquaponics provides for flexibility and scalability in the future for larger farming or commercial practices should the need arise. The fewer input resources required, such as water, land, the central point of system management, and income generation opportunities, are huge benefits of this type of farming (Fruscella et al., 2021:1661).

There has been a global drive to use technology to improve existing farming practices. The recent Coronavirus (COVID-19) pandemic revealed the need to automate and have agricultural machinery and farming methods self-regulate. This allows a farmer or organisation to remotely manage farms accurately and achieve productivity with minimal need for human involvement or intervention.

With the advent of the 4th industrial revolution (4IR), there has been a giant leap in the Internet of Things (IoT) usage and Artificial Intelligence (AI) technologies to achieve higher productivity and efficiency in daily operations (Alhnaity et al., 2019; Tai, 2020). However, combining IoT, machine learning (ML), and artificial intelligence (AI) within IIoT technology can significantly enhance data-driven decision-making, thereby enabling smart agricultural practices. Likewise, aquaponics will benefit when driven towards integrated, smarter technologies.

Smart aquaponics models generally process data via sensors, and the data collected is compared to pre-determined optimal range parameters (Reddy et al., 2020; Sridevi et al., 2024; Perumal et al., 2024). A different study used various Grove sensors to monitor the following environmental and aquaponics parameter values, namely: sunlight, pH, water, water level, water temperature, electrical conductivity, ammonia, etc. When the monitored values fell above or below the optimal range, the microcontroller kicked into action (Khaoula et al., 2021). In the study conducted by Valiente et al. (2018) when sensor data showed a value outside the optimal range, a message was triggered and sent to the programmed contact via phone or web. Kumar et al. (2016) and Khaoula et al. (2021) demonstrated that cloud storage services were used to store collected data. The collected data allows for trends and patterns to be established, which enables different forecasting and prediction capabilities (Debroy & Seban, 2022b; Liu & Jiang, 2024). Many aquaponics studies have focused on monitoring and reporting via emails, SMS, notifications and so on (Manju et al., 2017; Hsiao & Sung, 2020). These methods reduce the need for manual intervention to help maintain an efficient aquaponics system performance.

Aquaponic units can be installed in the field, greenhouse, tunnel or even indoors (Mchunu et al., 2018; Reyes-Yanes et al., 2020). Factors influencing the aquaponics system are compatibility of fish and plants, fish stocking density, amount of fish feed, nitrifying bacteria, climate factors, water quality and so on (Nair et al., 2025). A South African survey conducted by Mchunu et al. (2018) concluded that most aquaponics farmers required knowledge of technology to increase aquaponics food production. Accurate crop yield is crucial for making decisions related to agricultural risk management as well as for feasibility calculations. Start-up farmers do not know or understand which parameters need to be regulated, by how much, or the roadmap on how to manage an aquaponics system optimally for maximum yield (Mchunu et al., 2019).

Aquaponics is a rapidly growing farming method due to reduced resource consumption, such as water, soil, and land. The improvement of aquaponics productivity is possible through the application of IIoT, which entails integrating IoT technologies and AI to optimise the aquaponics critical parameters for maximum yield (Khaoula et al., 2021; Abbasi et al., 2022; Liu & Jiang, 2024; Sridevi et al., 2024).

1.3 Research problem

Aquaponic systems are relatively complex to monitor and manage due to a lack of expert knowledge (Hsiao & Sung, 2020; Karimanzira & Rauschenbach, 2021). Despite aquaponics having the potential to aid sustainable food production, there is still limited research on plant growth data trends within aquaponics (Channa et al., 2024). This makes it difficult to understand and optimise plant growth performance in aquaponic systems (Chowdhury & Asiabanpour, 2024). Careful monitoring and control of key parameters in an aquaponic system help maintain optimal conditions for fish health, plant growth, and the activity of microorganisms (Debroy et al., 2025; Nair et al., 2025). Identifying these influential parameters is essential, as it would allow stakeholders to make informed decisions that optimise the plant growth, fish growth, resource usage, and overall system performance (Khandakar et al., 2024; Nair et al., 2025). The lack of uncertainty in making decisions and the implementation of corrective actions timeously directly affect aquaponics fish and crop yield (Hsiao & Sung, 2020). In both startup and commercial-level farming, there exists a need for an informed decision support tool to optimise aquaponics productivity (Pechlivani et al., 2025).

So far, systems that provide clear, data-driven insights to stakeholders, particularly those new to the field, to optimise aquaponics productivity and improve system outcomes are not common.

1.4 Aim and objectives

1.4.1 Aim

This study aimed to develop a decision support system for aquaponics prediction that offers data-driven insight into plant growth and water quality parameters using Intelligent Internet of Things.

1.4.2 Objectives

The objectives of this study are to:

1. Identify the key parameters used to measure plant growth and the monitored water quality parameters in aquaponics systems.

2. Develop a machine learning (ML) prediction model to determine the optimal levels of key parameters for the aquaponics system.

3. Evaluate the performance of the different ML algorithms using suitable regression metrics.

4. Develop an ML-based data-driven decision support system for aquaponics.

5. Assess the usability (encompassing effectiveness, efficiency and satisfaction) of a decision support system from the perspective of aquaponics stakeholders.

1.5 Research Questions

The main research question for this study is:
How can a decision support system for plant growth and water quality prediction in aquaponics be developed through the application of Intelligent Internet of Things?

The sub-research questions are:

1. What are the parameters required for measuring plant growth, and which water quality parameters are essential for monitoring in aquaponics systems?
2. How can an ML prediction model for aquaponics be developed?
3. What is the comparative performance of the different ML algorithms for aquaponics prediction?
4. How can an ML-based data-driven decision support system for aquaponics be developed?

5. How can the usability of the decision support system for aquaponics prediction be determined from the perspective of stakeholders?

1.6 Delineation of the study

This study focused on developing a decision support system for aquaponics stakeholders capable of predicting plant growth and water quality while providing actionable insights through the integration of machine learning (ML) and Explainable AI (XAI). The data for the study were collected from a single field located at the University of Johannesburg, Johannesburg, South Africa, under the supervision of the field manager.

1.7 Significance of the study

Rising transport costs, and the high costs of owning vast farming land, has many challenges in South Africa and globally. This study encourages startup and subsistence farmers to consider aquaponics as a potential food and income source. The developed decision support (DSS) tool) and its regular use will provide the users and stakeholders with valuable assurance that they are on the correct path to realise the maximum plant growth.

The study will also provide insights on the parameters with the most impact on plant growth so that aquaponics stakeholders can focus on them. Wider adoption of the DSS deliverable from this study when embraced, will enhance aquaponics farming both in South Africa and globally.

1.8 Thesis outline

The entire thesis is organised into six chapters. A brief description of the chapters is given below.

Chapter 1: This chapter provides a brief explanation of the following components: motivation of the study, background, the problem statement, the aim and objectives of the study, the research questions and the significance of the study.

Chapter 2: This chapter provides an overview of the theoretical background and related work of the study.

Chapter 3: This chapter illustrates the methodology followed to accomplish the research objectives.

Chapter 4: This chapter explains how the data was collected for the experiment.

Chapter 5: This chapter presents the machine learning experimentation performed on the data collected for the study.

Chapter 6: This chapter explains how the decision support system (DSS) was developed and the usability evaluation of the DSS.

Chapter 7: This chapter presents the contribution and recommendations for future work.

1.9 Chapter summary

This chapter presented the motivation for the study, provided a brief background, outlined the study aim and objectives, and formulated research questions to guide the investigation toward the objectives. Additionally, the chapter highlighted the delineation and significance of the study and, finally, provided an overview of the thesis structure.

CHAPTER TWO
LITERATURE REVIEW

A literature review is a systematic process of reviewing, collecting, and synthesising previously written works (Snyder, 2019). It can be articulated in the form of a comprehensive previous scholarly work, a concise report of the latest primary data, or the result. (Cooper, 1998:3; Cresswell, 2014:24).

This chapter consists of three parts: theoretical background, related work, and research gaps. The theoretical background provides theoretical knowledge about key topics that provide the foundation for this study. The related work focuses on the review of previous scholarly work on smart aquaponics. The research gaps summarise the gaps in the reviewed work.

2.1 Theoretical background

This section provides background knowledge on relevant key topics such as hydroponics, aquaculture, aquaponics, the Internet of Things, machine learning, explainable artificial intelligence (XAI), Intelligent Internet of Things, expert systems and decision support systems.

2.1.1 Hydroponics

The word hydroponics originates from two Greek words: ‘hydro’, meaning water, and ‘ponos’, meaning labour (Shrestha & Dunn, 2010; Rajaseger et al., 2023; Reddy et al., 2024). Hydroponics is a soilless cultivation approach to growing agriculture in a nutrient-rich water solution with or without a medium (Shrestha & Dunn, 2010; Kumar & Savaridassan, 2023). Commonly used supporting mediums are wood fibre, expanded clay, coir, perlite, vermiculite, brick shards, polystyrene packing peanuts, gravel, etc. (Roberto, 2003:16; Shrestha & Dunn, 2010; Somerville et al., 2014; Rajaseger et al., 2023). The selection of a medium is based on the following characteristics: surface area, pH, cost, weight, life span, water retention, plant support, and ease of working with the medium (Somerville et al., 2014). Compared to in-ground cultivation, soilless cultivation has various benefits such as: requiring less land, less water, minimal fertiliser loss due to chemical, biological or physical processes, minimal human intervention and better yield (Shrestha & Dunn, 2010; Somerville et al., 2014; Kumar & Savaridassan, 2023). There are different types of hydroponic systems in use (Kumar & Savaridassan, 2023).

2.1.1.1 Types of hydroponic systems

The hydroponics growing method involves two ways: either a liquid system/solution culture or an aggregate system/solid media. There is no physical support for the plant root in the liquid system, and the nutrient solution is directly transferred to the plant. The aggregate system uses a support/growing/substrate medium to hold plant roots. If the excess nutrient solution is circulating/recycling/recovering in the hydroponic system, then the system is a closed/recirculating system, or else it is an open system (Shrestha & Dunn, 2010; Mason et al., 2018:12; Resh, 2013:2; Rajaseger et al., 2023; Rajendran et al., 2024). The implementation of a mechanical device in the hydroponics system for recirculating the nutrient water makes the system an active one. A passive system is where the roots absorb nutrients from the water without any mechanical device (gravity) and make use of capillary action (Roberto, 2003:20; Jones Jr., 2005:121; Shrestha & Dunn, 2010; Blancaflor et al., 2022; Reddy et al., 2024). There are different types of hydroponics techniques: floating/raft system, ebb and flow (flood and drain), Nutrient Film Technique (NFT), drip system, wick system, Deep Water Culture system (DWC) and aeroponic system (Shrestha & Dunn, 2010; Maucieri et al., 2019:90-93; Rajaseger et al., 2023; Rajendran et al., 2024; Naresh et al., 2024). The types of hydroponics are shown in Figure 2.1.

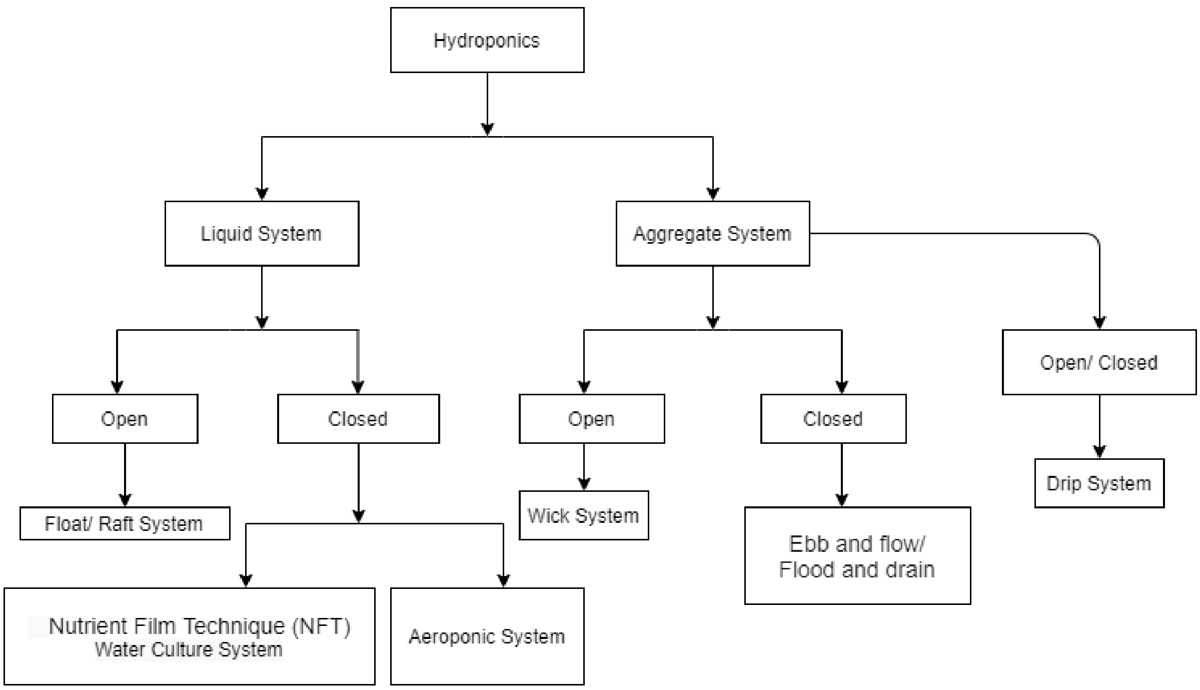


Figure 2.1: Types of hydroponics (Shrestha & Dunn, 2010)

a. Floating/Raft system

In a floating/raft system, plants are grown in plastic cups, which are placed into a Styrofoam sheet. Styrofoam floats on aerated nutrient water, and the plants' roots grow into the nutrient water (Jones Jr., 2005:248; Shrestha & Dunn, 2010; Velazquez-Gonzalez et al., 2022; Luta & Siregar, 2023).

b. Ebb and flow/ flood and drain

This system works by flooding the grow tray with a nutrient-rich water solution for a short period. Thereafter, the solution is drained back into the reservoir. Hence, this system is also referred to as a flood and drain system. This process is completed using a timer-controlled pump. The frequency of the process can be timer-controlled and depends on the size and type of plants. Factors such as temperature, humidity and the type of growing medium also serve to determine the timer settings (Jones Jr., 2005:143; Shrestha & Dunn, 2010; Rajaseger et al., 2023; Rajendran et al., 2024).

c. Nutrient Film Technique (NFT) system

Nutrient Film Technique was developed in England during the ’60s by Dr Allen Cooper (Sharma et al., 2018; Bhat et al., 2023). It was mainly developed to address the shortcomings of the ebb and flow system (Kannan et al., 2022). This system is designed in such a way that a nutrient-rich water system is completely circulated. This nutrient water is pumped to the growth tray using a water pump that operates without time control (Shrestha & Dunn, 2010; Sharma et al., 2018; Kannan et al., 2022; Blancaflor et al., 2022).

d. The drip system

The submersed pump is controlled with the help of a timer. The function of the timer is to activate the pump, thus letting the nutrient-rich solution be dripped into the base part of each plant via a tiny drip line (Shrestha & Dunn, 2010; Kannan et al., 2022; Rajaseger et al., 2023; Rajendran et al., 2024).

e. The wick system

This is a passive system and is one of the simplest hydroponic systems. This system uses a wick to link the roots of the growing media and the nutrients. There are no mechanical or moving parts in this system. It uses capillary action (Shrestha & Dunn, 2010; Subakti et al., 2022; Rajaseger et al., 2023; Prianka et al., 2024).

f. Deep Water Culture (DWC)

Deep Water Culture is the simplest of all active hydroponic systems. In this system, plant roots are submerged in the nutrient solution. This allows the roots to have a continuous supply of oxygen and water. Plants are supported on a floating platform or base made of Styrofoam. This base floats on top of the nutrient solution. An air pump is used to provide air to the air stone, which oxygenates the nutrient solution, ensuring the plant roots receive sufficient oxygen (Shrestha & Dunn, 2010; Saaid et al., 2013; Kannan et al., 2022; Rajaseger et al., 2023; Rajendran et al., 2024).

g. Aeroponic system

An aeroponic system has an enclosed growing chamber. A mist of nutrient-rich solution is sprayed at regular intervals. This aeroponic system is the most high-tech type of hydroponic gardening currently. Timer controls are used to pump the nutrients, but rather on much shorter bursts of a few seconds every couple of minutes (Shrestha & Dunn, 2010; Rajaseger et al., 2023; Rajendran et al., 2024). Nutrient-rich water or nutrient solution is required for better crop yield and quality.

2.1.1.2 Hydroponic plant nutrition

In hydroponics, plants receive all essential nutrients through a nutrient solution (Velazquez-Gonzalez et al., 2022). The hydroponic system uses dissolved fertiliser salts to supply essential nutrients to plants, excluding carbon, hydrogen and oxygen, which are obtained from the air (Resh, 2013:31; Velazquez-Gonzalez et al., 2022). Nutrient solution components are divided into macro- or micronutrients based on the quantity of the plant's nutrient requirements (Trejo-Téllez & Gómez-Merino, 2012:1; Maucieri et al., 2019:94; Kannan et al., 2022; Rajaseger et al., 2023). The purpose of these essential nutrients in plants, their roles, and the symptoms of deficiency are discussed in the following sections. This information assists the researcher in observing whether plants are growing healthily or not. If any deficiency symptoms are identified, this contributes to the understanding of which nutrients are lacking.

a. Essential elements/nutrients

In general, there are 17 essential elements/nutrients required for optimal plant growth and quality (Schwarz, 1995:8; Resh, 2013:9; Kannan et al., 2022; Veazie et al., 2022; Rajaseger et al., 2023). The selection of essential elements strictly falls under three criteria (Schwarz, 1995: 5; Arnon, 1950, 1951, cited in Resh, 2013:9; Veazie et al., 2022).

1. The plant cannot complete its life cycle if the element is not present.
2. The element activity must be specific, and the element must not be replaceable by other elements.
3. The element must act within the plant and not result in another element being more easily accessible.

Some elements are required in larger quantities, known as macro elements/macronutrients, while some require relatively smaller quantities, known as minor elements/micronutrients/trace (Maucieri et al., 2019:94; Blancaflor et al., 2022; Thakur et al., 2023). The macro elements are carbon (C), hydrogen (H), oxygen (O), nitrogen (N), phosphorus (P), potassium (K), calcium (Ca), sulfur (S), and magnesium (Mg). The microelements are iron (Fe), chlorine (Cl), manganese (Mn), boron (B), zinc (Zn), copper (Cu), molybdenum (Mo) and Nickel(Ni) (Schwarz, 1995:8; Resh, 2013:9-10; Blancaflor et al., 2022; Rajaseger et al., 2023; Thakur et al., 2023). Plants also get the following macronutrients: carbon(C), oxygen(O) and hydrogen(H) from carbon dioxide (CO_2) and water (H_2O) (Resh, 2013:9; Blancaflor et al., 2022; Rajaseger et al., 2023).

The appropriate balance between the macronutrients is required for crop growth. Most of the plant's dry weight contains an overall 90-95 % of carbon (C), oxygen (O), and hydrogen (H) and the remaining 5-10 % is the rest of the six elements (Schwarz, 1995:7; Jones Jr., 2005:37-38; Resh, 2013:9; Kannan et al., 2022).

b. The role of the essential nutrients in plants

Each nutrient plays a major role in plant growth. Even with an adequate supply of nutrients, plants may still experience nutrient deficiencies. The nutrients, along with their associated deficiency symptoms, are depicted in Table 2.1.

Table 2.1: Role of plant nutrients and deficiency symptoms

Nutrients	Roles	Deficiencies
Oxygen (O)	Oxygen is an essential nutrient for plant growth and the formation of sugar, starches and cellulose. It is further used in the process of respiration as well (Roberto, 2003:27; Rajaseger et al., 2023).	Hinders healthy plant growth (Velazquez-Gonzalez et al., 2022; Rajaseger et al., 2023).
Hydrogen (H)	Hydrogen is vital for the chemical reaction process, whereby plant roots can absorb nutrients. Hydrogen is readily available from water and air. It assists with the formation of starches and sugars (Roberto, 2003:27; Rajaseger et al., 2023).	Impairs healthy plant growth (Velazquez-Gonzalez et al., 2022; Rajaseger et al., 2023).
Carbon (C)	Carbon is found in cell walls and forms the backbone of most plant biomolecules, including proteins, starches, and cellulose, which are composed of carbon, hydrogen, and oxygen. It serves as both a building block and a source of energy, helping plants generate sugars through photosynthesis, a process driven by chlorophyll (Roberto, 2003:27; Ahluwalia, 2022:477).	Leads to poor plant growth (Velazquez-Gonzalez et al., 2022; Rajaseger et al., 2023).

Nitrogen (N)	Plants absorb nitrogen through their roots to generate amino acids, proteins, enzymes and chlorophyll, which are essential for plant growth (Resh, 2013:11; Roberto, 2003:27; Mason et al., 2018:85; Maucieri et al., 2019:94; Rajaseger et al., 2023). Nitrate and ammonium are the two forms of nitrogen that plants can absorb (Schwarz, 1995:10, Maucieri et al., 2019:94; Rajaseger et al., 2023; Thakur et al., 2023; Hong et al., 2024). For most plants, nitrate is the primary nitrogen source. It is non-toxic and can be stored in the plant. A large quantity of ammonia intake affects plant growth (Maucieri et al., 2019:94; Daiane et al., 2021).	Growth is constrained, with shorter and leaner stalks. Plant leaves have a yellowish colour overall and reduced fruit yield (Roberto, 2003: 29; Mason et al., 2018:85; Jones Jr., 2005:388; Maucieri et al., 2019:94; Rajaseger et al., 2023).
Potassium (K)	Potassium is essential for plant health, enhancing disease resistance and nutrient absorption (Rajaseger et al., 2023).	The tips and outer edges of the leaves die in monocot plants. Leaves of dicots are chlorotic at first; however, dead areas soon start to develop. Also, it causes weak stems (Mason et al., 2018:88; Rajaseger et al., 2023).
Calcium (Ca)	Calcium assists in permeating the membrane, assisting in the division of cells as well as the formation of the cell wall (Maucieri et al., 2019:96; Rajaseger et al., 2023; Thakur et al., 2023).	Spotted young leaves with irregular margins. Distorted young leaves, small-sized leaves, shoot and root tip death, and restricted bud development are other symptoms (Mason et al., 2018:89; Veazie et al., 2022). Stunted plant growth, deformation of younger leaf margins, stunted root systems without fine roots are further symptoms (Maucieri et al., 2019:96; Rajaseger et al., 2023).
Magnesium (Mg)	Magnesium is useful for building up the wall of chlorophyll molecules (Maucieri et al., 2019:96; Rajaseger et	Chlorosis begins to form in the vein areas of the leaves, leading to yellowing between the veins and a

	al., 2023). Magnesium boosts glucose synthesis and influences enzyme activity, which supports healthy leaf development and efficient energy production in plants (Rajaseger et al., 2023; Thakur et al., 2023).	reduction in chlorophyll concentration (Rajaseger et al., 2023). This progresses towards the death of the tissue (necrosis). The severely affected leaves eventually fall off (abscise) (Veazie et al., 2022).
Phosphorus (P)	Phosphorus promotes the fast growth of buds and several flowers, and encourages root development of the plants (Maucieri et al., 2019:95; Rajaseger et al., 2023).	Plant development and maturity are often delayed. Plants are a dark green colour and more often than not, advance to get a reddish or purple colour and display stunted growth in the vegetative apex (Mason et al., 2018:87; Maucieri et al., 2019:95; Veazie et al., 2022; Rajaseger et al., 2023).
Sulfur (S)	Essential for protein production and maintaining plant strength and health Rajaseger et al., 2023).	Light yellow leaves, stunting plant growth and woody stems (Mason et al., 2018:89; Veazie et al., 2022; Rajaseger et al., 2023).
Chlorine (Cl)	Chlorine helps maintain osmotic pressure within plant cells and supports cell turgor pressure, which is essential for optimal water and nutrient transfer, as well as overall plant health and growth (Thakur et al., 2023).	Insufficient chlorine can cause leaf chlorosis and necrosis, as well as leaf wilting, restricted root growth, and stunted development (Mason et al., 2018:90; Maucieri et al., 2019:97; Thakur et al., 2023).
Iron (Fe)	Iron is required for chlorophyll formation and enzyme functions, which are critical for photosynthesis (Maucieri et al., 2019:96; Rajaseger et al., 2023).	Chlorosis between the veins, especially in younger leaves, and can spread to older leaves, reducing root system growth (Mason et al., 2018:90; Maucieri et al., 2019:96; Rajaseger et al., 2023).
Manganese (Mn)	Manganese helps prevent pathogens and increases the root cells (Maucieri et al., 2019:97). Required for photosynthesis and enzyme activities (Rajaseger et al., 2023; Thakur et al., 2023).	Stunted growth. Chlorosis in the vein areas of the leaves begins to form. This progresses from the leaf ends or periphery and moves inwards (Mason et al., 2018:90; Rajaseger et al., 2023).

Boron (B)	Assist with the setting of the fruit and developing seed cells, cell division, pollen formation, and sugar transport (Maucieri et al., 2019:97; Rajaseger et al., 2023).	Delicate leaves and stems, irregularity in plant growth, and stem and root tip death. Incomplete growth of young light green leaves of the terminal bud and twisted leaves when it grows back (Mason et al., 2018:91, 94; Rajaseger et al., 2023). Young leaves increase their thickness and have a leathery consistency (Maucieri et al., 2019: 97; Rajaseger et al., 2023).
Zinc (Zn)	Zinc is essential for enzyme activation and hormone regulation (Maucieri et al., 2019:97; Rajaseger et al., 2023).	Chlorosis in-between the veins, especially in young leaves, inhibited growth, distorted leaf margins, and spots spread around the entire plant (Mason et al., 2018:91; Maucieri et al., 2019:97; Rajaseger et al., 2023).
Copper (Cu)	Copper assists with the respiratory process involving photosynthesis and is important for enzyme functions (Maucieri et al., 2019:97; Rajaseger et al., 2023).	Twisted young leaves and yield reduction. Restricted growth and the growing tip may die (Mason et al., 2018:91). Interveinal chlorosis leads to the collapse of the leaves' tissues (Maucieri et al., 2019:97; Rajaseger et al., 2023).
Molybdenum (Mo)	Helps in nitrogen metabolism and protein synthesis (Jones Jr., 2005:400; Maucieri et al., 2019:98; Rajaseger et al., 2023).	Chlorosis and necrosis in-between the veins in older leaves evolve into younger leaves and deformed younger leaves (Jones Jr., 2005:400; Mason et al., 2018:92; Maucieri et al., 2019: 98; Rajaseger et al., 2023).
Nickel (Ni)	Plays a role in nitrogen metabolism and enzyme function (Rajaseger et al., 2023).	Reduces plant growth and causes leaf deformation (Rajaseger et al., 2023).

When adding fish into a hydroponics reservoir, it becomes an aquaponics system, which is an integrated farming technology (Maucieri et al., 2019:77; Luta & Siregar, 2023). The main difference between aquaponics and hydroponics is that fish waste makes reservoir water very nutrient-rich, which is critical for plants (Lennard & Goddek, 2019:114; Rajaseger et al., 2023).

2.1.2 Aquaculture

Aquaculture is the process of cultivating fish or other aquatic organisms in water (Krishna et al., 2023; Kathuria et al., 2024). Fish are great, globally demanded aquatic organisms in aquaculture and are rich sources of protein and omega-3 (Kusuma et al., 2023; Krishna et al., 2023; Kathuria et al., 2024).

Many factors influence the growing state of aquaculture, such as aquaculture organisms' species choice, aquaculture organisms' density based on the aquaculture water capacity, the number of organisms, and food uses in aquaculture, including water quality parameters management (Deng et al., 2010). However, in aquaculture, water quality plays an important role in the growth of aquatic organisms (Deng et al., 2010; Krishna et al., 2023). Insufficient water quality can lead to stress, diseases, including the death of aquatic organisms, thereby negatively impacting productivity, the inability to harvest in the desired time, and industry profit as well (Dupont et al., 2018; Krishna et al., 2023). Various parameters are considered to assess the water quality, such as pH, hardness, dissolved oxygen, water temperature, carbon dioxide, nitrate, nitrite, salinity, Total Dissolved Solids (TDS), turbidity, water colour and so on (Bhatnagar & Devi, 2013; Yildiz et al., 2019: 445; Krishna et al., 2023; Kathuria et al., 2024). However, the most commonly monitored parameters are temperature, dissolved oxygen, and pH (Abbink et al., 2012; Dupont et al., 2018; Krishna et al., 2023; Khandakar et al., 2024). Thus, it is important to keep these parameters in an optimal range for growth performance. Table 2.2 specifies the optimum range of water temperature, dissolved oxygen and pH, the reason for regulation, and what will happen if it deviates from the optimum range (Wongkiew et al., 2017; Dupont et al., 2018; Espinal & Matulić, 2019:39; Lennard & Goddek, 2019:130; Verma et al., 2022).

Table 2.2: Popularly monitored parameters in aquaculture

	Temperature	**Dissolved Oxygen**	**pH**
Optimum range	Depends on the fish species	>5ppm	7 - 8.5
Reason for regulation	✓ To control disease ✓ To control oxygen consumption ✓ Fish productivity	Fish survival	✓ Controls fish metabolism ✓ Microbial activities

Deviation from optimal range triggers	The higher temperature required frequent microbiota biochemical activity that demanded more oxygen	✓ Can lead to fish death ✓ Fish growth became slow ✓ High stress ✓ Nitrifying biofilter failure	✓ Fish stress ✓ Fish growth became slow

2.1.3 Aquaponics

Aquaponics is a combination of aquaculture and hydroponics (Murdan & Joyram, 2021). The word “Aquaponics” is a blend of “Aqua” and “Ponics”. “Aqua” refers to water or aquaculture, which is fish farming and “Ponics” branches from Hydroponics, which refers to growing plants in water without soil (Thorarinsdottir et al., 2015:9; Murdan & Joyram, 2021). Compared to traditional farming, aquaponics farming has the following benefits: less space for farming required, no soil-borne diseases, eliminates pesticides, reduces insect infection and pests, produces healthy organic food, hydroponic cultivars can be harvested in less time, increases food production, minimal chemical usage, and reduces water consumption massively (Rakocy et al., 2006; Shafeena, 2016; Manju et al., 2017; Yanes et al., 2020; John & Mahalingam, 2021; Ubayasena et al., 2023; Sridevi et al., 2024). Aquaponics is a reliable and sustainable solution for global food security (Murdan & Joyram, 2021; Friuli et al., 2021).

Fish consume the fish feed and excrete waste, primarily in the form of ammonia, through their gills, enriching the water with nutrients beneficial for plant growth (Sallenave, 2016; Ru et al., 2017; Kamil et al., 2020).

The nitrogen cycle plays a major role in the aquaponics system because it converts fish waste into nutrients that are beneficial for plants, resulting in better production (Petrea et al., 2013; Ru et al., 2017; Kim et al., 2022). Thus, nitrogen is the main source of nutrients for fish, plants and micro-organisms. In this integrated system, water is reused multiple times. This frequent water reuse causes the generation and collection of non-toxic nutrients and organic matter, which is useful for plants. This non-toxic nutrient and organic matter can contribute to the efficient and optimal growth of plant crops. The Nitrosomonas and Nitrospira bacteria convert ammonia to nitrite and then to nitrate, respectively (Rakocy et al., 2006; Gnanasagar & Vivek, 2020; Eneh et al., 2023).

The aquaponic system can cultivate plants or veggies and fish concurrently (Rakocy et al., 2006; Sridevi et al., 2024). The fish grow in the fish tank, and the plant grows in the hydroponic grow system.

2.1.3.1 Types of aquaponic systems

An aquaponic unit is a combination of an aquaculture unit and a hydroponic unit. There are two main types of aquaponic systems: coupled and decoupled. Coupled aquaponics operates with a single closed-loop water recirculation system. There is a direct transfer of the nutrient-rich water from the fish tank to the hydroponic unit and back (Palm et al., 2019:163; Chandramenon et al., 2024). Whereas, in a decoupled aquaponic unit, there are separate loops for the aquaculture and hydroponic units. Water does not circulate back from the hydroponic unit to the fish tanks, providing independent control over each system (Goddek et al., 2019:202; Chandramenon et al., 2024). A coupled vs decoupled aquaponics system is shown in Table 2.3 (Chandramenon et al., 2024).

Table 2.3: Coupled vs decoupled aquaponics system

Type	Features	Benefits	Demerits
Coupled	Mainly used at a mini/ hobby/domestic/ backyard/ demonstrative/ small and semi-commercial level. May have short-term nutrient peaks and variations. Production depends on feed demand, no of plants and fish. Gravity influences water flow. Single loop systems/ scaling from small-medium-large	Easy to implement, maintain, and manage. Requires less infrastructure Simple architecture	pH, temperature, and nutrient concentration are compromised Less profitable Lower commercial profile
Decoupled	Mainly used at a semi/ full commercial level. Multiloop systems detached units	More profitable Improved nutrient stability Improved pest management	Complex design Implementation needs expertise Hard system maintenance

Commonly used hydroponic grow systems in aquaponics are Media-based systems (MBS), Deep Water Culture (DWC), also known as the floating or raft method, and Nutrient Film Technique (NFT) (Goddek et al., 2015; Shafeena, 2016; Kledal et al., 2019:489; Singh et al., 2021; Arakkal Thaiparambil & Radhakrishnan, 2022).

2.1.3.2 Parameters affecting aquaponics plants and fish growth

The parameters that affect the production of both plant and aquatic animals are the concentration of macro- and micronutrients, water, pH, dissolved oxygen, water temperature, light, air temperature, and CO_2 in the air (Thorarinsdottir et al., 2015:42; Chandramenon et al., 2024). However, the fish and plants ratio is an important factor for balanced nutrient distribution (Goddek et al., 2015; Dharshan et al., 2024).

a. Nutrients

Nutrients are essential for plant growth; plants get nutrients from fish waste, and the required quantity may be moderate or significant (Rakocy et al., 2006; Ru et al., 2017; Chandramenon et al., 2024). Nutrients absorbed by plants can be classified as micronutrients and macronutrients. Micronutrients require only smaller quantities, whereas macronutrients require larger quantities. All these nutrients must be balanced for optimal plant growth (Thakur et al., 2023). Plants absorb all required micronutrients and macronutrients from cultured water. Nevertheless, water (H_2O) and carbon dioxide (CO_2) supply carbon (C), oxygen (O) and hydrogen (H) to the plants (Rakocy et al., 2006; Ru et al., 2017; Blancaflor et al., 2022; Rajaseger et al., 2023). The macronutrients and micronutrients required for aquaponics plants are shown in Table 2.4 (Rakocy et al., 2006; Blancaflor et al., 2022; Rajaseger et al., 2023; Thakur et al., 2023).

Table 2.4: Macronutrients and micronutrients required for aquaponics plants

Macronutrients	**Micronutrients**
Carbon(O)	Chlorine (Cl)
Oxygen(O)	Iron (Fe)
Hydrogen(H)	Manganese (Mn)
Nitrogen (N)	Boron (B)
Potassium (K)	Zinc (Zn)
Calcium (Ca)	Copper (Cu)
Magnesium (Mg)	Molybdenum (Mo)
Phosphorus (P)	
Sulfur (S)	

b. Water pH level

The Potential of Hydrogen (pH) level of a solution indicates the concentration of hydrogen ions present in the solution and relative acidity (Alselek et al., 2022; Lindholm-Lehto, 2023; Chandramenon et al., 2024). pH is a vital parameter in the aquaponics system as it directly impacts the lifecycle and health of both fish and cultivated plants, including the performance of the nitrifying bacteria (Maulini et al., 2022; Kumar et al., 2023; Kok et al., 2024; Channa et al., 2024). In the aquaponics system, it is essential to maintain a pH within an acceptable range, 6-8, to achieve a stable growth balance among fish, plants, and nitrifiers (Hsiao & Sung, 2020; Kumar et al., 2023). If pH varies from the optimal range, it affects the nitrification process, fish metabolism, increases the risk of fish diseases, and hinders plant growth by reducing the nutrient absorption rate (Hsiao & Sung, 2020; Chandramenon et al., 2024).

c. Dissolved Oxygen (DO)

Dissolved Oxygen (DO) refers to the amount of free and non-compound oxygen present in the water (Thorarinsdottir et al., 2015:34; Lorenzo et al., 2019; Channa et al., 2024). DO level is a crucial parameter for indicating water quality (Eze & Ajmal, 2020). The maintained DO level helps aquaponics plants with root respiration, transpiration and root growth (Rakocy et al., 2006).

DO plays a major role in aquaponics fish growth and bacteria, and restricts the fungal growth and rotting of roots (Sallenave, 2016; Channa et al., 2024). DO intensity in the water is based on the fish type and water temperature (Eze & Ajmal, 2020). However, the required concentration of DO to keep good health and maximise the warm water fish is 5 ppm (parts per million) or 5 mg/L (milligrams per litre), whereas for cold-water fish it is 6.5 ppm or 6.5 mg/L (Sallenave, 2016). The ideal DO range for a fish is 4-5 ppm (Hsiao & Sung, 2020). Nevertheless, the aquatic species will go under stress if the DO concentration goes below 3 ppm, which causes disease and death (Eze & Ajmal, 2020; Hsiao & Sung, 2020). Kumar et al., 2023).

d. Water temperature

Water temperature in aquaponics is a major factor that influences fish and plant growth (Sallenave, 2016; Kumar et al., 2023).

Aquatic species depend on water temperature. The acceptable water temperature for warm water fish is 22 – 29 °C, whereas for cold water fish it is less than 18 °C (Chandramenon et al., 2024). The optimal temperature for fish is 18°C to 30°C, which is also acceptable for crop and nitrifier (Hsiao & Sung, 2020). However, tilapia can tolerate a wide range of water temperatures from 9 °C–42.5 °C (Obirikorang et al., 2021).

e. Light

Light is a critical requirement for plant growth and for the photosynthesis process to be carried out (Hsiao & Sung, 2020; Yanes et al., 2020). Sunlight availability for indoor plants poses a challenge; however, studies suggest that artificial lighting can effectively replace natural sunlight (Yanes et al., 2020; Ghandar et al., 2021; Gnanasagar & Vivek, 2020).

f. Air temperature

Air temperature influences fish and plant growth (Khaoula et al., 2021; Yang et al., 2023). Extreme temperature influences aquatic biological activity, photosynthetic rate, and transpiration rate. Also, it causes plant stress and hinders plant growth (Bhat et al., 2023; Morchid et al., 2024).

g. Carbon dioxide (CO_2)

CO_2 is included in plant respiration (Morchid et al., 2024).

h. Nitrification process

In water, ammonia can be in two forms: unionised ammonia (NH_3) and ionised ammonia (NH_4^+), together ($NH_3 + NH_4^+$), called total ammonia nitrogen (TAN), also known as ammonia (Francis-Floyd et al., 2009; Somerville et al., 2014; Espinal & Matulić, 2019:41; Lindholm-Lehto, 2023; Mohamed Ramli et al., 2024). Water temperature, pH, and salinity control the proportion between unionised and ionised ammonia (Lindholm-Lehto, 2023). Fish excrete liquid waste through gills or urine in the form of ammonia (Thakur et al., 2023).

Ammonia toxicity depends on water temperature and pH. Higher temperature and pH affect the fish's life (Francis-Floyd et al., 2009; Somerville et al., 2014; Lindholm-Lehto, 2023; Thakur et al., 2023). Ionised ammonia is not toxic to the fish, whereas unionised ammonia is (Pillay, 2004: 4; Espinal & Matulić, 2019: 41).

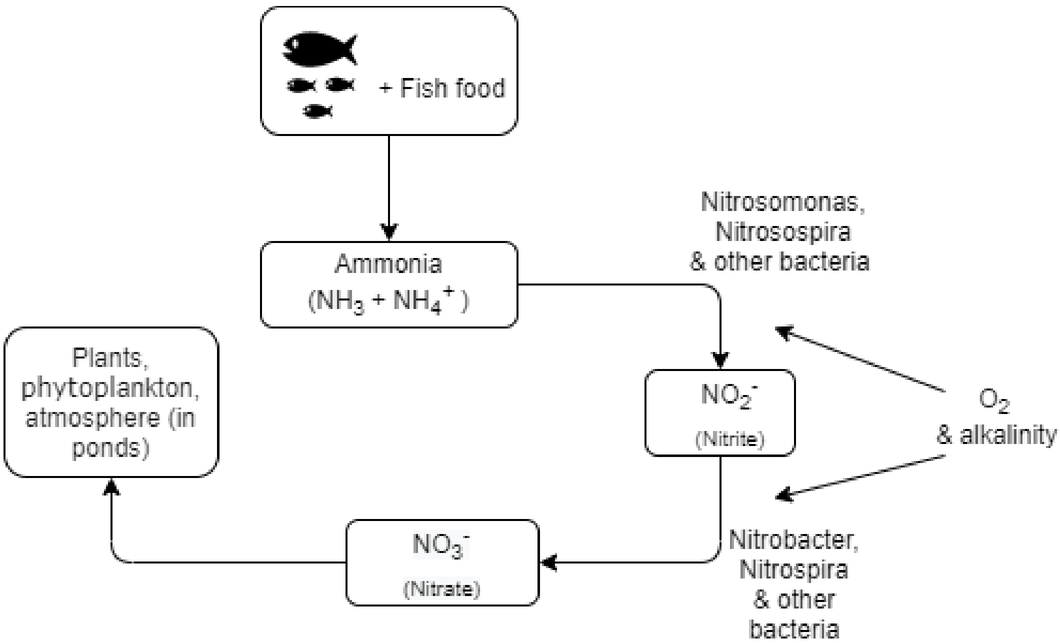


Figure 2.2: Nitrogen cycle in aquaponics (Francis-Floyd et al., 2009)

The nitrogen cycle is a biological process that helps to eliminate ammonia. The nitrogen cycle in aquaponics is shown in Figure 2.2. Ammonia is eliminated from the water by converting it into another form of nitrogen, such as nitrite (NO_2^-) and nitrate (NO_3^-), with the help of nitrifying bacteria, Nitrosospira, Nitrosomonas, Nitrospira, Nitrobacter, and other bacteria (Francis-Floyd et al., 2009; Prosser, 1989 citated in Espinal & Matulić, 2019:41; Lindholm-Lehto, 2023; Jiang & Liu, 2024; Kok et al., 2024). Nitrifying bacteria use oxygen and alkalinity to convert ammonia and nitrite into a less toxic byproduct, nitrate (NO_3^-). In the pond, nitrate is used as plant fertiliser, for microalgae (phytoplankton) or returned to the atmosphere.

For a healthy fishpond/tank, the total ammonia nitrogen (TAN) concentration should be maintained at less than 1 mg L^{-1} (Mohamed Ramli et al., 2024). If it is more than zero, the monitoring of UIA in TAN must be undertaken, as this is highly toxic to the fish. UAI that is more than 0.05 mg/L (ppm) is harmful to the fish and can even cause death. UIA concentration in water is determined by water pH and temperature (Thakur et al., 2023; Eneh et al., 2023).

Aquaponics systems involve three main organisms, namely: plants, fish, and bacteria. Each organism has specific tolerance ranges for key parameters, which can vary depending on the plant and fish species. Table 2.5 provides a summary of the specified organisms' tolerance range (Sallenave, 2016; Kurian et al., 2019; Singh et al., 2021). However, maintaining optimal ranges for these parameters can significantly enhance the overall yield of the aquaponics system. Table 2.6 shows the optimal water quality range of (Sallenave, 2016; Shafeena, 2016).

Table 2.5: Aquaponics fish, plant and bacteria water quality parameters and tolerance range

Organism Type		**Temp (°C)**	**pH**	**Ammonia (mg/litre)**	**Nitrite (mg/litre)**	**Nitrate (mg/litre)**	**DO ppm (parts per million) / (mg/L)**
Warm water fish		22-32	6-8.5	<3	<1	<400	4 - 6
Cold water fish		10-18	6-8.5	<1	<0.1	<400	6 – 8
Plant	**Leafy**	14- 20	5.5-7.5	<30	<1	-	>3
	In general	18- 30					
Bacteria		14-34	6-8.5	<3	<1	-	4-8

Table 2.6: Optimal water quality range of general and tilapia-based aquaponics systems

Type	**Temperature**	**pH**	**TAN**	**NO_2 Nitrite**	**NO_3 Nitrate**	**DO**
General Aquaponics System	65 - 85 ^{0}F (18.33- 29.44 ^{0}C)	6-7	<1ppm	<1ppm	5 - 150 ppm	>5ppm
Tilapia-based Aquaponics System	81 - 84 ^{0}F (27.22 – 28.88 ^{0}C)	7	<1ppm	<1ppm	5 - 150 ppm	>5ppm

It is important to monitor some water quality and environmental parameters to help maximise plant and fish growth in your aquaponics system (Hadi et al., 2022).

2.1.3.3 Aquaponics monitoring parameters

Water is a common medium for the three living organisms of an aquaponics system, namely: fish, plants and bacteria (Thorarinsdottir et al., 2015:33; Shafeena, 2016; Sallenave, 2016; Lennard & Goddek, 2019:124; Singh et al., 2021). Therefore, it is essential to continuously monitor and control water quality parameters, including nitrogen, pH, electrical conductivity (EC), dissolved oxygen (DO), total dissolved solids (TDS), temperature, and light conditions, to maintain ideal conditions for the healthy and optimal growth of these organisms (Timmons & Ebeling, 2010:49; Roberto, 2003:34; Resh, 2013:78; Somerville et al., 2014; Lennard & Goddek, 2019:126; Maulini et al., 2022; Wibowo et al., 2019; Rozie et al., 2020). Among these, water temperature is particularly critical and requires close monitoring (Ekanayake et al., 2022). Similarly, pH is directly and indirectly related to other water quality parameters, making its monitoring equally important (Saha et al., 2018).

In a newly set up aquaponics system, parameters need to be tested daily to make the required parameter value/s corrections at the earliest stage. If the nutrient cycle were balanced in the aquaponics system, it would only require weekly testing (Sallenave, 2016).

For optimal growth and productivity of aquaponics systems benefiting both fish and plants, key water quality parameters such as pH, dissolved oxygen, and temperature, along with environmental factors like light, humidity, and ambient temperature, must be monitored. Remedial actions should be taken promptly if any parameter deviates from the expected values (Sallenave, 2016; Hsiao & Sung, 2020; Yanes et al., 2020; Hadi et al., 2022).

Other water quality parameters, such as TDS and EC, have also been highlighted in various studies as essential for effective monitoring (Pappu et al., 2017; Saha et al., 2018; Yanes et al., 2020; Rozie et al., 2020). Monitoring and controlling can be done either manually or electronically (Shafeena, 2016; Manju et al., 2017; Hsiao & Sung, 2020).

Electrical Conductivity (EC): Electrical Conductivity (EC) measures the ability of water to conduct an electric current, which is directly correlated with salinity levels. The optimal EC range for fish in aquaponics systems is between 100 and 2000 µS/cm (Yanes et al., 2020). However, the broader acceptable range extends from 30-5000 µS/cm (Saha et al., 2018). A high EC reading typically indicates water pollution, which can adversely affect the aquatic environment. Additionally, the fish population is closely linked to EC levels, as higher densities of fish can influence salinity and, consequently, EC readings (Yanes et al., 2020).

Total dissolved solids (TDS): TDS levels represent the concentration of organic matter, dissolved materials, and inorganic salts in water. The ideal TDS level in water is 1000 mg/L. Exceeding this optimal range can create a toxic environment for aquatic organisms (Yanes et al., 2020).

Relative humidity: Relative humidity refers to the amount of moisture in the air. It is essential for plant growth as it helps plants thrive. The considerable relative humidity for plants ranges from 50% to 80%, although it may vary depending on the plant variety (Yanes et al., 2020; Morchid et al., 2024).

Ambient temperature: Ambient temperature significantly influences plant health. The optimal temperature range for most vegetables in aquaponics is between 18°C and 30°C (Yanes et al., 2020).

Aquaponics farming is a multidisciplinary field where knowledge about plants, fish and micro-organisms is required (Goddek et al., 2015; Channa et al., 2024). Good training, skills, and management will lead to successful aquaponics farming. Aquaponics daily management is essential as an aquaponics unit has three different living organisms, whilst the common medium is water (Goddek et al., 2015; Valiente et al., 2018). Management embraces fish feed, fish tank, grow bed, water flow and monitoring, and maintenance of the environmental parameters: pH, temperature, humidity level, water level and many more (Dutta et al., 2018; Valiente et al., 2018).

Aquaponics has many variables and complexities; thus, one needs to be meticulous in monitoring the chemistry throughout the circulating water to ensure optimal ratios and concentrations of nutrients. Ammonium is a very toxic component. It is thus imperative to watch it carefully. Water quality parameter reading is continuously required in aquaponics to check whether the system maintains a controlled environment or not (Goddek et al., 2015; Sallenave, 2016; Deshpande et al., 2024). The controlled environment guarantees the optimal growth of fish, vegetables and bacteria simultaneously.

Monitoring and controlling an aquaponics system manually/traditionally is time-consuming and might not be accurate (Shafeena, 2016; Naser et al., 2019; Channa et al., 2024). If there is any abnormality in the parameter reading, the value from the optimal value of the parameter needs to be adjusted to maintain the environment and keep it under control.

Hence, human intervention is intensively required to monitor and control these constantly changing values, which are critically required for plant and fish growth production. This shows the necessity of a smart aquaponics system to reduce the burden of human intervention, labour, and monotonous tasks (Goddek et al., 2015; Shafeena, 2016; Jerry, 2020; Raman & Vasmatkar, 2024).

Smart aquaponics is an integrated system that uses advanced technologies like IoT, machine learning (ML), and automation for real-time monitoring, controlling, and optimising aquaponics farming. Recent developments and trends include IoT devices for real-time system management and AI-driven machine learning algorithms to enhance sustainability and efficiency (Liu & Jiang, 2024).

2.1.4 Internet of Things

The term " Internet of Things" was devised by Kevin Ashton in 1999 (Corcoran, 2016; Mouha, 2021). The Internet of Things is a framework that provides a structure to interconnect physical devices, sensors, electronics, or additional technologies to collect and exchange data with other devices or systems over the internet (Mouha, 2021). As technology evolves, the definition of "Things" also changes (Gubbi et al., 2013). The "Thing" in IoT can be an object having a sensor installed in it that can collect data and transfer it across the network, which helps to implement, monitor and control operations without human involvement (Jamali et al., 2020:1; Mouha, 2021). The main aim of IoT is to monitor and control things/objects from anywhere in the world, which makes the devices "Smart" (Jamali et al., 2020:1; Maity et al., 2023).

The IoT integration in certain areas makes it more efficient, practical, safe and intelligent, such as smart agriculture, smart water, smart cities, smart cars, smart farming, smart homes, smart glasses, smart postal, precision farming, industries, health monitoring, education, security, media and many more (Vashi et al., 2017; Reddy et al., 2020; Ammayappan & Smys, 2020; Jamali et al., 2020:2; Mouha, 2021; Maity et al., 2023).

2.1.4.1 IoT architecture

IoT architecture encompasses a collection of physical objects, sensors, cloud services, actuators, communication layers, users, business layers and IoT protocols (Jamali et al., 2020:3). The integration of hardware and software over the network is grounded in the anticipated solution. Therefore, the implementation of IoT architecture may vary; depending on the study, it can be a three-layer, four-layer, five-layer or even seven-layer architecture (Vashi et al., 2017; Mouha, 2021; Kumar & Sharma, 2023). However, a widely accepted IoT technology architecture has three layers, namely, the perception layer, network layer and application layer (Lin & Shi, 2014; Jamali et al., 2020:3; Mouha, 2021; Kumar & Sharma, 2023; Prasetya et al., 2024). Figure 2.3 portrays the three-layered IoT architecture.

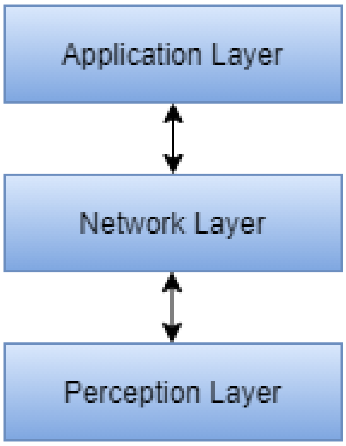


Figure 2.3: Three-layer IoT architecture (Mouha, 2021)

Perception Layer: The lowest layer in the standard IoT architecture. This layer involves various sensors, actuators and any physical devices. The primary purpose of this layer is to collect data from the environment (Jamali et al., 2020:3; Mouha, 2021).

Network Layer: This layer is responsible for transmitting data between the perception layer and the application layer. It also establishes a connection between other smart things, network devices, and servers (Jamali et al., 2020:3; Mouha, 2021).

Application Layer: This layer establishes the connection between the user and the application and provides services according to the user's needs (Jamali et al., 2020:4; Mouha, 2021).

2.1.4.2 IoT in aquaponics

Integrating technologies such as the internet, sensors, automation systems, robotics, and AI in agriculture creates smart agriculture. It aims to enhance crop quality and quantity while minimising manual labour (Kassim, 2020; Arjune & Kumar, 2022; Lynda et al., 2023). The applications of IoT in agriculture include weather monitoring, disease monitoring, soil condition monitoring and irrigation management (Kassim, 2020; Ismaili et al., 2024). The IoT smart devices can sense the variations in data, collect, store and send the data over the network (Kassim, 2020; Lynda et al., 2023). Various sensors, such as water temperature, Total Dissolved Solids (TDS), soil moisture, pH, air humidity, air temperature, precipitation, rain detection sensors, dew point sensors and so on, were used in agriculture for monitoring, controlling and data collection purposes (Saini & Saini, 2020; Kassim, 2020; Lynda et al., 2023; Ismaili et al., 2024).

Similarly, IoT technology in aquaponics enables continuous monitoring and autonomous control, optimising management and transforming it into a "smart aquaponics" system (Mohd Ali et al., 2021; Wan et al., 2022). The aquaponics farmers monitor the real-time data, which enables them to maintain optimal conditions for both fish and plant growth. Renewable energy, like solar energy, was also used to run electronic devices connected to the aquaponics unit (Murdan & Joyram, 2021; Mohd Ali et al., 2021).

According to IoT's three-layer architecture, the sensors, cameras and actuators are placed in the perception layers for monitoring and controlling the parameters. The parameters monitored in the aquaponics studies using sensors include water temperature, water level, pH, turbidity, electrical conductivity (EC), ammonia, nitrate, Total Dissolved Solids (TDS), plant growth condition, soil moisture, planting environment such as light intensity, temperature, Carbon Dioxide, etc. (John & Mahalingam, 2021; Udanor et al., 2022; Ekanayake et al., 2022; Wan et al., 2022; Abdullah & Mazalan, 2022; Mahmoud et al., 2023; Naputol et al., 2024; Prasetya et al., 2024; Abidin et al., 2024; Perumal et al., 2024). The key factors controlled in aquaponics studies include water circulation using a pump to regulate water level in the tank, ambient light for plant growth, automated fish feeding, heaters to maintain water temperature, and fans to regulate ambient temperature for cooling. These controls are essential to achieve optimum growth of both plants and fish (John & Mahalingam, 2021; Hadi et al., 2022; Wan et al., 2022; Mahmoud et al., 2023; Prasetya et al., 2024; Abidin et al., 2024).

The collected data from the perception layer and information from the application layer are transmitted between the layers using network communication technologies and protocols that belong to the network layer. For example, technologies are Wireless Fidelity (Wi-Fi), 5G communication, LongRange (LoRa), LoRaWAN, and Wireless Sensor Networks (WSNs). Protocols are Message Queuing Telemetry Transport (MQTT), Internet Protocol, and ZigBee (Zaini et al., 2018; Nichani et al., 2018; Wang et al., 2020; Ghandar et al., 2021; Wan et al., 2022; Silalahi et al., 2022; Alselek et al., 2022; Mahmoud et al., 2023; Abidin et al., 2024; Prasetya et al., 2024).

The collected data is processed in the application layer to provide the user with insight via a user interface. For prediction, machine learning, Artificial Intelligence or deep learning were used. Web interfaces, mobile applications or dashboards are used to monitor and control the real-time parameters from anywhere. Data is stored in the cloud and the database (Kyaw & Ng, 2017; Pasha et al., 2018; Barosa et al., 2019; Taha et al., 2022; Abdullah & Mazalan, 2022; Kim et al., 2022; Taha et al., 2022; Mahmoud et al., 2023; Abidin et al., 2024; Prasetya et al., 2024; Perumal et al., 2024).

The technologies that were used by Anila and Daramola (2024) in their Systematic Literature Review study in various aquaponics research are depicted in Figure 2.4.

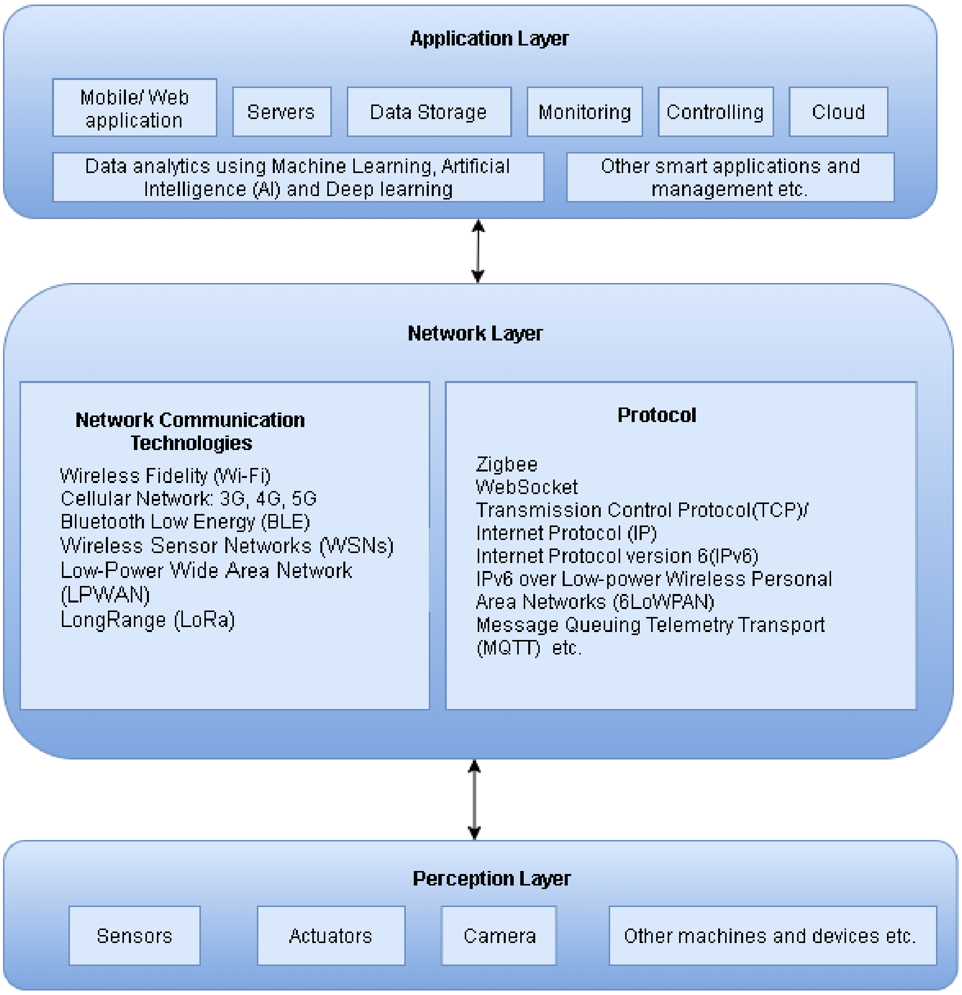


Figure 2.4: IoT architecture with various technologies in aquaponics studies (Anila & Daramola, 2024)

2.1.5 Machine Learning

Machine learning is a core area of Artificial Intelligence (AI) (Ray, 2019; Janiesch et al., 2021). Artificial intelligence is the ability of a machine to behave like a human and solve complex computer-based problems using large data in a very short time (Joshi, 2020:4). The term "Machine Learning" was coined by Arthur Samuel in 1959 (Joshi, 2020:4). Machine learning refers to a computer program's ability to learn from experience and enhance its performance or behaviour over time. According to Tom Mitchell, who in 1997 defined machine learning as: "A computer program is said to learn from experience *E* regarding some task *T* and some performance measure *P*, if its performance on *T*, as measured by *P*, improves with experience *E*" (Géron, 2019:3; Abdel-Fattah et al., 2021). Machine learning extracts actionable knowledge from data by applying machine-learning techniques (Lantz, 2013:10; Müller & Guido, 2016:1).

In machine learning, the functionality involves mapping input data to output results as predictions through a systematic machine-learning process.

2.1.5.1 Machine learning process

The general machine-learning process is described in 7 steps, which are data collection, data preparation, model selection, model training, model evaluation, hyperparameter tuning, and model deployment (Panigrahi et al., 2023). The general machine learning process is depicted in Figure 2.5.

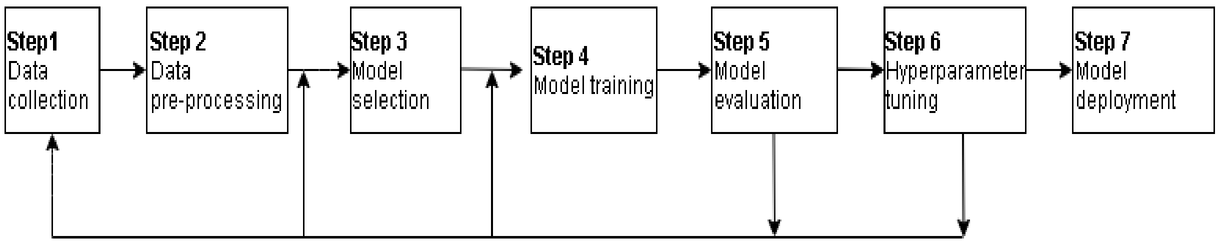


Figure 2.5: The general machine-learning process (Lantz, 2015:17; Panigrahi et al., 2023)

Step1. Data collection

Data collection is the primary step in the machine learning process (Alzubi et al., 2018). Data collection involves gathering relevant information from various sources based on the nature. Once the collected data is preprocessed, it is fed into a machine learning model to generate actionable insights (Lantz, 2015:16).

Step 2. Data pre-processing

Data quality plays a major role in machine learning (Badillo et al., 2020). Data pre-processing is intended to prepare the collected data for data analysis (Abdelaziz et al., 2025). The collected data may contain noisy, redundant or missing data and inconsistent data (Alzubi et al., 2018; Abdelaziz et al., 2025). During the data cleaning, insignificant or redundant data can be disregarded (Yang & Shami, 2020). Missing values and outliers can be treated by replacing them with calculated statistical measures, such as the mean, mode or median (Dangeti, 2017:11; Joshi, 2020:151). Once the data is cleaned, the next step is feature selection. Although a dataset may contain many features, it is crucial to select only those that are relevant to the study's objectives to ensure effective model training (Alzubi et al., 2018). Additionally, the features in the dataset may have different ranges, which can reduce the accuracy and performance of the model. To address this, data transformation techniques such as normalisation or standardisation are applied to scale the features appropriately, thereby improving model performance. To address it, the data transformation process, normalising or standardising, can be applied to the data. Normalising maintains data to a specific range, in standardising, the data have a mean of zero and a standard deviation of one, which helps improve the accuracy and performance of the model (Abdelaziz et al., 2025). Finally, the dataset is split into the train and test sets. The training set is used to train the model, while the test set is used to evaluate the model's ability to generalise unseen data (Panigrahi et al., 2023; Abdelaziz et al., 2025). Generally, the data is split into higher portion ratios to train the models, whilst a smaller portion is used to test the models.

Step 3. Algorithm selection

After the data is prepared for analysis, the researcher will most likely gain insight into what can be learnt from the collected data (Lantz, 2015:16). Trends and patterns in the data can be uncovered during the analysis of the data, and the type of algorithm that needs to be selected. Algorithms enable computers to learn behaviours and patterns based on the given data (Chitralekha & Roogi, 2021). The selection of the appropriate algorithm, such as supervised, unsupervised, semi-supervised and reinforcement learning, depends on the type of problem to be solved (Chitralekha & Roogi, 2021). The problems are, namely, classification, regression, anomaly detection, clustering and reinforcement (Alzubi et al., 2018). Selecting the most suitable machine learning algorithm can be challenging, as it directly affects prediction accuracy and overall model performance. Once a suitable machine learning algorithm is chosen, it represents the data in the form of a model (Lantz, 2015:17).

Step 4. Model training

The most critical phase in machine learning is model training. In this phase, the model will be trained using the training dataset to learn the trends in the given dataset. The prepared data is input into the selected machine learning algorithm to train the model and finally make predictions. (Panigrahi et al., 2023).

Step 5. Model evaluation

It is vital to assess the model's performance to understand how effectively the algorithm learns from experience and to estimate the accuracy of the model's results on unseen data (Lantz, 2015:17). Different metrics are used to evaluate the model. However, the choice of metrics depends on the algorithm selected.

Step 6. Hyperparameter tuning

Machine learning algorithms adjust the model parameters based on the given data during the training process. Model parameters focus on covering the input data to the desired output data. Whereas, other parameter types that are pre-configured before the training process is initialised and cannot change during the training process are known as hyperparameters. Hyperparameters are involved in building the structure of a model (Elgeldawi et al., 2021; Yu

& Zhu, 2020).

Hyperparameter tuning, or the process of finding optimal hyperparameters, is crucial because it involves developing methods to systematically and formally identify the best hyperparameter configurations. This process, which is considered an optimisation problem, facilitates better learning and understanding of the model's performance (Yu & Zhu, 2020).

Hyperparameter optimisation determines which hyperparameters to tune and systematically adjusts the hyperparameter values to evaluate the model's performance across various hyperparameter sets (Yu & Zhu, 2020). The main goal is to determine the best hyperparameter combinations effectively and efficiently (Yu & Zhu, 2020). This achieves minimum loss or maximum accuracy on a validation set. Fine-tuning a model's hyperparameters is vital for adapting a machine-learning model to different problems (Yu & Zhu, 2020; Yang & Shami, 2020; Elgeldawi et al., 2021).

The performance of the machine learning model changes based on the choice and values of its hyperparameters. However, it is also important to know how well a model can perform on unseen data. For the cross-validation, a statistical method is used to assess the machine learning model's accuracy. This will determine how well a model can perform on unseen data. One of the popular cross-validation methods is K-fold cross-validation (Elgeldawi et al., 2021).

Grid Search and Random Search are two hyperparameter optimisation techniques used to determine the optimal combinations of hyperparameters (Bischl et al., 2023).

Grid search performs an exhaustive search over a specified set of hyperparameters defined by the user and evaluates every possible combination of hyperparameter values using cross-validation. This method is popularly used to tune model hyperparameters to obtain the best combination for determining the best fit (Géron, 2019:79; Dangeti, 2017:286; Yu & Zhu, 2020; Bischl et al., 2023).

Random search is an improved version of grid search. It performs a randomised search over hyperparameters to find optimal combinations for the model under consideration. The random search is usually computationally intensive compared to the grid search (Yu & Zhu, 2020; Elgeldawi et al., 2021; Bischl et al., 2023). The search continues until the entire allocated budget is exhausted or the desired accuracy is achieved (Yu & Zhu, 2020).

Step 7. Model deployment

The final step in the machine learning process is model deployment. Once the model is performing well, it can be deployed for prediction purposes (Lantz, 2015:17; Dangeti, 2017: 12). When the model is in use for its intended task, it is essential to regularly assess whether the model is performing well with the new data and update it accordingly to ensure getting the optimised results (Pruneski et al., 2022).

In machine learning, specific terms are frequently used to describe various aspects of the field, models, and processes. Table 2.7 addresses the commonly used terms in machine learning.

Table 2.7: Machine learning terms

Term	Meaning	Source
Datasets	The data set is ideal when all the data is numerical and it does not contain any missing values.	(Joshi, 2020:22)
Model	A machine learning model is a mathematical representation(rule, formula, or equation) trained to identify data patterns.	(Fenner, 2019:8)
Entities	In machine learning, an entity represents a digital storage of data commonly stored in a CSV file format.	(Joshi, 2020:22)
Attribute	An attribute represents the column of an entity. A group of attributes is an entity.	(Géron, 2019:9; Joshi, 2020:22)
Data type	The stored format of an attribute in an entity uses different types. For example, integer, string, datetime, etc.	(Joshi, 2020:23)
Features	In machine learning, a feature means a set of attributes used for prediction. It may vary based on the context.	(Géron, 2019:9)
Predictors	In machine learning, the predictors are input variables that predict an output.	(Géron, 2019:9)
Labels	Labels are expected results, target variables or predictions from a trained algorithm. The features are used for prediction. Labelled data, together with input data to train an algorithm, produces a model in supervised learning.	(Géron, 2019:8)
Training dataset/training set/ training data	It is a dataset that is inputted into a selected machine-learning algorithm to train the model.	(Müller & Guido, 2016:17)
Test dataset/ test set/ test data	It is a dataset used to validate the accuracy of the model. It is not the same as a training dataset.	(Müller & Guido, 2016:17)
Overfitting	In machine learning, overfitting means the model performs well on the training dataset. However, the model does not perform well during the testing period or for generalising the model. This will lead to high variance. Variance is how the data is scattered from the average value.	(Lantz, 2013: 16; Géron, 2019: 28; Joshi, 2020: 50; Molin, 2021:653)
Underfitting	It is the opposite of overfitting, where the model performs poorly on the training dataset. This is because the model is too simple to learn the underlying structure of the data. This	(Géron, 2019:30; Molin, 2021:653)

	leads to high bias. Bias is an error due to the difference between the actual value and the predicted value.	

Machine learning techniques use algorithms to learn patterns from given data and provide decisions or predictions without being explicitly programmed (Mahesh, 2020; Obaido et al., 2024). Machine learning depends on several algorithms to solve specific problems. The performance of the algorithm depends on the nature of the problem that needs to be solved (Mahesh, 2020; Abdel-Fattah et al., 2021). However, the efficacy of machine learning is determined by the type and characteristics of input data and the performance of the learning algorithms (Sarker, 2021).

2.1.6 Types of machine learning

There are three main types of machine learning categories, namely: supervised, unsupervised and reinforcement learning (Géron, 2019; Shrestha & Mahmood, 2019; Ray, 2019; Joshi, 2020:10; Janiesch et al., 2021). However, some authors addressed semi-supervised learning in their machine-learning category (Géron, 2019:8; Sarker, 2021; Chitralekha & Roogi, 2021; Richardson et al., 2022).

2.1.6.1 Supervised machine learning

A supervised learning algorithm uses a labelled data set as a pair of inputs and expected output to train the model (Müller & Guido, 2016:2; Sarker, 2021). After training, the algorithm will learn a pattern (Mahesh, 2020). The pattern can apply to a new data set/test data set, which helps with prediction. Supervised learning algorithms are further divided into two: regression and classification (Ray, 2019; Janiesch et al., 2021; Sarker, 2021; Obaido et al., 2024). The regression algorithm predicts the numeric value. Classification algorithms classify the input data set into two or more classes (Russell, 2018:14; Janiesch et al., 2021). Forecasting, corn crop yield prediction, car price prediction, annual income prediction and trend analysis are some of the uses of regression algorithms. Image classification, cancer detection, spam filters, text classification, weather forecasting, and face recognition are some of the uses of classification algorithms (Müller & Guido, 2016:26; Géron, 2019:9; Ray, 2019; Sarker, 2021; Alnuaimi & Albaldawi, 2024). Differentiating between regression and classification tasks can be done by questioning if there is a pattern of continuity in the output. If continuity is identified in the possible outputs, this points towards a regression problem. The classification task, however, has completely different categorisations and thus shows no signs of continuity (Müller & Guido, 2016:26; Sarker, 2021; Alnuaimi & Albaldawi, 2024).

Various algorithms come under regression and classification problems. k-Nearest Neighbors, Linear Regression, Naïve Bayes, Logistic Regression, Support Vector Machines (SVMs), Linear Discriminant Analysis (LDA), Decision Trees and Random Forests are some of them (Müller & Guido, 2016:22; Géron, 2019:10; Ray, 2019; Sarker, 2021).

2.1.6.2 Unsupervised machine learning

Unsupervised learning algorithms learn by themselves using unlabelled datasets without any target variable or supervision provided (Dangeti, 2017:9; Fenner, 2019:445; Sarker, 2021). It requires finding hidden patterns and relations in the given data (Dangeti, 2017:9; Naeem et al., 2023). It involves a model that is fit for observations. In unsupervised learning, a data set of input objects is collected. Unsupervised learning then typically treats input objects as a set of random variables. Thereafter, a joint density model is built for the dataset (Ayodele, 2010b: 13-14).

Unsupervised learning provides the unknown output and uses an unlabelled dataset without a training dataset to find hidden patterns or structures of data in which no target variable exists (Dangeti, 2017:304; Géron, 2019:10; Ray, 2019; Janiesch et al., 2021).

Social network analysis, software fault prediction, segmentation of customers, search engine, data mining and knowledge extraction, etc., are some examples of unsupervised learning applications (Dangeti, 2017:304; Janiesch et al., 2021; Naeem et al., 2023).

Clustering, dimensionality reduction, self-supervised learning, density estimation, and association rules are five major types of unsupervised learning tasks (Dangeti, 2017:9; Géron, 2019: 10; Joshi, 2020:133; Ren et al., 2023; Obaido et al., 2024).

2.1.6.3 Semi-supervised learning

Semi-supervised learning is a combination of supervised and unsupervised learning (Géron, 2019:14; Sarker, 2021). It is suitable when there is insufficient labelled data and the dataset contains more unlabelled data than labelled data (Chitralekha & Roogi, 2021; Richardson et al., 2022). The model trains based on a small amount of labelled data and predicts on a large set of unlabeled data (Richardson et al., 2022). Using labelled data guides the model to learn the pattern and then make a prediction using unlabelled data (Chitralekha & Roogi, 2021; Richardson et al., 2022; Obaido et al., 2024). The two main types of semi-supervised learning methods are self-training and co-training. In the self-training method, the dataset is split into three parts such as train data, unlabeled data, and test data. The model trains using the training dataset and then makes predictions with the unlabeled data. Then, select the data points with the highest prediction probabilities and add them to the training dataset. These selected points will no longer be part of the unlabeled dataset. This process repeats until no more high-probability predictions remain. Finally, evaluate the model's performance using the test dataset. In the co-training method, the dataset is split into two views under the sufficiency and independence assumptions. Each view is sufficient to train a classifier, and the views are independent of each other. After training, predictions are made on the unlabeled dataset, and, according to each view, the high-confidence unlabeled data are added to the new training set. Repeat the process until the predictions are optimised (Richardson et al., 2022; Ning et al., 2023). The applications of semi-supervised learning are machine translation, speech detection, fraud detection, text classification and so on (Sarker, 2021; Richardson et al., 2022).

2.1.6.4 Reinforcement learning

Reinforcement learning is one of the machine learning categories in which an intelligent program, learning system or software agent learns from environmental interaction feedback and takes action to move to the next stage to achieve a goal (Kaelbling et al., 1996; Sutton & Barto, 2018:1; Nandy & Biswas, 2018:1; Géron, 2019:14-15; Elguea-Aguinaco et al., 2023; Alnuaimi & Albaldawi, 2024). If the feedback is positive, this is classified as a reward, and if the feedback is negative, it is known as a punishment or penalty (Nandy & Biswas, 2018:1; Géron, 2019:14; Obaido et al., 2024; Alnuaimi & Albaldawi, 2024).

Reinforcement learning depends on trial-and-error experiments (Alnuaimi & Albaldawi, 2024). The interaction deals with the environment in which real-world scenarios are portrayed. Taking the environment into consideration brings about a lot of factors, and more learning is thus required (Nandy & Biswas, 2018:2; Alnuaimi & Albaldawi, 2024). The agent trains itself from the learning occurring in the environment. Due to the volume of information the agent learns, it can have different paths to choose from.

The main elements of reinforcement learning are agent, environment, action, state, reward and policy (Dangeti, 2017:361-362; Sutton & Barto, 2018:6; Jia & Wang, 2020; Sarker, 2021). An agent is a model that is being trained via reinforcement learning. The environment is the training situation that the model must optimise within. An action is a possible step that the model can take. A state can be described as a condition or current position given by the model. Reinforcement learning focuses on increasing the aggregate and collective reward, i.e. all the rewards accumulated and received by the agent from the environment, instead of the immediate reward received from the current state. The software agent understands the current state of the environment and takes action to move to the next stage. It also determines how an agent will behave at any given time. In a nutshell, a policy is a decision-making process that allows changes from the action taken to the present state (Dangeti, 2017:361-362; Sutton & Barto, 2018:6; Nandy & Biswas, 2018:54; Jia & Wang, 2020; Elguea-Aguinaco et al., 2023). The objective of reinforcement learning is to create an optimal or close to optimal policy based on the rewards received.

Markov designed a framework to simplify the manner of illustrating features of an intelligence problem. The Markov decision process (MDP) framework is used to define the interaction between the environment and the learning agent in terms of rewards, actions, and states (Sutton & Barto, 2018: 13; Jia & Wang, 2020; Elguea-Aguinaco et al., 2023).

Compared to supervised and unsupervised learning, reinforcement learning is used to design optimal or near-optimal policies based on rewards received (Dangeti, 2017:359, Sutton & Barto, 2018:1; Jia & Wang, 2020).

The proposed study aims to address a regression problem by predicting numerical values from unseen data. Hence, various supervised-based learning regression models were explored, namely linear regression, random forest and eXtreme Gradient Boosting (XGBoost), K-Nearest Neighbors (KNN), and a multilayer perceptron (MLP). These algorithms were explored because they represent the various regression methods, such as linear, tree-based ensembles, instance-based, and deep learning. This allows for a comprehensive comparison between the algorithms and identifies the most suitable algorithm for the problem to be solved.

2.1.7 Linear regression

Regression encompasses the relationship between the value to be predicted and one or more predictors. This model is used for representing the relationship between one or more numeric input variables and one output variable. The input variable is also known as the independent variable or predictor, whereas the output variable is known as the dependent variable/predictor. The relationship between the independent and dependent variables is assumed to follow a straight line. Regression equations model data using a slope-intercept format. Regression analysis usually models complex relationships among data elements. It is also used to estimate the impact of a treatment on an outcome. Regression analysis is a pool of many methods that can be adapted to almost any machine-learning task. Linear regression is the most basic regression models that use straight lines. If there is only a single independent variable, this is known as simple linear regression, otherwise, it is known as multiple regression. Both the simple linear regression and multiple regression models take it that the dependent variables are continuous (Ray, 2019; Joshi, 2020:34).

Regression analysis aims to build mathematical models that explain the existing relationships between variables (Seber & Lee, 2012:2; Roustaei, 2024).

2.1.7.1 Simple linear regression

Simple linear regression is one of the simplest forms of regression, which uses the input and output variables as a dataset. If the relationship between the input and output variables is linear, the dataset can fit into a straight line (Roustaei, 2024). For this to be achieved, it uses the formula below:

$$y = \alpha + \beta x \tag{2.1}$$

Where: y = dependent variable; x = independent variable; α = intercept (the point where the line crosses the 'y' axis); β = slope (the slope b indicates how much the line rises for each increase in x).

The regression analysis is aimed at finding the estimated value for 'α ' and 'β' (Lantz, 2013:163; Ray, 2019; Roustaei, 2024). To find the optimal estimated values of the intercept 'α ' and slope 'β', the ordinary least squares (OLS) regression estimation method is used. The aim is to minimise the sum of the residual, which is the sum of the squared error. In this regression estimation method, the intercept and slope are chosen in a certain way to minimise the sum of the squared errors. These errors refer to the difference between the predicted dependent variable and the actual dependent variable (Lantz, 2013:164; Roustaei, 2024).

2.1.7.2 Multiple linear regression

Multiple regression is an extension of simple linear regression (Roustaei, 2024). Multiple linear regression has a many-to-one relationship between many input (independent) variables and one output (dependent) variable (Ray, 2019; He, 2023).

The multiple linear regression is represented in the following equation:

$$y = \alpha + \beta 1x1 + \beta 2x2 + \beta 3x3 + \dots\dots.. + \beta ixi + \varepsilon \tag{2.2}$$

Where: y = dependent variable; x= independent variable; α = intercept (the point where the line crosses the 'y' axis); β= slope (the slope b indicates how much the line rises for each increase in x); ε represents the residual (error); and i represents the total number of features.

Both simple and multiple linear regression have the same goal, which is to determine the values of the coefficient that reduce the prediction error of a linear equation (Priya, 2021). The dependent variable y represents the sum of an intercept term added to the product of the estimated slope 'β' value and the independent variable 'x' value for all 'i' features (Lantz, 2013:169; He, 2023; Roustaei, 2024).

The multivariate technique provides insight into the relationship between the set of independent variables and dependent variables. It also sheds light on the relationship between the independent variables using multiple regression, partial correlation, and tabulation techniques (Ray, 2019).

2.1.8 Ensemble learning

Algorithms generate a model using training data through learning; hence, the algorithms are learners. Linear regression, decision tree, logistic regression, neural network, etc., are some examples of individual learners or predictors. Ensemble learning is a technique that generates a model using a collection of single learners/base learners or weak learners (Zhou, 2012:15; Géron, 2019:191; Joshi, 2020:60; Thomas & Gupta, 2020). Ensemble learning algorithms are known as ensemble methods (Géron, 2019:191). Weak learners together build a strong learner who can perform predictions more accurately and has a better ability to generalise (Joshi, 2020:60; Li et al., 2020).

Different types of ensemble methods are Bagging, Boosting, Stacking, and Voting (Géron, 2019:191; Zhang et al., 2022; Mahajan et al., 2023; Khan et al., 2024).

Bagging: The name bagging is an acronym for "Bootstrap aggregation". The bagging method is applicable for both classification and regression problems. This method generates several weak learners in parallel, which are independent, thereafter, averaging the outcome of each weak learner if the solution belongs to the regression problem. In the alternative, find the majority vote for the classification problem (Breiman, 1996; Géron, 2019:195-196; Joshi, 2020:62).

Boosting: The boosting ensemble method is used for improvement. This method generates multiple weak learners in sequence (Géron, 2019:201). The first weak learner generates a model with a training dataset. The second weak learner checks the outcome of the first weak learner. If the first weak learner provides poor performance, then the second weak learner selects the training data to reduce the error of the previous model. This process is continued until it reaches the desired result. Thus, the stronger learner model will be generated using the improved weaker learners (Dangeti, 2017:52; Géron, 2019:201; Joshi, 2020:62). Gradient Boosting, Extreme Gradient Boost (XGBoost), and Adaptive Boosting (AdaBoost) are examples of boosting algorithms. Boosting may cause overfitting, and this method takes time, compared to the bagging ensemble method, because of the sequencing process (Dangeti, 2017:52; Joshi, 2020:62).

Stacking: Stacked generalisation, known as stacking, is an ensemble method used to accomplish generalisation accuracy by minimising the generalisation error when combining various generalisers (Zhang et al., 2022). This method uses various machine learning algorithms to generate predictions using a training set. Later, the predicted outputs will be an input for the final predictor or meta-learner to train the model and provide a final prediction (Lantz, 2013:338; Naimi & Balzer, 2018; Géron, 2019:210; Zhang et al., 2022; Mahajan et al., 2023).

Voting: Voting involves summing the predictions for classification and averaging the predictions for regression. Hard voting and soft voting are two types of voting in classification problems. Hard voting selects the prediction with the most votes, whereas soft voting combines the probabilities from each model and chooses the prediction with the highest overall probability (Mahajan et al., 2023; Khan et al., 2024).

2.1.8.1 Random Forest

A random forest is a supervised machine learning algorithm and is an ensemble or collection of decision trees, depending on the ensemble technique (Khan et al., 2024; Zhao et al., 2025).

A decision tree is composed of nodes and edges. To form a decision tree structure, the dataset needs to be split into smaller datasets based on the feature value. Dataset splits take place in a node. A node represents a decision point where the feature value is selected from the dataset to split and perform testing on it. There are different types of nodes, namely: root nodes, internal nodes and leaf nodes, which are connected by edges (Lantz, 2013:120; Song & Lu, 2015; Prajwala, 2015; Obaido et al., 2024).

The random forest algorithm results in a prediction for the regression problem and the category/class for the classification problem. Random Forest can deal with continuous variables and categorical variable datasets for regression problems and classification problems, respectively. A single decision tree's drawback is overfitting; however, the collection of decision trees and the aggregated result reduce overfitting by changing high variance to low variance. Random forest is a solution for overfitting (Ray, 2019; Molin, 2021:653).

The first step in the random forest process randomly selects a subset of features. In the decision tree formation split, the feature selection will be from the selected subset (Breiman, 2001; Zhou, 2012:58; Mienye & Jere, 2024; Khan et al., 2024). Thereafter, different decision trees will use the bagging method, which trains different decision trees in parallel and aggregates the results (Molin, 2021:653). In the regression algorithm, the final prediction result is the average of the output results of all decision trees, whereas in classification, it is determined by the majority voting method. The random forest algorithm is given in Table 2.8 (Wei, 2023; Mienye & Jere, 2024).

Table 2.8: Random forest algorithm (Mienye & Jere, 2024)

Step 1	for *i* = 1 to *T* do
	1.1 Randomly sample n instances from D with replacement
	1.2 Randomly select m features from the total p features (where $m<p$).
	1.3 Build decision tree h_i based on the sampled instances and attributes.
	end for
Step 2	To make predictions for a new instance x:
	if a classification task, then:
	$f(x) = \arg \max_c \frac{1}{T}\sum_{i=1}^{T} I\ (h_i\ (x) = c)$
	else if regression task then
	$f(x) = \frac{1}{T}\sum_{i=1}^{T}[h_i\ (x)]$
	end if

Where: n = number of samples; T = the number of decision trees in the random forest model; p = total number of features; m = randomly selected features; h_i = single decision tree; c = output of class; $I(.)$ = an indicative function; $arg\ max_c$ = select the class c corresponding to the highest vote; $f(x)$ = majority vote across trees (classification); $f(x)$ = average of tree predictions (regression).

In the bagging method, the correlation between the trained decision trees for the prediction can be high due to a strong feature selection at the node by all the trees. This will limit the improvement of prediction accuracy. Since the decision trees are not correlated in random forests, they can improve the prediction accuracy (Mekonnen et al., 2020). The random forest can handle missing and noisy data, and it performs well in most problems.

2.1.8.2 eXtreme Gradient Boosting (XGBoost)

Gradient boosting is one of the boosting algorithms. This algorithm develops new base learners or weak learners in a sequence and accumulates them into an ensemble. This method tries to reduce the errors of the preceding models (Géron, 2019:205; Mokhtar et al., 2022; Khan et al., 2024).

Chen and Guestrin (2016) proposed an improved gradient boosting decision tree algorithm, which is eXtreme Gradient Boosting (XGBoost). The difference, however, is that XGBoost has far better performance and speed due to its efficient utilisation of the CPU core of the machine and less complexity (Ramraj et al., 2016; Parsa et al., 2020). It takes a multithreaded method instead of a sequential one. XGBoost is a supervised machine-learning algorithm for tree boosting, and it is also scalable and quick to execute (Chen & Guestrin, 2016; Mitchell & Frank, 2017; Desdhanty & Rustam, 2021). This makes it suitable for both regression and classification problems (Pan, 2018). XGBoost supports arbitrary differentiable loss functions, together with the prevention of overfitting (Mitchell & Frank, 2017; Desdhanty & Rustam, 2021).

The XGBoost model generates a weak learner or decision tree (DT) in each iteration and predicts the values. Each iteration uses the previous result to boost the current result. The result is generated by accumulating weak/base learners. Assume a dataset D= $\{(x_i, y_i)\}$ ($|D|$ = n, $x_i \in \mathbb{R}^m$, $y_i \in \mathbb{R}$) with *n* examples and *m* features (Chen & Guestrin, 2016; Li et al., 2020).

Weak learner/ decision tree representation:

$$F = \{f1, f2, f3 \ldots \ldots . fm\} \tag{2.4}$$

Where: F is a feature; f is a base learner/ weak learner/decision tree; m is the total number of features.

The main task of the XGBoost model is to build *t* trees so that the predicted value $\hat{y}_i^{(t)}$ up to the *t*th tree (Li et al., 2020).

Predicted *t*th tree value:

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) \tag{2.5}$$

Where: $f_k(x_i)$ is *k*th decision tree score in *i*th observation.

The mathematical derivation is given below:

$$\hat{y}_i^{(0)} = 0 \tag{2.6}$$

$$\hat{y}_i^{(1)} = f_1(x_i) = 0 + f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i)$$

$$\hat{y}_i^{(2)} = f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i)$$

$$\ldots\ldots.$$

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i)$$

Where: $\hat{y}_i^{(t)}$ is the predicted value of the *i*th iteration; $\hat{y}_i^{(t-1)}$ is the total predicted value from the previous iteration; $f_t(x_i)$ is the decision tree result of *i*th round.

However, it is also important to consider how to split the leaf nodes, how to determine the leaf nodes' predicted value on each decision tree, and how each decision tree connects to the previous decision tree. All these are determined by the Objective function (Li et al., 2020).

a. Objective function

The observation function helps to check how well the model can fit into a given sample dataset. Fewer errors means the best fit. The main aim when creating a model is to minimise the error. In XGBoost, the objective function has two parts: the loss function and the penalty/regularisation term, respectively. The loss function helps to prevent the complexity of the model and evaluates how well the model can predict based on the training data. The regularisation term helps reduce overfitting (Chen & Guestrin, 2016; Pan, 2018; Li et al., 2020; Li et al., 2021).

The model objective function is shown below:

$$Obj(\theta) = L(\theta) + \Omega(\theta) \tag{2.7}$$

Where: L is a loss function; Ω is a regularisation term.

The expanded expression is given below:

$$minL^{(t)}\left(y, \hat{y}^{(t)}\right) = min\left(\sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{k=1}^{t} \Omega(f_k)\right) \tag{2.8}$$

Where: $l\left(y_i, \hat{y}_i^{(t)}\right)$ is a loss function which measures the difference between the actual value (y_i) and the predicted value $\hat{y}_i^{(t)}$; *t* is the number of trees; $\sum_{k=1}^{t} \Omega(f_k)$ is the regularisation term, which measures the complexity of the whole model.

The regularisation term is defined as:

$$\Omega(f_k) = \gamma T_k + {}^{1}/_{2}\, \lambda \sum_{j=1}^{T_k} w_{kj}^2 \tag{2.9}$$

Where: ***Ω***= Regularisation term to evaluate the complexity of the model and to prevent being too difficult; $\Omega(f_k)$ is an objective function to avoid overfitting; f_k is the *k*th decision tree; The parameter γT_k is used to control the number of leaf nodes T, whereas $\lambda\ is$ used to control the weight of the leaf node *j*. T_k is the number of leaf nodes in the *k^th^* tree;

w_{kj} is the result of the *j*th leaf node in the *k*th tree.

To optimise the objective function, substitute the predicted value $\hat{y}_i^{(t)}$ of the ith sample in the tth iteration in the objective function. The simplified objective is given below:

$$minL^{(t)} = min\left(\sum_{i=1}^{n}\left[g_i\, f_t\,(x_i) + \frac{1}{2}\, h_i\, f_t^2(x_i)\right] + \Omega(f_k)\right) \tag{2.10}$$

Where: g_i is the first derivative of the loss function; h_i is the second derivative of the loss function. $\Omega(f_k)$ is a regularisation term.

b. Build the next learner

To obtain an even fixed tree structure q(x) of leaf node j, compute the weight w_j^* of *j* leaf (Chen & Guestrin, 2016).

The formula is below:

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda'} \tag{2.11}$$

Where: $I_j = \{i | q\ (x_i) = j\}$ as the set of leaf nodes *j* in the decision tree; λ' is the regularisation parameter; g_i and h_i represent the first and second derivatives of the loss function, respectively.

To evaluate the quality of the tree structure (q) use the scoring function.

$$\tilde{L}^{(t)}(q) = -\frac{1}{2}\sum_{j=1}^{T} \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \tag{2.12}$$

Where: $\tilde{L}^{(t)}(q)$ is a scoring function, which is used to measure the quality of the tree structure q.; λ is the regularisation parameter; First and second loss function derivatives are g_i and h_i ; γT to control the number of leaf nodes T.

c. Best Split

The XGBoost model used a greedy algorithm to divide the leaf node into left and right nodes and iteratively add the branches (Li et al., 2021).

The following formula is used to find the best split on any given node.

$$L_{(split)} = \frac{1}{2}\left[\frac{\left(\sum_{i \in I_L} g_i\right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i\right)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{\left(\sum_{i \in I} g_i\right)^2}{\sum_{i \in I} h_i + \lambda}\right] - \gamma \tag{2.13}$$

Where: I_L = Left node of the sample set after the split of leaf node$I$; I_R = Right node of the sample set after the split of leaf node I; g_i and h_i are the first and second derivatives of the loss function; γ is the regularisation parameter.

The XGBoost performs well compared to other tree-boosting methods due to the regularised loss function, which controls overfitting. It can reduce the weight of each tree using a given constant, which scales down the impact of individual trees on the last score, and column sampling performs the same as random forest (Pan, 2018; Desdhanty & Rustam, 2021).

2.1.9 K-Nearest Neighbor (KNN)

KNN is a non-parametric algorithm because it does not make assumptions about the elementary data it uses (Lantz, 2013:67; Alzubi et al., 2018; Ray, 2019; Taunk et al., 2019). It was introduced by Fix and Hodges in 1951 (Imandoust & Bolandraftar, 2013; Taunk et al., 2019). KNN is also known as instance learning or lazy learning because the KNN model does not learn during training; instead, the model observes and stores the training data and memorises the dataset. In the testing phase, compare the test observation with the training observation (Dangeti, 2017:187; Taunk et al., 2019). KNN solves problems based on neighbouring training examples in a given region (Taunk et al., 2019; Abdel-Fattah et al., 2021). KNN can be applied to both regression and classification problems without making any changes in the architecture (Alzubi et al., 2018; Joshi, 2020:38). In KNN, 'K' denotes the number of neighbours that need to be considered to predict the test data point (Bhatia & Vandana, 2010). The nearest neighbour is the point with the lowest distance between the training and sample points (Bhatia & Vandana, 2010). To measure the distance between the query point (target) and cases from the example sample (training data points), a metric known as the distance metric is used. Euclidean distance is a popularly used distance metric to calculate the nearest neighbours by measuring the similarity between two distances (Lantz, 2013:70; Imandoust & Bolandraftar, 2013; Taunk et al., 2019).

Euclidean distance dist (p,q) between two data points p and q is calculated (Lantz, 2013:70; Taunk et al., 2019; Sudheer et al., 2022).

$$\mathrm{dist}(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \tag{2.14}$$

Where: p and q are the examples to be compared, each having n features. The term p_i refers to the value of the i^{th} feature of example p, while q_i refers to the value of the i^{th} feature of example q.

Once the value of K is selected, it can make the prediction based on KNN examples. In classification, the prediction for a new data point is considered by its closest neighbour(s) in the training set (Müller & Guido, 2016:24). Whereas in regression, the predicted value will be the average of its K- nearest neighbours(Taunk et al., 2019; Imandoust & Bolandraftar, 2013). The formula is shown below (Imandoust & Bolandraftar, 2013; Sudheer et al., 2022).

$$y = \frac{1}{K}\sum_{i=1}^{k} y_i \tag{2.15}$$

Where: y_i is the i^{th} case of the example sample (nearest neighbour); y is the predicted value for the query point, calculated as the average of the y_i values of the k nearest neighbours; k is the number of nearest neighbours considered.

Other distance functions are Manhattan distance, Chebyshev distance, Mahalanobis distance, Bhattacharyya distance, Hamming distance, Cosine distance, Minkowski distance and so on (Lantz, 2013:273; Joshi, 2020:134; Zhang et al., 2023). The selection of distance is based on the problem that needs to be solved. KNN is simple, easy to implement, and builds a model cheaply (Ray, 2019; Joshi, 2020:38).

2.1.10 Multi-Layer Perceptron (MLP)

Deep learning is a subset of machine learning. The implementation of artificial neural networks into deep learning generates a model for supervised or unsupervised problems using structured and unstructured datasets, respectively. Video, image, voice, etc., are examples of an unstructured dataset (Dangeti, 2017:267; Janiesch et al., 2021). There is a significant improvement in the performance of classifiers when deep learning is used, as opposed to more conventional machine learning methods (LeCun et al., 2015; Mathew et al., 2021:600). Deep learning is capable of learning from a large amount of data (Alzubaidi et al., 2021). Deep learning techniques have achieved great strides and a lot of success in pattern recognition, speech recognition, handwritten classification, image analysis, Natural Language Processing (NLP), and many more (Liu et al., 2017; Alzubaidi et al., 2021).

An animal's body has millions of neurons. Neurons are biologically specialised to send and receive electrical signals called action potentials between other neurons through the connections known as synapses (Awad & Khanna, 2015:129; Géron, 2019:279).

A neural network is a union of neurons (Awad & Khanna, 2015:130). Aleksander and Morton (1990), cited in Haykin (1994:24), defined the neural network as:

A neural network is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for sorting experiential knowledge and making it available for use. It resembles the brain in two respects:

1. *Knowledge is acquired by the network from its environment through a learning process.*
2. *Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge.*

The artificial neuron model replicates biological neurons. An artificial neural network consists of many interconnected processors called neurons (Vui et al., 2013; Janiesch et al., 2021; Emmanuel et al., 2022). Artificial neurons are also known as nodes or units (Lipton et al., 2015; Janiesch et al., 2021). Figure 2.6 depicts the architecture of the neural network.

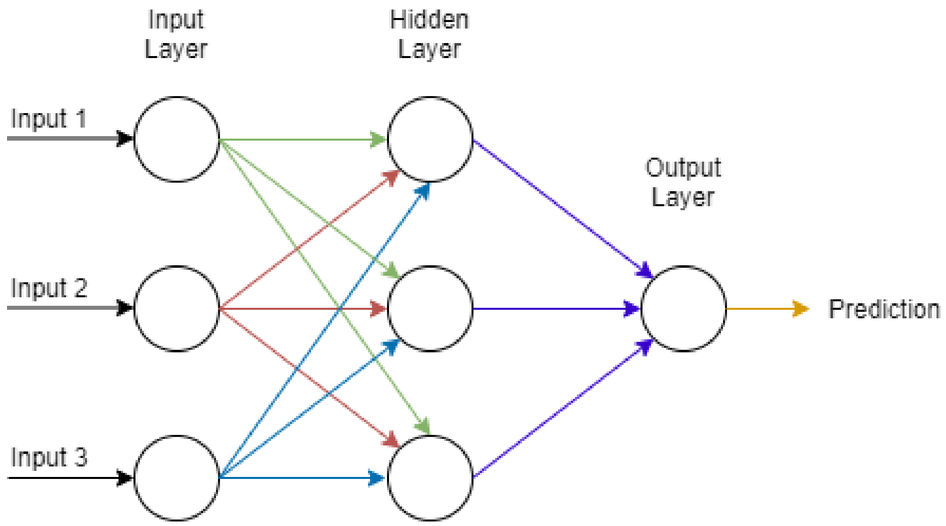


Figure 2.6: A neural network architecture (Nielsen, 2015:11)

An ANN consists of one input layer, one or more hidden layers and one output layer. Input layer nodes receive input signals from the environment. The layers/s between the input layer and output layer are known as hidden layer/s, which are neither input nor output layers. The input layer nodes are connected to the hidden layer nodes, which are neurons. The last layer is an output layer, which is generated by final nodes that can provide the result or prediction (Nielsen, 2015:11; Taud & Mas, 2018:454; Zaras et al., 2022:17; Abdolrasol et al., 2021).

Artificial Neural Networks became popular after the introduction of a computational model for neural network activity using propositional logic by neurophysiologist Warren McCulloch and the mathematician Walter Pitts in 1943. This model explains how biological neurons/ artificial neurons activate based on the given inputs and perform complex computations to provide an output. Also, it is possible to build a network of artificial neurons (Walczak, 2018: 121; Géron, 2019:278-281).

In the late 1950s, Frank Rosenblatt proposed a computational model known as the perceptron (Rosenblatt, 1962 cited in Bishop, 1995:98; Awad & Khanna, 2015:128; Nielsen, 2015:2-3; Wang & Raj, 2017; Rosenblatt, 1958 cited in Walczak, 2018:121; Géron, 2019:281). A perceptron is a single logic unit in an artificial neural network. A simple perceptron is identified as an ANN (Walczak, 2018:121). A multilayer perceptron (MLP) is an ANN composed of one or more layers of neurons or multiple perceptrons (Obiora et al., 2023). It is also known as a multilayer feedforward neural network (Lai et al., 2022).

The processing of information takes place through neurons. Each neuron contributes to the operation and functioning of a neural network. In the input layer, nodes receive input signals in the machine learning model that will be featured in. The nodes in the input layer will connect to the neurons in the hidden layer. Each neuron performs a function that includes two activities: Backpropagation isadded together with a bias to arrive at the sum of the net input. Secondly, calculate the activation function using the sum of the net input result to generate the output signal or result or prediction. The activation function takes a single number and performs a certain fixed mathematical functional mapping on it (Dangeti, 2017:243). The current layer function output will be the input for the next layer, and the decision of whether the neurons need to be fired or not. It is used to learn and model complex datasets (Zhou, 2012:8; Vui et al., 2013; Awad & Khanna, 2015:129; Dangeti, 2017:2, 268; Zaras et al., 2022:18-19).

Sigmoid function, Rectified Linear Unit (ReLU) function, Exponential Linear Unit (ELU), SoftMax, Tanh, Hyperbolic tangent sigmoid function, and Linear are some of the activation functions. The selection of the activation function is critical, which encompasses the generation of other neurons for the network performance and accuracy (Haykin, 1994:36; Zhou, 2012:7; Awad & Khanna, 2015:130; Dangeti, 2017:242-243; Géron, 2019:288; Zaras et al., 2022:19-21). A basic artificial neuron is shown in Figure 2.7.

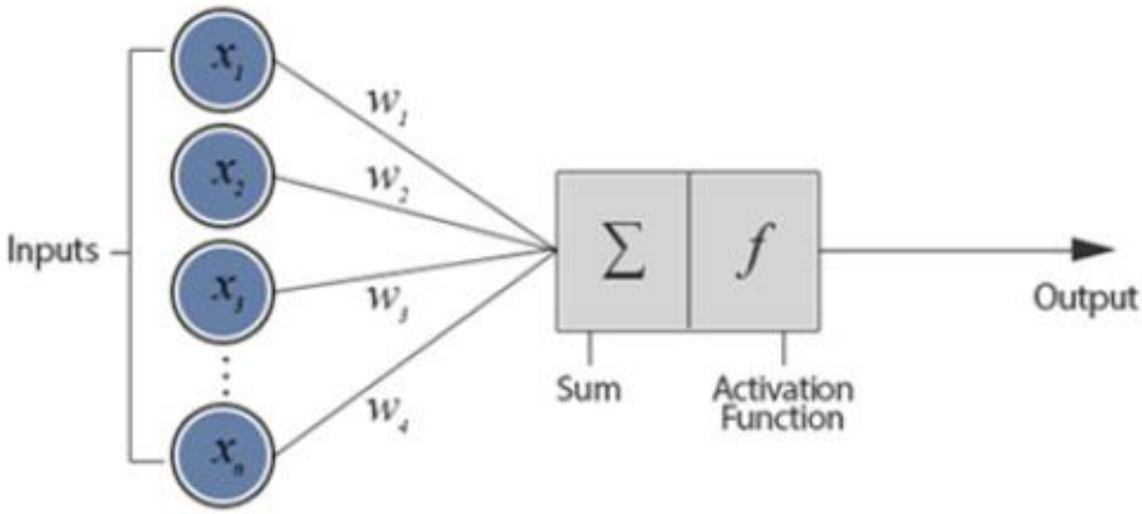


Figure 2.7: Artificial neuron (Dangeti, 2017:241)

The mathematical representation of neuron operation is:

$$Y = \theta\left(\sum_{i=1}^{n} \mathrm{WiXi} + \mathrm{b}\right) \tag{2.16}$$

Where: output Y = Active function (Sum of net input + bias); θ =activation function; Xi= input features; Wi= respective weight; b= bias.

a. Sigmoid

The sigmoid function is used in logistic regression to squash the real-valued number between 0 and 1. The mathematical representation of the sigmoid activation function is represented in Equation (Dangeti, 2017:243; Géron, 2019:144; Misra & Dinker, 2025).

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.17}$$

Where: $\sigma(x) =$ sigmoid function; x = input value (feature); *e* = Euler's number.

b. Hyperbolic Tangent (Tanh)

Tanh is a type of activation function and is very similar to a sigmoid function (Rasamoelina et al., 2020). Tanh squashes a real-valued number between -1 and 1 (Dangeti, 2017:243; Géron, 2019:288; Joshi, 2020:45).

The function can be represented as (Géron, 2019:288; Rasamoelina et al., 2020).

$$tanh(x) = 2\sigma\,(2x) - 1) \tag{2.18}$$

Or

$$tanh(x) = \frac{2}{1 + e^{-2x}} - 1$$

Where: $tanh(x)$ = tanh function; $\sigma(x)$ = sigmoid function; x = input value (feature); *e* = Euler's number.

c. A rectified linear unit (ReLU)

ReLU is a simple nonlinear activation function which is computationally efficient. It performs well compared to Sigmoid or Tanh due to the convergence property, improves model computational speed, and fixes the vanishing gradient problem (Dangeti, 2017:243; Habibi Aghdam et al., 2018:74; Géron, 2019:288; Joshi, 2020:122). ReLU is linear for all positive values and zero for all negative values. However, the output values range from 0 to infinity (Rasamoelina et al., 2020).

The representation is given below (Habibi Aghdam et al., 2018: 74; Géron, 2019:288).

$$ReLU(x) = \max(0, x) \tag{2.19}$$

Where: $ReLU(x)$ = ReLU function; x = input value (feature).

ANN constituted of multiple hidden layers, is a deep neural network (DNN) (Dangeti, 2017:268; Albawi et al., 2017; Géron, 2019:286; Zaras et al., 2022:17). DNN is also known as a deep multilayer perceptron (Abdolrasol et al., 2021). DNNs are suitable for supervised, unsupervised, reinforcement and hybrid learning types (Mathew et al., 2021:602). Each layer is connected to several other layers, where each layer can extract features as it channels to the next layers (Mathew et al., 2021:599). DNN algorithms aid in generating a model using complex datasets. With the help of non-linear activation, the model maintains a non-linear relationship between the input and the expected result. Each layer performs a volume of computation (Zaras et al., 2022:19).

Deep Learning implements different architectures to solve problems within different domains. Deep Belief Networks, Convolutional Neural Networks, Restricted Boltzmann Machine (RBM), Recurrent Neural Networks and Long Short-Term Memory (LSTM) are examples of deep learning architecture (Liu et al., 2017; Shrestha & Mahmood, 2019; Mathew et al., 2021:600; Alzubaidi et al., 2021).

2.1.11 Evaluating model performance

To assess how well the predicted model values align with the actual values, it is essential to evaluate the model's performance. If improvement is required, it can be achieved through hyperparameter tuning (Dangeti, 2017:286; Elgeldawi et al., 2021; Sarker, 2021; Janiesch et al., 2021).

In regression problems, several key metrics are commonly used to evaluate model performance by comparing predicted values with actual values. The metrics explored in this study include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2 or R-squared), and Adjusted R-squared (Chicco et al., 2021; Kumar et al., 2023; Olafadehan & Ahaotu, 2023). These error metrics depict the difference between the predicted and the observed values. The best algorithm is chosen based on the combination of minimal errors (MSE, RMSE, MAE) and the highest values of R-squared and Adjusted R-squared (Debroy & Seban, 2022b).

a. Mean Squared Error (MSE)

Mean squared error is an important evaluation metric for the algorithm's optimisation in regression models, as it minimises the squared differences between predicted and actual values. MSE is useful for detecting outliers. The best value for MAE is 0, and the worst value is +∞ (Chicco et al., 2021; Kumar et al., 2023).

The MSE formula is given below (Chicco et al., 2021).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2.20}$$

Where: n= total number of observations; y = actual value of the observation; $\hat{y}$ =predicted value of the observation.

b. Root Mean Squared Error (RMSE)

Root mean squared error is one of the most frequently used metrics to assess the accuracy of predictions. RMSE is the square root of the Mean Squared Error, MSE. It evaluates the standard deviation of the predictions from the actual value. The best value for RMSE is 0, and the worst value is +∞ (Chicco et al., 2021; Kumar et al., 2023).

The RMSE formula is given below (Chicco et al., 2021).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(yn_i - \hat{y}_i)^2}{n}} \tag{2.21}$$

Where: n= total number of observations; y = actual value of the observation; $\hat{y}$ =predicted value of the observation.

c. Mean Absolute Error (MAE)

Mean absolute error is the magnitude of the difference between the predicted value and the actual value. MAE can be used if outliers represent corrupted parts of the data. The best value for MAE is 0, and the worst value is +∞ (Chicco et al., 2021; Kumar et al., 2023).

The MAE formula is given below (Chicco et al., 2021).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2.22}$$

Where: n= total number of observations; y = actual value of the observation; $\hat{y}$ =predicted value of the observation.

d. Coefficient of Determination (R^2 or R-squared)

R-squared is the measure of the percentage of the variance explained by the model. The best value for R-squared is +1, and the worst value is -∞ (Dangeti, 2017:29; Chicco et al., 2021).

The R-squared formula is given below (Chicco et al., 2021).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2.23}$$

Where: n= total number of observations; y = actual value of the observation; $\hat{y}$ =predicted value of the observation.

e. Adjusted R-squared

Adjusted R-squared statistic explanation is very similar to R-squared, but it penalises the R-squared value if more variables without a strong correlation are included in the model (Dangeti, 2017:29). The range is less than or equal to R-squared (Sudheer et al., 2022).

The adjusted R-squared formula is given below (Dangeti, 2017:29).

$$R^2_{adjusted}=1-\frac{(1-R^2)(n-1)}{n-k-1} \tag{2.24}$$

Where: R^2 = sample R-squared; n= total number of observations; *k* = number of predictors (or) variables.

2.1.12 Explainable AI

Artificial Intelligence models often have operational behaviours that are difficult to understand and explain due to a lack of transparency; hence, these models are considered “black boxes” (Adadi & Berrada, 2018; Machlev et al., 2022). However, while the black-box nature of AI can produce powerful predictions (Adadi & Berrada, 2018). Enhancing the explainability of machine learning models has become essential. This need led to the development of Explainable Artificial Intelligence (XAI) (Machlev et al., 2022). XAI, also known as AI explaining or AI explainability, is a technique that explains the underlying processes of AI algorithms and can depict the reasoning behind the prediction (Rothman, 2020:3; Mohseni et al., 2021; Saranya & Subhashini, 2023; Lee et al., 2023). The XAI process is demonstrated in Figure 2.8. The explanations generated by XAI are presented to users through an interactive interface, allowing them to easily understand and interpret the AI model's insights (Rothman, 2020:3).

This improves the transparency, credibility, and accountability of AI systems, as transparency is essential for building trust (Saranya & Subhashini, 2023; Héder, 2023).

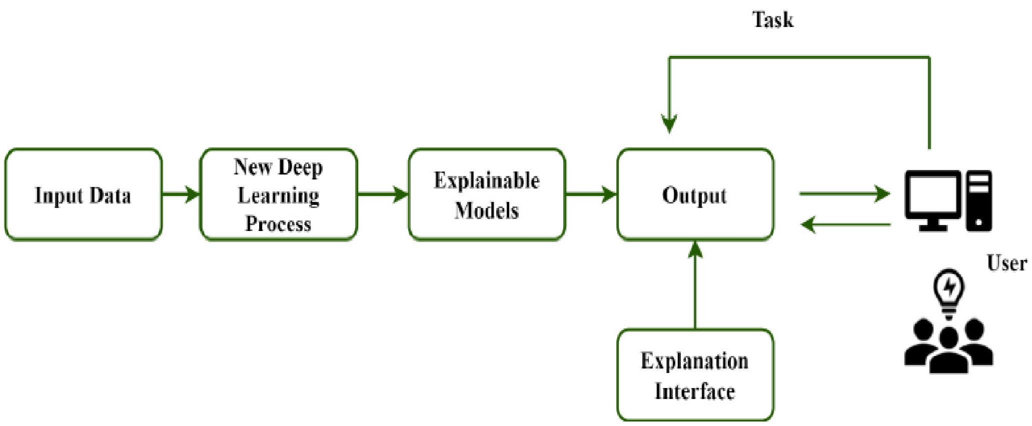


Figure 2.8: Explainable AI process (Saranya & Subhashini, 2023)

Adadi & Berrada (2018) addressed four key reasons for explaining AI systems: explaining to justify, explaining to control, explaining to improve, and explaining to discover. They concluded that explainability is a powerful tool for justifying AI-based decisions, aiding in prediction verification, model improvement, and uncovering new insights into the problem at hand.

The two main strategies to interpret (or explain) models are local and global interpretation (or explanation). Local interpretation (or Instance explanation) focuses on explaining a specific output of the system, while global interpretation (Model explanation) involves understanding the model as a whole (Das & Rad, 2020; Machlev et al., 2022; Héder, 2023; Mohseni et al., 2021). In the local explanation, map g (explanation of a model f) is generated each time for an individual data point $x \in X$ (a single instance of input data from a population X). However, in the global explanation is a group of data instances x and generating an explanation map g based on the given group of inputs (Das & Rad, 2020). Locally and globally explainable algorithms are described in Figure 2.9 and Figure 2.10, respectively.

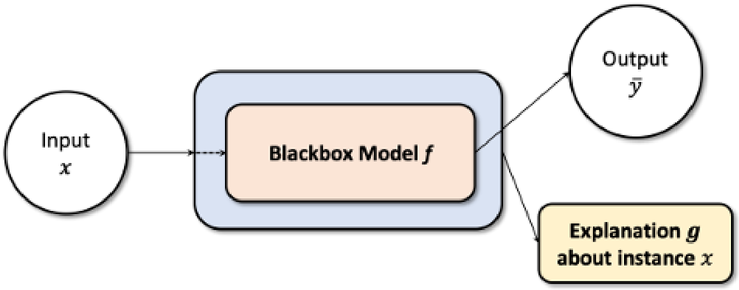


Figure 2.9: Explainable algorithm (Das & Rad, 2020)

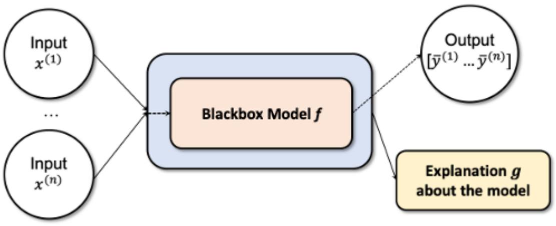


Figure 2.10: Globally explainable algorithm (Das & Rad, 2020)

XAI system purposes vary based on user demands (Mohseni et al., 2021). Their study identified six commonly used explanations in the design of Explainable AI (XAI) systems, as exhibited in Table 2.9 (Mohseni et al., 2021).

Table 2.9: XAI system design and purposes

Explanation types	**Purpose**
How Explanations	To explain how the model works. This provides a general overview of how the machine learning algorithm works.
Why Explanations	To explain the reason behind the prediction based on a given input.
Why-Not Explanations	To clarify the reason for the difference between the user's expected output and the model's prediction.
What-If Explanations	To demonstrate how the output changes with new inputs across different algorithms and data.
How-to Explanations	To explain what adjustments to the model or input data would be required to get the desired result.
What-Else Explanations	To provide examples of inputs that produce similar or identical outputs from the model.

The reasoning behind machine learning model explanations can be designed in various ways, depending on user preferences and objectives. Visual explanations, verbal explanations, and analytic explanations are among the different types available (Phillips et al., 2020; Mohseni et al., 2021).

Phillips et al. (2020) organise explanations into two main categories: self-interpretable models and post-hoc explanations. A self-interpretable model is an algorithm that can explain its overall structure globally and provide local explanations for individual decisions. Examples of self-interpretable models are decision trees and regressions, which include logistic regression models. A post-hoc explanation utilises software tools to provide insights into how an algorithm works. Post-hoc explanations are grouped into two: local and global explanations. Commonly used local explanation algorithms are Local Interpretable Model- Agnostic Explainer (LIME), SHapley Additive exPlanations (SHAP), Counterfactual, Saliency Pixel Algorithm, Class Activation Maps (CAM), Gradient-weighted Class Activation Mapping (Grad-CAM) and Individual Conditional Expectation (ICE). Global explanation algorithms are: Partial Dependence Plots (PDPs) and Testing with Concept Activation Vectors (TCAV). Two well-known XAI methods are LIME and SHAP (Kalasampath et al., 2025). LIME provides local and intuitive explanations. However, it is computationally expensive and may produce inconsistent interpretations when the model behaves in a complex manner (Linardatos et al., 2021; Kalasampath et al., 2025). Compared to LIME, SHAP provides both local and global explanations. SHAP is a game theory–based method that enhances the interpretability of individual predictions by computing the contribution or significance of each feature. It is more natural regarding interpretation (Saranya & Subhashini, 2023). SHAP is reliable and consistent, with mathematically grounded explanations, making it well-suited for decision-making processes (Kalasampath et al., 2025). This study focuses on SHAP to interpret machine learning based predictions because it is recognised as a unified measure of feature importance (Ekanayake et al., 2022).

2.1.12.1 SHapley Additive exPlanations (SHAP)

Interpreting a model’s predicted result is crucial in machine learning models, especially to understand which features contribute the most to making certain predictions. Explainable AI (XAI) technology allows users to understand, interpret and analyse the features that contribute to a model's training and its results (Linardatos et al., 2021; Lee et al., 2023). SHapley Additive exPlanations (SHAP) is one of the AI analysis techniques. Lundberg and Lee (2017) proposed SHAP values, recognised as a unified measure of feature importance that also enables the user to interpret the model’s behaviour for better decision-making (Ergün, 2023). Additionally, variations of SHAP, such as Kernel SHAP, Deep SHAP and TreeSHAP can be used for specific model categories (Ekanayake et al., 2022). A classic equation to compute the SHapley value is adopted from cooperative game theory. The SHAP framework helps to understand the contribution of each feature by assigning a value to each one for a particular prediction (Lundberg & Lee, 2017; Liu et al., 2024). The computed Shapley value is used as a feature attribute. SHAP values have various properties which are valuable for model interpretation. The properties are local accuracy, missingness and consistency (Lundberg & Lee, 2017; Ergün, 2023).

Local accuracy: When approximating the original model f for a specific input x, local accuracy requires the explanation model to at least match the output of f for the simplified input x' (Lundberg & Lee, 2017).

$$f(x) = g(x') = \emptyset_0 + \sum_{i=1}^{M} \emptyset_i x_i' \tag{2.25}$$

The explanation model $g(x')$ matches the original model $f(x)$ when $x = h_x(x')$.

Where $f(x)$ = Original model; $g(x')$ = explanation model; *M*= Number of simplified input features; $\emptyset_0$ = attribute an effect; $h_x(x')$ = mapping function; $\emptyset_i$ = *i*th SHapley value; x_i' = *i*th simplified input (Shapley, 1953; Lundberg & Lee, 2017).

Missingness: If the simplified inputs represent feature presence, then missingness requires features missing in the original input to have no impact (Lundberg & Lee, 2017).

$$x_i' = 0 \Rightarrow \emptyset_i = 0 \tag{2.26}$$

Where x_i' = *i*th simplified input ; $\emptyset_i$ = *i*th SHapley value.

Missingness constrains features where $x_i' = 0$ to have no attributed impact.

Consistency: Consistency states that if a model changes so that some simplified input's contribution increases or stays the same regardless of the other inputs, the input's attribution should not decrease (Lundberg & Lee, 2017).

Let $f_x(z') = f(h_x(z'))$ and $z' \setminus i$ denotes setting $z_i' = 0$. For any two models f and f', if

$$f_x(z') - f_x^{'}(z' \setminus i) \geq f_x(z') - f_x(z' \setminus i) \tag{2.27}$$

for all inputs $z' \in \{0,1\}^M$, then $\emptyset i(f', x) \geq \emptyset i\,(f, x)$.
Where : z' = a vector of features

For a model f and a set of features M, the Shapley value of the feature i is defined as: (Lundberg & Lee, 2017; Ergün, 2023):

$$\emptyset_i(f,x) = \sum_{z' \subseteq x'} \frac{|z'|!\,(M - |z'| - 1)!}{M!} \left[f_x(z') - f_x'(z' \setminus i)\right] \tag{2.28}$$

Where: $|z'|$ is the number of non-zero entries in z' and $z' \subseteq x'$ represents all z'vectors, where the non-zero entries are a subset of the non-zero entries in x'. M is the total number of features.

A fundamental approach to interpreting machine learning models is to analyse input features (attributes) to understand the training process and prediction outcomes. Feature importance highlights each feature's contribution to the model, indicating how valuable a specific feature is for accurate predictions. The SHAP value is used to calculate the influence of each feature on the predicted outcome, providing a clear measure of feature importance in the model's predictions (Lee et al., 2023).

The importance of each feature in the machine-learning model can be analysed through the SHAP. Feature importance is calculated by averaging the absolute SHAP values for all instances of the dataset (Lee et al., 2023).

$$S_0 = \frac{1}{N}\sum\nolimits_{i=0}^{N} |\emptyset_i| \tag{2.29}$$

Where: S_0 = mean absolute SHAP value; N= the number of instances in the dataset; $\emptyset_i$= the SHAP value of the feature for the i^{th} data instance.

Different methods, namely, Kernel SHAP, Deep SHAP, and Tree SHAP, are used to calculate the SHAP value for general machine learning models, deep learning models, and tree-based models, respectively (Lee et al., 2023).

Kernel SHAP: Kernel SHAP is a combination of linear LIME and Shapley values. LIME provides local interpretations of machine learning models, whereas Shapley values represent the importance of each feature for every individual observation in the prediction. Kernel SHAP can be applied to certain deep learning and machine learning models (Keleko et al., 2023). The purpose of this algorithm is to perform additive feature attribution by randomly sampling coalition vectors, masking features from the input data, and approximating the model’s influence through kernel SHAP linearisation (Das & Rad, 2020; Keleko et al., 2023).

Deep SHAP: The Deep SHAP is applicable for deep neural network explanation (Lundberg & Lee, 2017). Deep SHAP is considered a combination of Deep Learning Important FeaTures (DeepLIFT) and Shapley values. DeepLIFT method is for computing importance scores in a neural network by comparing a neuron's activation to its reference. Deep SHAP provides both local and global explanations of the features (Yang, 2021; Keleko et al., 2023). DeepLIFT leverages deep learning features to enhance computational performance and extract deep information (Keleko et al., 2023).

Tree SHAP: Tree SHAP is one of the approaches that reduces the time and memory costs of implementing SHAP. It is a specific implementation for the decision-tree-based ensemble models like random forest, gradient boosted trees. However, Tree SHAP can be imprecise locally because of the intrinsic uncertainty of the decision-tree models in Extreme Gradient Boosting (XGBoost) (Yang, 2021; Keleko et al., 2023).

2.1.13 Intelligent Internet of Things

The general concept of the "Internet of Things" is network connectivity that can send, receive, and analyse data. Whereas, "Intelligent Internet of Things" is the same as the IoT concept, along with the ability to take action based on analysed results (Prince & Prince, 2018:1). According to Zhang (2021), the blend of the IoT and Artificial Intelligence (AI) produces an Intelligent Internet of Things.

Artificial Intelligence (AI) is a key to tapping into IoT potential (Schatsky et al., 2017). AI technology, especially machine learning, can extract insights from the huge amount of collected data and help in pattern identification, prediction, and machine failure early warning. and so on (Schatsky et al. 2017; Firouzi, et al., 2020:15).

IoT is capable of exchanging data, whereas Artificial Intelligence technology provides worthwhile information. The Intelligent Internet of Things (Intelligent IoT) can make an impact in different disciplines. Business operations will improve in operational efficiency, decision-making, innovation and productivity. The implementation of AI, especially machine learning, in the Industrial Internet of Things can predict problems affecting industrial production by reducing maintenance and downtime costs, thereby increasing production output, etc. Data collection is much faster due to the IoT, which reduces the labour force and the data collection time needed. The Intelligent Internet of Things plays an important role in the improvement of consumer fulfilment of a need (Zhang, 2021).

2.1.14 Expert system

Intelligence refers to the ability to compute, reason, perceive, learn and solve novel problems, along with the ability to act like humans (Gupta & Nagpal, 2020:4, 11). An expert system (ES) is a knowledge-based intelligent information system (Liao, 2005; Dubey et al., 2013; Nagori & Trivedi, 2014; Rajabi et al., 2019; Zhang & Lu, 2021; Aslem & Abu-Naser, 2022; Megdad et al., 2022). It is a computer program that can provide information from the knowledge base to the user without the presence of human domain experts. This insight is useful during the decision-making process and for delivering recommendations to a user who may not be a domain expert (Bohanec et al., 1990; Tripathi, 2011; Yelapure & Kulkarni, 2012; Dubey et al., 2013; Aslem & Abu-Naser, 2022). Expert systems behave and judge like an experienced domain expert (Aslem & Abu-Naser, 2022; Tan et al., 2022).

The main components of expert systems are the user interface, rules/inference engine and knowledge base (Aslem & Abu-Naser, 2022; Megdad et al., 2022; Tan et al., 2022). A diagrammatic representation of an expert system is shown in Figure 2.11.

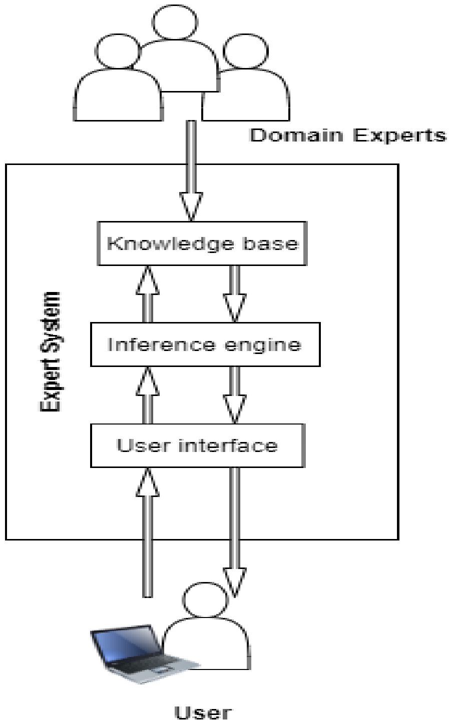


Figure 2.11: Schematic representation of an expert system (Janjanam et al., 2021)

Knowledge base: Expert systems' performance is dependent on the knowledge that is stored in the knowledge base (Janjanam et al., 2021; Aslem & Abu-Naser, 2022; Tan et al., 2022). Two forms of knowledge are stored in the knowledge base, namely: factual or declarative and heuristic or productive knowledge (Tripathi, 2011; Ogidan et al., 2019; Janjanam et al., 2021). Factual knowledge is acceptable facts about a particular domain that can be from experts, books, etc., whereas heuristic knowledge is generated based on individual judgement, good experience, practice or tacit knowledge and so on (Tripathi, 2011; Ogidan et al., 2019; Mohammed et al., 2019).

Inference engine: The inference engine is considered the brain of an expert system, where analysis and logical reasoning processes take place and find the inference for a specific domain problem based on the knowledge that is stored in the knowledge base (Tripathi, 2011; Ogidan et al., 2019).

User interface: To make the interaction between the user and the system (Tripathi, 2011; Joy & Sreekumar, 2014; Ogidan et al., 2018).

The knowledge in the knowledge base can be represented in different ways, helping to distinguish between the expert systems. Hence, the inference engine uses different approaches to speed up the inference process in concluding. Rule-based systems (RBS), fuzzy expert systems, frame-based expert systems, Knowledge-based systems (KBS), Artificial Neural Network Systems, hybrid expert systems, etc., are the categories of expert systems methodologies that are used when developing an expert system (Liao, 2005; Nagori & Trivedi, 2014; Ogidan et al., 2019; Mohammed et al., 2019; Janjanam et al., 2021).

An expert system uses knowledge and inference procedures to solve a domain-specific problem (Ogidan et al., 2019). The inference procedure helps to find the solution for a complex problem using knowledge and presents it to the user based on user input (Bohanec et al., 1990; Tripathi, 2011; Aslem & Abu-Naser, 2022).

2.1.15 Decision Support System (DSS)

A Decision Support System (DSS) is an interactive computer system that can support decision-makers by making an operational, planned or strategic decision to solve an unstructured and semi-structured problem using data and models (Ford, 1985; French & Turoff, 2007; Lu et al., 2007:53; Darbi & Saleh, 2022). A DSS can ease and advance the productivity, effectiveness, and efficiency of decision-making (Ford, 1985; Lu et al., 2007:56; Souha et al., 2024). Support may be in the form of providing a data summary, future prediction based on the current situation, assisting decision-makers to find insights and values, accounting for uncertainties, etc. (French & Turoff, 2007). The concept of a DSS was first formulated by Michael S. Scott Morton in the early 1970s (Sprague Jr, 1980; Ford, 1985; Power, 2008). Michael S. Scott Morton published his book "Management Decision Systems: Computer-Based Support of Decision Making" in 1971 (Power, 2007).

Keen and Scoot-Morton (1978, cited in Lu et al., 2007:54) mentioned the balance between Decision (D), Support (S) and System (S). Decision (D) focuses on the application selection criteria and the aspects of DSS, which are non-technical, functional and analytical concerns. Support(S) revolves around the implementation of the system, it tries to figure out how people operate the system, and how to provide help for the users. In summary, System (S) focuses on technology design and development. A successful DSS implementation can support individuals, groups or organisations (Phillips-Wren, 2017:5; Ghandar et al., 2021; Darbi & Saleh, 2022; Ali et al., 2023; Papazoglou et al., 2024; Senapaty et al., 2024).

According to Forgionne (2003:5), a typical DSS can be divided into three sections: input, process and output. The problem-related data and model are stored as input. The available data can be from internal and external sources. In the second segment, the decision-maker, with the aid of computer technology, processes the data by organising and attaching it to the model. Thereafter, the model is used to conduct an experiment or simulation. This helps in finding the best solution from the available alternatives. The process segment results, parameter requirements, experimental forecast, and recommended actions are reported. In the later stage, the obtained feedback from the decision-makers is stored as additional input for upcoming opportunities or other processes.

A single-user DSS provides early-stage support based on Simon's decision-making process. It includes input, process and output for each phase. In the intelligence phase, the system focuses on input, which involves problem definition, data collection, exploration, and preprocessing. In the design phase, the processing section helps in generating alternative solutions. Finally, in the choice phase, the output section assists in selecting the best solution or action within the problem context (Forgionne, 2003:5; Phillips-Wren, 2017:5; Milutinovíc et al., 2021; Hak et al., 2022). The DSS architecture is shown in Figure 2.12, which details how inputs, processing, and outputs interact to support decision-making.

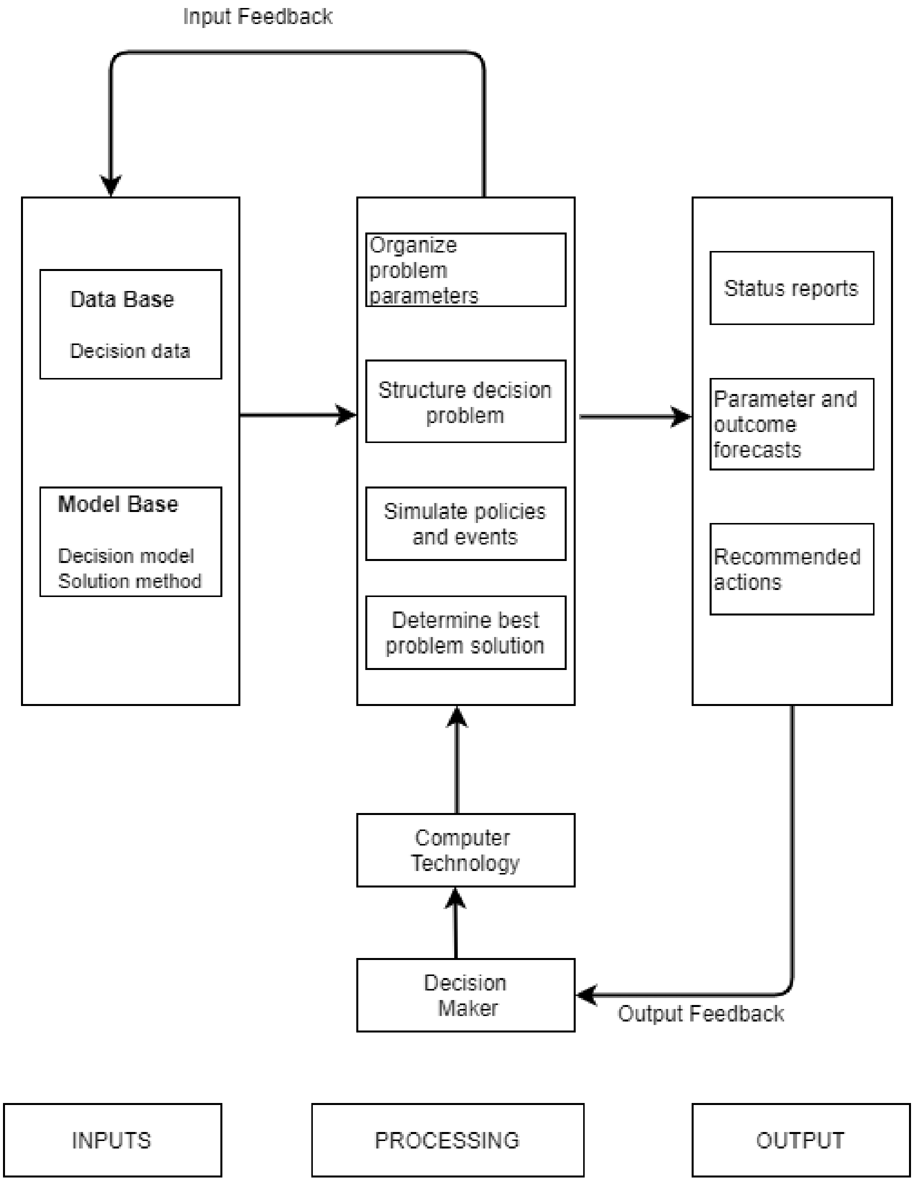


Figure 2.12: Decision support system architecture (Forgionne, 2003:6)

Problem-related data and models are stored as input in the DSS system. The decision maker uses a computer or capable device to process the inputs into problem-relevant outputs. Processing of the data involves: (a) organising the data into problem parameters, (b) structuring the parameters into a model, (c) using the model to experiment with policies and events, and (d) identifying the ideal solution to the problem. After completion of the processing, status reports, forecasts and recommendations are generated as output. The processing phase provides input feedback as additional data, knowledge, and models that could be used as a reference for future decision-making. Output feedback helps to extend, revise or modify the original analysis and evaluations. (Forgionne, 2003:6, 15; Phillips-Wren, 2017:5).

The following sections present a brief overview of the different types of DSS and the application of DSS. This provides the foundation for identifying the kind of DSS that should be developed to support stakeholders in the decision-making process.

2.1.15.1 Types of Decision Support Systems

A Decision Support System (DSS) can fall under any of the following categories based on purpose, users, and delivery namely, communications-driven DSS, data-driven DSS, document-driven DSS, knowledge-driven DSS and model-driven DSS (Power, 2008; Darbi & Saleh, 2022; Souha et al., 2024).

a. Communications-driven Decision Support Systems

Communication-driven DSS is heavily reliant on hybrid networks and electronic communication networks and their technologies. These technologies help connect and allow communication between collaborating resources and decision makers. Some of the communication-driven technologies are bulletin boards, audio and video conferences, as well as groupware (Power, 2008; Zeebaree & Aqel, 2019; Darbi & Saleh, 2022).

b. Data-driven Decision Support Systems

A data-driven decision support system (DSS) focuses on data retrieval and the manipulation of organisational internal or external data, and real-time data (Power, 2008). Once the user's requirements are established, this type of DSS will provide queries and management reports. This can be taken a step further with more advanced DSS providing online analytical processing and data mining. Thus, it can be used for analysing past data and establishing patterns and relations (Lu et al., 2007: 58; Darbi & Saleh, 2022). In addition, data-driven DSS often incorporate machine learning models to perform predictions (Gaftandzhieva et al., 2023).

c. Document-driven Decision Support Systems

A document-driven DSS combines a range of storage and processing technologies to provide extensive document retrieval and analysis. The World Wide Web and cloud computing technologies have been hailed as platforms for the use of decision support systems (Power, 2008; Darbi & Saleh, 2022).

d. Knowledge-Driven Decision Support Systems

A knowledge-driven DSS focuses on problem-solving by recommending actions to the decision-makers with the help of problem-solving expertise (Power, 2002:24). Knowledge-driven DSS embraces a rule-based system to assist decision-makers in making decisions (Lu et al., 2007:58). The expert understands the problem within the particular domain. An expert system technology connected to relational databases using web-based user interfaces has widened the use of knowledge-based DSS (Power, 2008; Darbi & Saleh, 2022).

e. Model-driven Decision Support Systems

A model-driven DSS focuses on access to and manipulation, simulation and optimisation of models. It relies more on mathematical models and optimisation than on huge amounts of data. Computer systems make model-driven DSS easy for managers to use. Artificial Intelligence applications help to overcome more complex problems (Chai & Jiang, 2011; Hızıroğlu et al., 2022). The models can be used for different purposes, such as accounting, financial, and optimisation. Statistical and analytical tools can be used to obtain basic-level functionality (Power, 2002:24; Zeebaree & Aqel, 2019; Darbi & Saleh, 2022).

2.1.15.2 Applications of Decision Support Systems

Decision Support Systems are used in various fields, namely, engineering, organisation, military, agriculture, health, tourism and so on (Senapaty et al., 2024; Souha et al., 2024).

Decision Support Systems assist users in enhancing their activities. For example, a knowledge-based DSS for predicting traffic crash events, a web-based DSS is used for human resource management for employee recruitment using Multi-Attribute Utility Theory, a web-based DSS for remote weather radar maintenance, predicting vegetable prices using a web-based DSS, recommending appropriate fertilisers to improve crop yield, and extracting information from disaster-related tweets for disaster management (Abou Elassad et al., 2020; Febriandirza et al., 2023; Papazoglou et al., 2024; Rao et al., 2024; Manju et al., 2024; Sinha et al., 2024).

Various technologies are used when developing the DSS, including Artificial Intelligence (AI), machine learning, Deep Learning (DL), Natural Language Processing (NLP), Docker, Flask for Application Programming Interface (API), a visual interface design tool, Balsamiq mock-up, and Google Translate API for real-time translation (Papazoglou et al., 2024; Febriandirza et al., 2023; Rao et al., 2024; Sinha et al., 2024).

A DSS integrates data, analytical models, and artificial intelligence to help both expert and non-expert users make well-informed decisions within their domain. DSS can integrate various models. The artificial intelligence-driven DSS serves for Data-Driven DSS and Knowledge DSS functionalities.

2.2 Related work

In this section, the existing body of scholarly work on prediction within the domain of smart aquaponics is reviewed. In this review process, related work was analysed critically to establish the basis which led to the proposed study.

Amano et al. (2022) focused on designing a Bok Choy Leaf Disease identification system in Smart Aquaponics. The dataset used for the research was a combination of the researcher's collected data and publicly available data from Kaggle. This study performed the identification and classification of diseases using machine-vision feature extraction. Data collection was performed using an IP Webcam. The study used IoT sensors to monitor water quality parameters, including pH, electrical conductivity (EC), and water temperature. The machine learning algorithms applied included Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbour (KNN). Based on the evaluation metrics, Precision, Recall, False Positive Rate (FPR), Specificity, and F1 Score KNN demonstrated superior performance compared to the other algorithms. The study successfully explored the use of machine learning algorithms for identifying Bok Choy leaf diseases in a smart aquaponics system.

However, the study focused only on the determination of the most effective ML algorithm and its comparison. The study could have had more benefit if it had built a decision support system that provided insights to the stakeholders on the detected disease and the prevention of the detected disease. This limitation prevented stakeholders from making informed and appropriate system management decisions.

Debroy & Seban (2022b) presented two prediction models for estimating tomato biomass within the aquaponics system, both the Artificial Neural Network (ANN) and its hybrid with fuzzy logic, known as Adaptive Neuro-Fuzzy Inference System (ANFIS). The ANN model utilised a Feed-forward backpropagation network, while the ANFIS model was also implemented. These models were developed to improve the accuracy and efficiency of tomato biomass estimation, offering potential advancements in agricultural management within aquaponic setups. The evaluation metrics: Mean Absolute Error(MAE), Root Mean Square Error (RMSE) and Coefficient of Determination (R^2) indicated that the ANFIS model had achieved the best prediction accuracy compared to the conventional ANN model. The data set included the input data information on recirculating water temperature (°C), dissolved oxygen (mg/L), nitrate (mg/L), and pH (ppm), while the output data represents tomato fruit biomass (g). Mathematical models were used for data collection. The study concluded that temperature, nitrate, and pH strongly correlated with tomato weight.

The study presented tomato biomass prediction to enhance economic management, improve production rates, and address market supply and demand challenges. However, it lacks effective communication of the findings to stakeholders for decision-making, particularly regarding the degree of influence of the various parameters that affect tomato weight, which could assist in better prioritising of those factors and management thereof.

Owusu et al. (2024) developed an aquaponics system to predict water temperature, where the system used heating elements operating concurrently at 5 watts, 10 watts, and 15 watts to observe water temperature changes over time. Based on Long Short-Term Memory (LSTM) RNN, the prediction model is particularly useful for reducing water temperature fluctuations, especially in outdoor aquaponics setups. The performance of the system was evaluated using the following metrics: Coefficient of determination (R2), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root-Mean-Square Error (RMSE). The study concluded that the LSTM model accurately predicted water temperature. Maintaining water temperatures between 20°C and 30°C was identified as essential for the optimal growth and health of bacteria, plants, and fish, ensuring a thriving aquaponics ecosystem.

The study focused its investigation only on water temperature as the predicted parameter and could have been broadened to predict other water parameters, as well as plant parameters, too. By integration of XAI, the study could enhance transparency and user trust, mainly for stakeholders who require not just predictions but also explanations for actionable insights.

Liu et al. (2024) proposed a fusion deep learning model (DLDL) for long-term prediction of dissolved oxygen (DO) concentration in aquaponics systems. Data was collected using IoT sensors, including water quality parameters such as DO, water temperature, pH, turbidity, conductivity, and salinity, with the aerator manually controlled to observe variations in DO levels. Before prediction, the raw time-series data were broken down using CNN-based decomposition to enhance the predictability of the data. Sequential features were then extracted using LSTM networks. A masked loss function was used to enable prediction at different temporal resolutions. The proposed model was evaluated against LSTM, Temporal Convolutional Network (TCN), CNN-LSTM, and Informer using Mean Squared Error (MSE) and Mean Absolute Error (MAE). The minimum errors are MSE = 0.199 and MAE = 0.355. The results indicate that the proposed DLCL model outperforms LSTM, TCN, and CNN-LSTM in long-term prediction accuracy. Finally, the predicted DO values were used to control the aerator automatically, ensuring sufficient oxygen concentration, promoting the health of cultured species, and reducing energy consumption through precise aeration management. The study demonstrated success in mitigating the challenges of traditional DO prediction models, such as poor stability and insufficient prediction accuracy, as well as the shortcomings of threshold-based aeration control.

In spite of the study's contributions, the study overlooked the effects of various factors that influence DO concentration. The model's absence of explainability restricts its integration with DSS, reducing transparency and interpretability for effective decision-making in system management.

Khandakar et al. (2024) focused on predicting the fish's length (cm) and weight (g) by analysing different parameters such as pH, ammonia, and nitrate levels, temperature (◦C), turbidity (NTU), dissolved oxygen (g/mml), pH, ammonia (g/mml), nitrate (g/ml), and the population of fish in the pond. The dataset for the study was sourced from freshwater aquaponics catfish ponds and was derived from the study titled "An Internet of Things Labelled Dataset for Aquaponics Fishpond Water Quality Monitoring System" by Udanor et al. (2022). The data, collected using IoT sensors, includes parameters such as temperature, pH, dissolved oxygen, turbidity, ammonia, and nitrate levels. The collected data is uploaded to the cloud in real-time and is publicly available on Kaggle under the *Sensor-Based Aquaponics Fish Pond Datasets* available at https://www.kaggle.com/datasets/e81da8b7666dc7af41cdc3aa5ef96c5547e4f412598a030f40d444550965e34f (Udanor et al., 2022). The study used several machine learning models for prediction, including Linear Regression, Lasso Regression, Ridge Regression, XGBoost, CatBoost, and LightGBM, which were evaluated using metrics such as R², Mean Squared Error (MSE), and Mean Absolute Error (MAE). Among these, the LightGBM model performed well in predicting fish length and weight. The incorporation of Explainable AI (XAI) Local Interpretable Model-Agnostic Explanations (LIME) for model interpretation represents a significant breakthrough, enhancing transparency and building confidence in machine learning predictions. This method enables stakeholders and domain experts to comprehend the model's results and leverage insights effectively, facilitating more informed and actionable decision-making.

Explainable AI (XAI) bridges the gap between high performance and interpretability. However, whilst insights are provided by the model, there has been oversight in communicating these insights to stakeholders effectively using a decision support tool. Incorporating plant growth prediction alongside fish growth could provide a more comprehensive view of the aquaponics system, leading to more efficient production and management.

Liu & Jiang (2024) implemented machine learning in their research to identify the most significant factors contributing to lettuce plant growth and their optimal levels. The study applied and evaluated several machine learning algorithms, including Linear Regression, Bagging Regressor, Decision Tree, Random Forest, XGBoost, and Artificial Neural Networks. The models were assessed using key performance metrics such as Accuracy, Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). XGBoost outperformed other models with 91.6% accuracy and the lowest MAE, followed by Random Forest with 90.9% accuracy and Bagging Regressor with 88.5%. A feature importance analysis of the best-performing XGBoost model revealed that Nitrogen had the most significant impact on plant growth, followed by nitrate, nitrite, light, and phosphorus. The dataset used for the study consisted of data collected from their previous research using IoT sensors. The parameters included air temperature, humidity, pH, light intensity, nitrogen, phosphorus, potassium levels, and Total Dissolved Solids (TDS). Additionally, camera module sensors were used to automatically capture and monitor plant growth, providing valuable insights into the system's effectiveness and optimisation (Jiang & Liu, 2024).

The study used feature importance to calculate ranking scores, identifying the parameters that most strongly influence plant growth. However, an XAI-based explainability approach can provide greater transparency on how and why a parameter influenced the prediction to the stakeholders for an informed decision-making process. This study's implementation had the potential to include a tool that enhances stakeholder support and decision-making.

Ghandar et al. (2021) designed a distributed, data-driven decision support system (DSS) for urban farming, designed to operate at two distinct scales. Firstly, on a large scale, the DSS supports urban agriculture planning by defining system structures, policies, and updates, while also enabling coordination among multiple stakeholders or users. It synchronises production with consumer demand in a data-driven way to minimise waste. Secondly, at the unit scale, a cyber-physical aquaponics prototype was developed to optimise production processes. A digital twin of the aquaponic system was implemented, providing a virtual model continuously updated with sensor data and real-time simulations. The proposed planning DSS was evaluated with the prototype on the collected data. The study compares the predictive performance of the digital twin with machine learning methods for predicting fish growth in aquaponics.

The study incorporated different machine learning algorithms to predict daily fish growth in grams and weekly plant growth rate in inches as a subsection of their study. They chose Regression (LR), Support Vector Regression (SVR), Decision Tree and ensemble method, and the eXtreme Gradient Boosting (XGBoost) decision tree. The parameters used were water temperature, room temperature, water pH, total dissolved salt (TDS), fish feed, fish weight, and plant length. The sensors used in the grow bed and fish tank to monitor the parameters were humidity, room temperature, pH level, and fish feeding. This data was collected over 3 months. The plant was the white tuberose bulbs, and the fish was the Nile tilapia. To evaluate the predictive model's performance, they used 10-fold cross-validation. After performance evaluation, the authors found that the best plant growth rate prediction model was simple linear regression, and the daily fish growth rate prediction was the decision tree or support vector regression. The model predicted the daily fish growth in inches and the weekly plant growth in inches. The study results showed that consumer satisfaction was highest when urban farms and retail locations were evenly distributed across the urban region in the planning of urban agriculture food production. Furthermore, the developed simulation model accurately predicted fish growth, particularly when it was frequently recalibrated with new sensor data.

The study demonstrated the necessity of a decision support system for planning in agriculture, with a focus on predicting fish growth using a simulation model. However, the system is limited in its ability to predict plant growth and lacks model explainability as well. Incorporating the most influential parameters for both fish and plant growth would enhance predictive accuracy and enable stakeholders to use the DSS more effectively for integrated urban agriculture planning and optimising aquaponics production.

This current study performed plant growth and water quality prediction using regression models that can predict numerical values. In similar plant growth prediction studies, the results showed that linear regression and XGBoost performed well (Ghandar et al., 2021; Liu & Jiang, 2024). Hence, this study used linear regression, XGBoost, along with random forest, K-Nearest Neighbors (KNN), and a multilayer perceptron (MLP) for prediction. Thereafter, evaluate the models using the commonly used metrics in various studies, namely Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2 or R-squared), and Adjusted R-squared (Debroy & Seban, 2022; Owusu et al., 2024; Khandakar et al., 2024). The machine learning model that had the best performance formed the basis for developing a decision support system for aquaponics prediction. The summary of related work is presented in Table 2.10.

Table 2.10: Summary of related work

Reference	Prediction	ML- models	Evaluation metrics	Best performance	Model Explainability	Decision support system
(Amano et al., 2022)	Bok Choy Leaf Disease	Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbor (KNN)	Precision, Recall, False Positive Rate (FPR), Specificity, and F1 Score	KNN	No	No
(Debroy & Seban, 2022)	Tomato biomass	Artificial Neural Network (ANN) and its hybrid with fuzzy logic, known as Adaptive Neuro-Fuzzy Inference System (ANFIS)	Mean Absolute Error(MAE), Root Mean Square Error (RMSE) and Coefficient of Determination (R^2)	ANFIS	No	No
(Owusu et al., 2024)	Water temperature	Long-Short-Term Memory (LSTM)	Coefficient of Determination (R2), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root-Mean-Square Error (RMSE)	-	No	No
(Liu et al., 2024)	Dissolved oxygen	LTSM, TCN, LTSF- Linear, CNN-LSTM, Informer and DLCL	Mean Square Error (MSE) and Mean Absolute Error (MAE).	DLCL	No	No
(Khandakar et al., 2024)	Fish's length (cm) and weight (g)	Linear Regression, Lasso Regression, Ridge Regression, XGBoost, CatBoost, and LightGBM	R^2, Mean Squared Error (MSE), and Mean Absolute Error (MAE)	LightGBM	Yes	No
(Liu & Jiang, 2024)	Plant growth	Linear Regression, Bagging Regressor, Decision Tree, Random Forest, XGBoost, and Artificial Neural Networks	Accuracy, Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE)	XGBoost	No	No

(Ghandar et al., 2021)	Fish growth & Plant growth	Regression (LR), Support Vector Regression (SVR), Decision trees and eXtreme Gradient Boosting (XGBoost) decision tree	MAE	Plant growth- Linear regression Fish growth – Support Vector Regression and Decision Tree	No	Yes
This study	Plant growth and water quality	linear regression, random forest and eXtreme Gradient Boosting (XGBoost), K-Nearest Neighbors (KNN), and a multilayer perceptron (MLP)	Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2 or R-squared), and Adjusted R-squared	Plant diameter prediction - Random Forest and XGBoost Plant height - Random Forest pH prediction – XGBoost TDS prediction - Linear regression	Yes	Yes

2.3. Research gap

In the examination of the body of the related work and literature, it is evident that most prediction studies in smart aquaponics have primarily focused on individual parameters, especially water quality, while giving limited attention to integrated or multi-parameter prediction. Although some studies explored plant growth prediction using image-based methods or measurement of plant height, other important growth indicators, such as plant diameter, remain underexplored. Furthermore, none of the previous studies have emphasised the need for model explainability and decision support for aquaponics stakeholders (see Table 2.10). Previous efforts have focused on developing predictive models to generate outputs, but have overlooked translating these insights into actionable knowledge for stakeholders in the decision-making process. To address these gaps, this study incorporates an explainable AI (XAI) model (SHAP) to improve model transparency, uses multiple evaluation metrics for robust assessment, and develops a decision support system that identifies key parameters influencing plant growth and water quality. By leveraging machine learning and empirical data from the South African context, this research enhances predictive capabilities, facilitates better decision-making, and contributes to the advancement of smart aquaponics management.

2.4 Chapter summary

The chapter covered essential components of this study, such as hydroponics, aquaculture, aquaponics, the Internet of Things, machine learning, XAI, Intelligent Internet of Things, expert systems and decision support systems, along with related work on prediction in the smart aquaponics domain. Finally, the review of related work reveals the research gaps in smart aquaponics that motivated this study.

CHAPTER THREE
RESEARCH METHODOLOGY

This chapter presents the methodology adopted for this study. The research onion framework, developed by Saunders et al. (2019:110), is used to describe the methodological choices made for this study. The research onion illustrates the various steps to create a research design. It contains different layers, starting from the outermost layer: philosophies, approaches, strategies, choices, time horizons, techniques and procedures (Saunders et al., 2019:110).

The researcher first selected the research philosophy's ontological and epistemological stance, which were then linked to the methodological approaches, including research design and data collection (Ugwu et al., 2021).

3.1 Research philosophy

Research philosophy embodies a critical assumption from the perspective of how the researcher sees the world (Ugwu et al., 2021). This influenced and affected the way the research was conducted. This relative view of the world meant the knowledge gained was influenced by the perception of the researcher (Khatri, 2020). Research philosophy is a belief or a set of beliefs about the ways data needs to be collected, analysed and used. The researcher had to be conscious and aware while forming beliefs and assumptions. According to Saunders et al. (2019:113), there are three major aspects: ontology, epistemology and axiology. Each of these aspects carries substantial differences, influencing the critical thinking and research procedure adopted. The research philosophy that was selected for the study is positivism.

3.1.1 Ontological stance

Ontology is a branch of metaphysics that stems from philosophy (Smith, 2012:47). Ontology is derived from two Greek words, “onto” and “logos”. “Onto” means “being” or “that which is everything that exists”, whereas “logos” means “knowledge” or “study” (Ni’mah et al., 2024). Put together, it can be referred to as the “Study of being” (Crotty, 2003:10). Ontology deals with “the nature of reality” or “nature of existence” (Saunders et al., 2019:133; Ugwu et al., 2021). It factors into the research assumptions the nature of the world and reality. Ontology is a core concept that guides data collection, analysis, and interpretation (Ugwu et al., 2021).

Objectivism was selected as the ontological stance of this study. Objectivism “ believes that there is only one true social reality experienced by all social actors” (Saunders et al., 2019:135). Objectivism was selected because the study intends to seek the nature of existence based on building the reality of whether a decision support system can be developed for aquaponics prediction, which would be independent of human thought or consciousness.

3.1.2 Epistemology of the study

Epistemology is one of the three aspects of philosophy. “Epistemology” is derived from two Greek words, “episteme” and “logos”. “Episteme” means “knowledge”, “understanding”, or “acquaintance”, whereas “logos” means “account”, “argument”, or “reason” (Ni’mah et al., 2024). Epistemology is “the theory of knowledge” (Ni’mah et al., 2024). The primary source of knowledge for this research was the literature review, while new knowledge was gained from analysing the data, the findings, and conclusions from the experiments. The study focused on a single reality that can be measured; hence, the epistemological stance of this study was positivism. The study made predictions based on quantifiable data.

The knowledge gained thus far within the aquaponics environment has been primarily through experimentation as well as surveys from previous studies. The experimental result is presented in different ways. It is believed that the proposed decision support system will assist aquaponics farmers in making decisions. To find the truth, experiments were conducted to predict the aquaponics output using different parameters, namely: pH, TDS, EC, water temperature and plant details such as plant height, number of leaves, plant diameter, ambient temperature, and humidity. Thereafter, the insights gained from the experiment were presented to the aquaponics stakeholders, including farmers, researchers and aquaponics practitioners in a meaningful and understandable way. The stakeholders provided feedback on the study to improve the prediction model. This feedback also determined whether the proposed decision support system was indeed helpful in making effective decisions for aquaponics farming, and thus, the study justified the belief. The source of knowledge was empirical knowledge.

3.2 Research approach

The research approach is a plan and roadmap for conducting research. The research approach guided how to collect, analyse, and interpret the collected data (Cresswell, 2014:3). The selected research approach for this study was deductive, as the literature helped to identify relevant theories and ideas that were subsequently tested with collected data (Saunders et al., 2019:78). The central research question explored in this study was whether a decision support system for aquaponics prediction could be developed to aid stakeholders in making better decisions and taking corrective actions. An experiment was carried out, and the results were analysed. The test outcomes provided insights into the research question.

3.3 Methodological choice

This study adopted a quantitative methodology because it required experiments with empirical, quantifiable data to develop a decision support system that can predict key aquaponics parameters. The parameter values for the experiment were collected from the field using measuring tools and IoT devices. The collected data was recorded numerically and analysed statistically (Creswell, 2014:4). Thereafter, the research sought to investigate the knock-on, ripple effects of the variable changes on one another. It was important to establish the interdependence of variables as the entire aquaponics system had to be in equilibrium; otherwise, it could affect the yield output adversely (Yildiz et al., 2017).

3.4 Research strategy

Saunders et al. (2019:173) defined research strategy as a "general plan of how the research questions of the study will be answered". The proposed research aimed to determine various parameter values that could support decision-making. A fundamental relationship existed between these parameters and the prediction process. To accomplish this objective, the research adopted an experimental research strategy and used machine learning techniques for experimentation (Saunders et al., 2019:178, 190). The collected data was pre-processed through data cleaning and feature engineering. For the experiment, supervised machine learning algorithms such as Linear regression, random forest, K-nearest neighbour, eXtreme Gradient Boosting, and multi-layer perceptron were adopted. Data were trained using built-in algorithms, and then an optimal model was generated (Takami et al., 2016). The experimentation process used the Jupyter Notebook on a web-based computing platform, incorporating various built-in libraries such as pandas, numpy, sklearn, matplotlib, etc (Fenner, 2019:20; Géron, 2019: 48).

3.5 Research design

The research design serves as a comprehensive blueprint, outlining how to approach the research questions (Saunders et al., 2019:173). Research design provides the structure to choose the correct research methods and techniques proposed for collecting and analysing the data (Saunders et al., 2019:173). A good research design helps to find accurate answers to the problem using collected data during the research. The purpose of the research was to develop a decision support system for aquaponics prediction to aid farmers in making decisions to achieve maximum productivity.

The study investigated the research objectives, identified the causes and effects, and observed how changes in one variable could affect one or more variables. The research questions were used to evaluate the accuracy of the developed prediction models. Thus, the researcher employed an experimental research design. The experimental research design of this study influenced the data collection, experimentation, presentation of results, evaluation, and interpretation of the findings of this study.

The insights derived from the study were used as a guide for decision-making and actions, resulting in gained insights and wisdom. The overview of the adopted research design is shown in Figure 3.1.

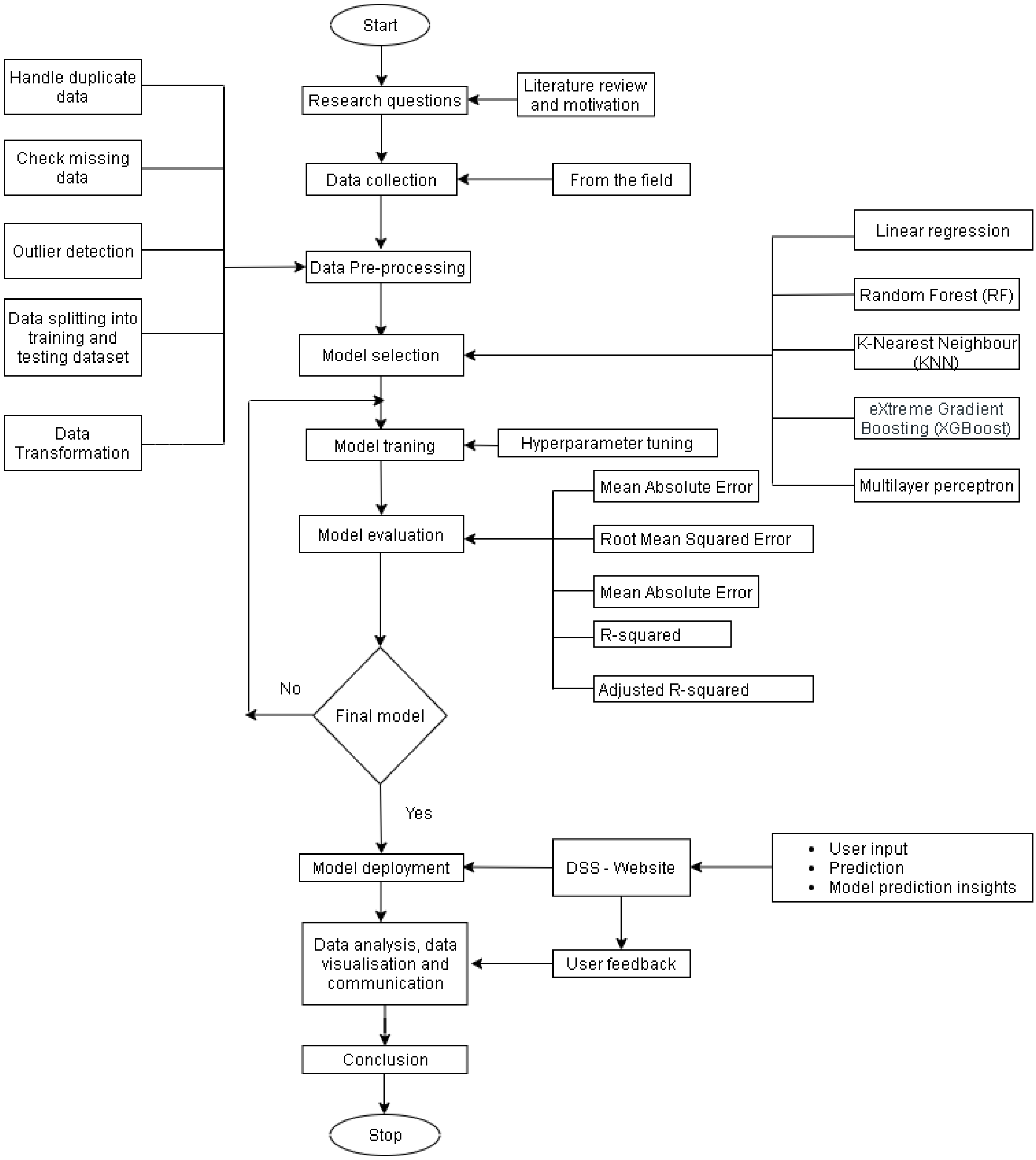


Figure 3.1: Overview of the experimental research design (Source: Researcher)

The literature review assisted with the basis for identifying the research problem. This helped in the formulation of the research aim. From the aim, specific objectives were derived that guided the study. The research questions were then developed to align with these objectives. The following sections explain the research design in detail.

3.5.1 Data collection

Primary data for this study are critical and were collected for plant growth and water quality parameters (Saunders et al., 2019:338). For the field experiment, a coupled grow-bed-based aquaponics unit was set up within a tunnel. A DHT22 IoT sensor was also installed alongside the unit (Ghandar et al., 2021; Sunardi et al., 2021).

During this study, the plant growth dataset comprised the plant height, number of leaves, and plant diameter to form the key indicators in assessing the plant growth (Frasetya et al., 2021; Mokhtar et al., 2022; Villanueva et al., 2022). The ambient humidity and temperature were also included as key parameters influencing plant growth (Dutta et al., 2018; Abdullah & Mazalan, 2022). Plant details were recorded once every week, and on the first day of the week (Villanueva et al., 2022). Plant details, height and diameter were measured using a ruler, and the number of leaves was physically counted and recorded in an Excel sheet (Mahkeswaran & Ng, 2020; Subakti et al., 2022; Udanor et al., 2022; Villanueva et al., 2022). The ambient humidity and ambient temperature data were stored in an SD card via Raspberry Pi 4 (Pappu et al., 2017; Varkey et al., 2021; Alselek et al., 2022). The humidity and temperature measurements were recorded within two-minute intervals. Various studies have used different time intervals to record the collected data, for example, one-minute or five-minute intervals. (Defa et al., 2019; Kjellby et al., 2019).

Water temperature and pH play a crucial role in determining water quality. The pH level reflects the acidity of the water. Temperature changes can influence various components of the aquatic environment, including the pH as well (Maulini et al., 2022). Temperature and the pH of the water also play a role in the nitrification process of breaking down the fish waste and other microorganisms (Channa et al., 2024). Fish growth parameters, such as height and weight, have also been reported to be significantly influenced by pH and water temperature (Khandakar et al., 2024). Dewangan et al. (2023), highlighted the importance of monitoring water temperature in water quality management due to the significant relationship between temperature and EC/TDS. Since water quality and fish growth performance are directly influenced by pH, temperature, TDS, and EC, these parameters are the selected water quality parameters for this study. A water quality measuring device was used for collecting all four water quality parameters (Yanes et al., 2020; Subakti et al., 2022; Liu et al., 2024). Water quality parameters were collected at four intervals on a daily basis and were recorded into an Excel sheet.

After the field experiment, a dataset of 709 plant detail records and another dataset of 526 records for the water quality were compiled. Data collection began on 2023-07-28 and concluded on 2024-07-08, spanning approximately 4 months and 13 days (133 days), with 4 cycles conducted during this period. The data collection cycle is shown in Table 3.1. In the next step, the collected data had to be cleaned for further analysis.

Table 3.1: Data collection cycle

Cycle	Starting date	Ending date
1	2023-07-28	2023-08-28
2	2023-10-24	2023-11-13
3	2024-02-27	2024-03-26
4	2024-05-13	2024-07-08

3.5.2 Data pre-processing

In this study, data preprocessing was performed using the Python programming language, which is a prime language for data science and machine learning applications (Oscar et al., 2023). The preprocessing code was executed within the code cells of Jupyter Notebook (Géron, 2019: 48). Duplicate records were identified and discarded (Dabool et al., 2024). Since manual recording was adopted, this has prevented the detection of missing data in the plant and water quality dataset. Furthermore, outliers were detected using boxplots and afterwards handled by replacing them with the mean value, which is a central measure of the data distribution (Molin, 2021:13; Wilson et al., 2021). The independent and dependent variables were identified from the collected datasets to predict plant diameter, plant height, water pH, and water TDS. Thereafter, the dataset was split into training and test sets using an 80:20 ratio, respectively (Wilson et al., 2021; Kumar et al., 2023; Daniel et al., 2025). Through the feature engineering process, the features were scaled to a uniform range to improve model performance, since the dataset contained features with fluctuating scales (Keerthana et al., 2021; Abdelaziz et al., 2025). Hence, MinMaxScaler normalisation was used in this study to ensure that all feature values were scaled to fall within the range of 0 to 1(Molin, 2021: 562; Seegobin et al., 2024). After data pre-processing, the datasets were ready to be input into various supervised algorithms to perform the prediction.

3.5.3 Model Selection

In this study, plant diameter and plant height are considered as dependent (output) variables for plant growth prediction, whereas pH and TDS are considered dependent variables for water quality prediction. Thus, the independent (input) variables for plant growth predictions are plant height, plant diameter, number of leaves, ambient temperature and ambient humidity.

Whereas, in water quality prediction, pH, temperature, EC and TDS are independent (input) variables. The following supervised machine learning models were selected for the prediction.

Linear regression: It was selected because it is computationally simple and easy to implement (Kadam et al., 2025). A linear regression model aims to identify a general important pattern that connects independent variables and dependent variables. Furthermore, linear regression strives to establish a relationship between these variables, extending towards predicting dependent variables for specified input values (Kim et al., 2022). It is typically used as a baseline model to obtain preliminary insights into the data (Kadam et al., 2025).

Random Forest (RF): It is an ensemble of decision trees that combines the outputs of individual trees to produce the final prediction (Wie, 2023). Due to the manual, physical collection of data, as well as exposure to varied climatic conditions, noise, or errors are present in the collected data. RF was chosen as it reduces overfitting and handles missing or noisy data (Molin, 2021:653; Kadam et al., 2025). RF is used in many prediction studies due to its high accuracy and robustness (Wie, 2023).

K-nearest neighbour (KNN): In this study, the KNN model was selected for the prediction process because it is a simple, non-parametric and instance-based machine learning algorithm that does not require a specific training phase (Ozaga et al., 2024). The dataset used in this study is relatively small. KNN is tailored for this type of collected data because it does not make prior assumptions about the input variables. It also provides good accuracy on a small data size (Seyghaly et al., 2024).

eXtreme Gradient Boosting (XGBoost): XGBoost is a decision tree-based ensemble ML algorithm (Desdhanty & Rustam, 2021). It is selected in this study because of its ability to handle missing data, high accuracy, control overfitting, computational efficiency and high scalability (Mahajan et al., 2023; Wen et al., 2024). It uses various techniques to improve model performance and efficiency. Techniques, namely, parallelisation, optimising objective functions and regularisation. Additionally, it is apt for finding the key features in the given dataset, which is useful for feature selection and understanding the relationships within the data (Khan et al., 2024).

Multilayer Perceptron (MLP): It is a deep artificial neural network that consists of several interconnected perceptrons. MLP is used in the study to predict plant growth and water quality. During the training process, a series of input-output pairings assists with learning to represent the dependencies between the input features and the expected output. MLP also has a good nonlinear fitting ability whilst being suitable for complex datasets, and it minimises training errors (Zaras et al., 2022:19; Taud & Mas, 2018: 454; Obiora et al., 2023).

The criteria for selecting the specific machine learning models for prediction are summarised in Table 3.2.

Table 3.2: Overview of the attributes of the selected ML Algorithms

Algorithm	SI	Acc	HO	HMA	NLR	CLD	References
Linear Regression	High	Medium	Prone to overfitting if not regularised	Poor	No	No	(Lantz, 2013:161,169; Rashidi et al., 2019; Joshi, 2020: 36; Kim et al., 2022; Kadam et al., 2025)
Random Forest (RF)	Medium	High	Robust to overfitting	Good	Yes	Yes	(Kadam et al., 2025; Wei, 2023)
KNN	High	Medium to High	Prone to overfitting if K is not chosen correctly	Poor	Yes	No	(Sudheer et al., 2022; Zhang et al., 2023; Ozaga et al., 2024; Seyghaly et al., 2024)
XGBoost	Medium to Low	High	Robust to overfitting with regularisation	Good	Yes	Yes	(Friedman, 2001; Chen & Guestrin, 2016 ; Mahajan et al., 2023; Khan et al., 2024; Daramola et al., 2025)
Multilayer Perceptron (MLP)	Low	High	Prone to overfitting if not regularised	Poor	Yes	Yes	(Taud & Mas, 2018: 454; Obiora et al., 2023)
SI: Simplicity of implementation Acc: Accuracy HO: Handling overfitting HMD: Handling of missing data NLR: Non-linear relationships CLD: Efficiency with complex and large datasets							

3.5.4 Model training

The preprocessed data was used to train the selected machine learning algorithms to identify trends in the dataset and finally make predictions (Panigrahi et al., 2023). After training, the algorithm represents the data in the form of a model (Lantz, 2015:16). The model was applied to the test dataset to evaluate how well it could generate accurate predictions. Thereafter, model performance was assessed using appropriate evaluation metrics

3.5.5 Model evaluation

The performance of the selected supervised machine learning models was evaluated using standard regression metrics, namely Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R-squared (R^2), and Adjusted R-squared (Adjusted R^2). These are the most frequently used regression metrics in the literature (Sudheer et al., 2022; Debroy & Seban, 2022; Owusu et al., 2024; Khandakar et al., 2024). Finally, the metrics results were compared by checking the minimal error in MSE, MAE and RMSE and the highest score in R^2 and Adjusted R^2 (Sudheer et al., 2022; Khandakar et al., 2024).

3.5.6 Hyperparameter tuning

Hyperparameter techniques help to enhance the model's performance. Hyperparameter tuning techniques, *GridSearch* and *RandomSearch*, were used in this study for a comparative analysis (Elgeldawi et al., 2021). To split the data into several combinations, both 5-fold and 10-fold cross-validation methods were used (Lauguico et al., 2020; Khandakar et al., 2024). Then, the best hyperparameters were selected to train the selected supervised algorithms to predict the output. Two options of 5-fold and 10-fold were used to create a basis to experimentally determine the cross-validation option that would produce the best performance.

3.5.7 Model Deployment

A data-driven Decision Support System (DSS) using machine learning was developed in order to deploy the best-performing models. It was developed as a Python web application using the Flask framework, deployed on PythonAnywhere (Mufid et al., 2019). The developed DSS was made to be accessible to the various participants from anywhere (Gao et al., 2021). Participants could provide input data via the website. The system provided the participant with the predicted output, together with insights of the most significant factors contributing to the predicted value.

3.5.8 User feedback and communication

The usability of the developed DSS was evaluated using the System Usability Scale (SUS), which is a standardised self-completed questionnaire (Setemen et al., 2019; Saunders et al., 2019:505, 506). The SUS questionnaire was created using Google Forms and integrated into the DSS website. Participants were invited to explore and evaluate the developed DSS (Setemen et al., 2019; Saunders et al., 2019: 505). The targeted population consists of aquaponics practitioners, researchers and aquaponics community members (Saunders et al., 2019:295). Invitations to participate in the survey were distributed through various communication channels, including email, WhatsApp, and Facebook. Once the respondent completes the questionnaire, their responses are saved automatically (Brooke, 1996; Saunders et al., 2019: 544). A convenient sampling technique was used, where each participant volunteered to participate in the evaluation (Saunders et al., 2019: 324). Due to the voluntary nature of participation and online questionnaires, the survey experienced a relatively low response rate; hence in the study used a less representative sample (Chelghoum, 2024). Out of 127 individuals invited, only 16 responded, and 14 of those completed the SUS questionnaire in full. The collected data were analysed quantitatively, and insights were communicated using graphs and tables.

3.6 Ethical considerations

The ethical approval for this study was obtained from the Faculty of Informatics and Design Research Ethics Committee of the Cape Peninsula University of Technology (CPUT). Since the study involved human participants, all participants of this study were informed of their rights before participation(Saunders et al., 2019:55).

3.6.1 Protection of people

The aquaponics unit is situated at the University of Johannesburg (UJ) in Johannesburg, South Africa. Officials at the aquaponics site were not subjected to any harmful chemicals or products that could adversely affect their health.

3.6.2 Protection of the environment

Aquaponics units are environmentally safe and made from food-grade plastic so as not to release toxins into the environment. The aquaponics unit also encompasses three living organisms: plants, bacteria and fish. This aquaponics system maintained a good ethical relationship and provided value to the natural environment. The study did not use any pesticides in the aquaponics unit that would harm the living organisms. The research ensured that the data collection did not cause any harm to the aquaponics unit's living organisms and the environment.

3.6.3 Data storage

The data were stored electronically on common digital storage devices and were kept at CPUT. The data did not contain sensitive information. It primarily included the collected data and feedback from the aquaponic farmers. Private information of aquaponics farmers was not stored.

3.6.4. Informed consent

The participants in the evaluation survey gave their informed consent, and participation was voluntary.

3.7 Chapter summary

In this chapter, a summary of how the researcher designed the study is provided, along with the justification for the choice. The chosen research aspects applied in the study were ontology and epistemology. The ontological stance of the study was objectivism, whereas positivism was adopted for epistemology. The research approach was deductive and used the quantitative methodological choice to conduct an experiment, which was selected as the research strategy. Towards the end of this chapter, the ethical considerations applied in this study were explained.

CHAPTER FOUR
DATA COLLECTION

A vital step in the research process is gathering data, which forms the basis for insightful analyses and conclusions. This chapter explains how data was collected from the field and recorded to conduct the experiment.

4.1 Data collection

Experimentation is one of the data collection methods (Taherdoost, 2021; Ganesha & Aithal, 2022). For this study, the primary data were collected from the field experiment. Aquaponics was set up in a real-world setting; thus, the field experiment provided a high external and ecological validity. Ecological validity indicates how much the outcome of the study can be generalised to real-life conditions (Taherdoost, 2021; Ganesha & Aithal, 2022). Continuous quantitative data were collected and recorded from the field (Ganesha & Aithal, 2022).

Data collection commenced on the 28th of July 2023 and concluded on the 8th of July 2024. This amounts to a period of 11 months (346 days). Data was collected in 4 cycles during this period. Table 4.1 presents the data collection cycle. The following sections explain how data were collected for this study.

Table 4.1: Data collection cycle

Cycle	Starting date	Ending date
1	2023-07-28	2023-08-28
2	2023-10-24	2023-11-13
3	2024-02-27	2024-03-26
4	2024-05-13	2024-07-08

4.1.1 Aquaponics setup

In this study, a media grow-bed-type coupled aquaponics system was set up within a tunnel in Johannesburg, South Africa, for data collection. Three 102 cm x 108 cm media grow bed units with one 1000 litre tank holding 27 Mozambique Tilapia fish were used. In each grow bed, nine leafy lettuces were planted within a 12 cm distance of one another, and numbers were assigned to the plants. The plant numbering is shown in Figure 4.1.

Grow bed - 1	Grow bed - 2	Grow bed - 3

Figure 4.1: Plant numbering in grow beds

The aquaponics setup and fish used for this study are displayed in Figure 4.2 and Figure 4.3.

Figure 4.2: Aquaponics setup

Figure 4.3: Mozambique Tilapia Fish

4.1.2. Data recording

Manual measuring instruments were used to determine the plant details and water quality details (Saunders et al., 2019:403). Humidity and ambient temperature measurements were done using an IoT device.

4.1.2.1 Plant details

The plant diameter and plant height were measured using a measuring scale and recorded in centimetres, and the leaves were counted manually. A sample of measuring the plant height and diameter is shown in Figures 4.4 and 4.5. The plant growth was visually inspected, taking into account of the plant height, number of leaves, and leaf area (diameter) (Pandey et al., 2017; Frasetya et al., 2021; Qadeer et al., 2020). Plant details were measured once a week. Plant height was measured from the grow bed level to the longest leaf of the plant (Valiente et al., 2018; Villanueva et al., 2022). Plant diameter/area was measured from one leaf end to another leaf end.

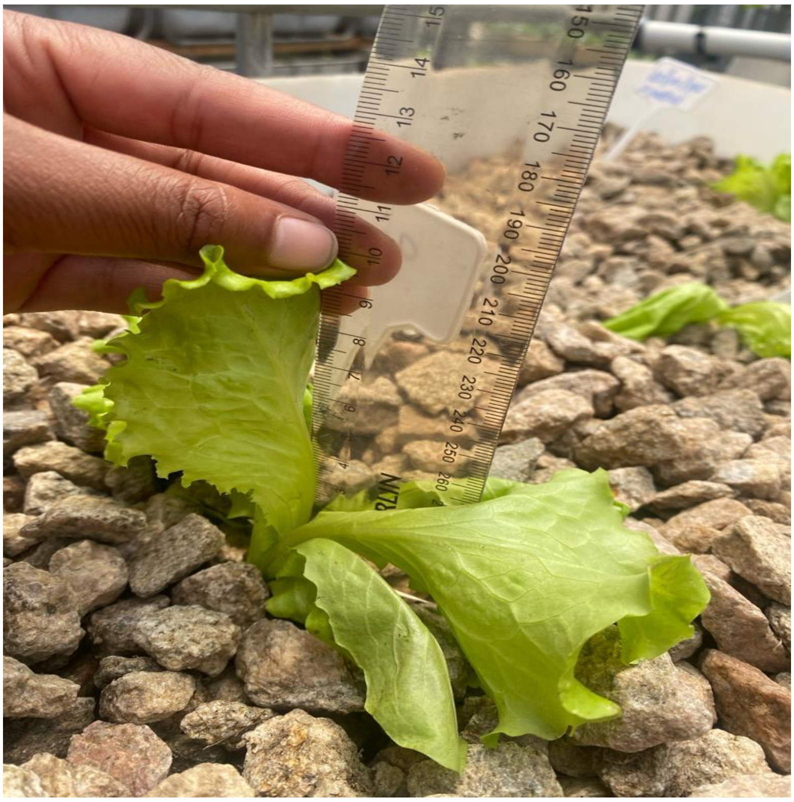


Figure 4.4: Plant height measuring

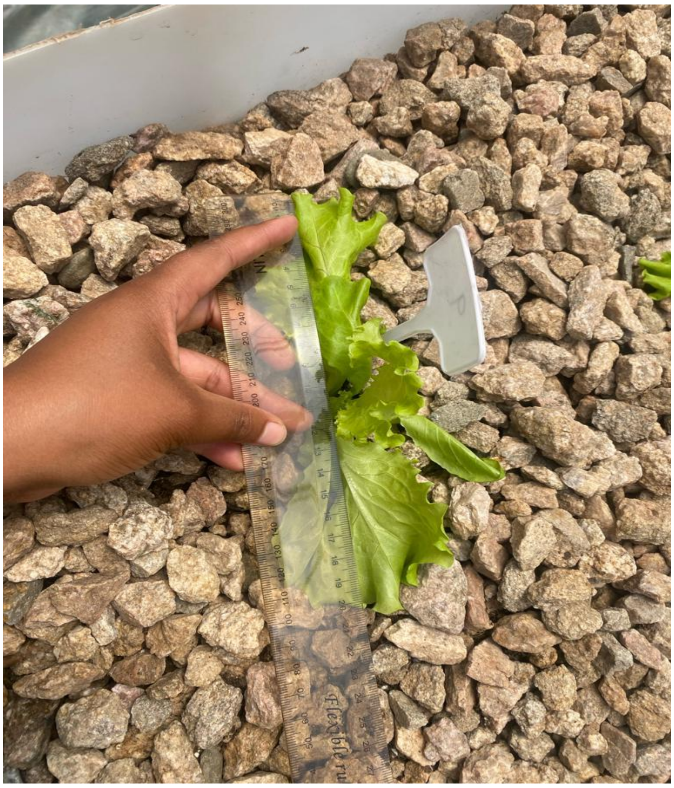


Figure 4.5: Plant diameter measuring

4.1.2.2 Water parameter

The selected water quality parameters for this study were pH, water temperature, TDS and EC. The Nf-7 in 1 water quality tester pen was used to measure the water quality. Water quality measurement was conducted using a test pen, is depicted in Figure 4.6. The units that were used in this study to measure the water quality are microsiemens per centimetre (µS/cm) for EC, parts per million (ppm) for TDS and Celsius (°C) for water temperature (Eneh et al., 2023; Abidin et al., 2024; Dewangan & Shrivastava, 2024). The water quality was measured daily, four times a day, between 9:00 AM to 12:00 PM. The results were recorded in an Excel sheet. More accuracy was ensured by calibrating the water quality tester each time before use. This procedure was followed as per the guidelines in the instrument manual located in the packaging (Mandap et al., 2018; Wibowo et al., 2019).

Figure 4.6: Water quality measurement

4.1.2.3 Ambient temperature and humidity

The sensor DHT22, Node-RED, a visual programming tool and Raspberry Pi were used in this study for ambient humidity and temperature data collection and visualisation (Lekić & Gardašević, 2018; Ekanayake et al., 2022; Arigela et al., 2024).

DHT22 Sensor: It is a low-cost, reliable, and effective sensor for measuring ambient temperature and humidity. It is important to collect the ambient temperature where plant growth is directly dependent (Ghandar et al., 2021).

Node-RED: This is a flow-based JavaScript development tool built on the Node.js platform and is used for visual programming (Lekić & Gardašević, 2018; Arigela et al., 2024). It is used for connecting hardware devices, API, and online services whilst providing a user-management interface (Lekić & Gardašević, 2018; Arigela et al., 2024). This development tool allows users to add or remove nodes and connect them for communication without writing code (Lekić & Gardašević, 2018; Garbev, 2022).

Raspberry Pi: It is an inexpensive, high-speed open-source computer device that consumes minimal power whilst being portable, making it ideal for IoT applications (Pappu et al., 2017; Dutta et al., 2018; Hosny et al., 2023). In this study, the Raspberry Pi 4 model B with 4GB of RAM and a 16 GB SD card was used (Hosny et al., 2023).

DHT22 was set up to collect the ambient temperature and humidity, which is shown in Figure 4.7. The collected data was recorded into an SD card in the Raspberry Pi unit. The data was recorded every two minutes and saved in a Comma-separated values file format.

Figure 4.7: DHT22 used to collect ambient temperature and humidity data

Installed Node-RED on the Raspberry Pi for visualising real-time data. When the user connected over the WIFI, ‘aqua_2’, they were able to access the Node-RED dashboard locally. It allowed the user to monitor the ambient humidity and ambient temperature data (Garbev, 2022). Figure 4.8 depicts the block diagram of the Node-RED, DHT22 and Raspberry Pi integration that was used in this study.

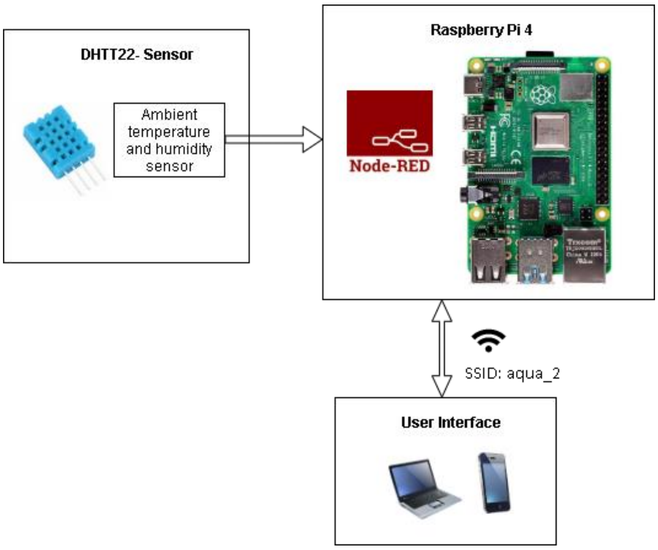


Figure 4.8: Architecture of IoT-based data collection and visualisation

A sample dashboard, which visualises ambient temperature and humidity at 10-minute intervals for 2023-11-16, is shown in Figure 4.9.

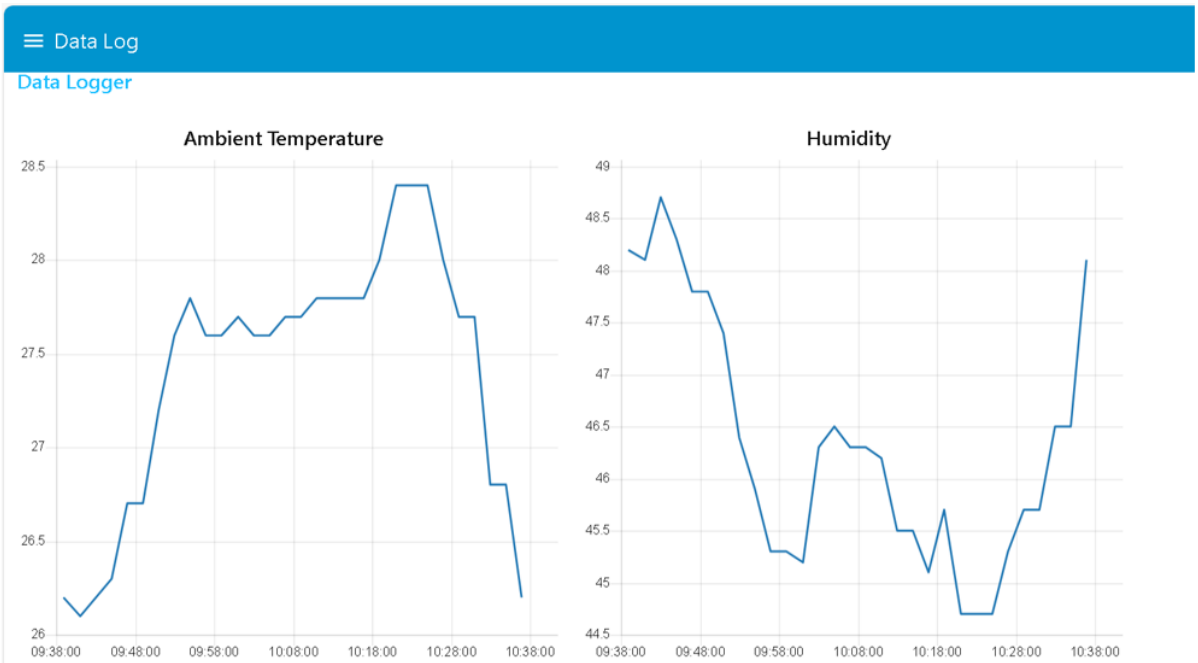


Figure 4.9: Node-RED dashboard display with ambient temperature and humidity

Data collection for this study was conducted over four cycles from the years, 2023 to 2024. To assist with prediction, the weekly average values of ambient temperature and humidity parameters were calculated and recorded to link with plant growth details. During the data collection, no missing data were identified for both plant and water quality parameters, due to data being recorded manually. However, anomalies were detected during the water quality data capturing, where the pH meter provided a strange reading in one instance. A UPS unit was installed to provide backup power for sensor data collection, mitigating the impact of load shedding. This was immediately replaced with a functioning unit and calibrated accordingly. The study ensured the precise and systematic documentation of all parameters. All recorded parameters were cross-checked by a senior field engineer onsite as well.

4.2 Chapter summary

This chapter describes the aquaponics setup used for the field study and explains how plant data, environmental data, and water quality data were all collected and recorded. Plant details, including height, diameter, and the number of leaves, were all manually measured using the appropriate tools. Ambient temperature and humidity, were measured and recorded using IoT devices. Water quality parameters such as, water temperature, TDS, EC, and pH, were manually collected using a water quality test pen and recorded accordingly.

CHAPTER FIVE

MACHINE LEARNING EXPERIMENTATION

This chapter describes how the experiment was conducted using selected machine learning models. Initially, the hardware and software specifications used in the study to conduct the experiments are explained. This is followed by the data preparation processes, model development, model training and finally the model evaluation. The chapter then presents and discusses the results obtained from the various experiments. The study finally explores the application of SHapley additive explanations (SHAP) which assisted to identify the most influential features and interpreting the models.

5.1 Hardware and software specifications

The hardware and software specifications required for the machine learning experiment are explained in the following sections.

5.1.1 Hardware

The hardware specification used in this study is given in Table 5.1.

Table 5.1: Hardware specification used for this study

Operating System	Windows 10 Home Single Language Version 22H2
CPU	Intel(R) Core(TM) i7-1165G7 @ 2.80GHz
System Type	x64
RAM	8,00 GB

5.1.2 Software

Python programming language was used for machine learning as it supports a wide range of libraries focussing on data science and machine learning, including NumPy, matplotlib, pandas, sklearn, and others. It is also well-suited for web application development (Molin, 2021:3; Castro et al., 2023). Jupyter Notebook was the platform used to develop the machine learning model which is an open-source, browser-based tool. This tool serves as a virtual lab notebook for coding, results execution, documentation, and visualisations (Prathanrat & Polprasert, 2018; Wang et al., 2021). The Python code was typed into the Jupyter Notebook’s code cell (Géron, 2019:48).

The main libraries used for ML model development are as follows:

NumPy: Library that performs mathematical operations on arrays, including shape manipulation, mathematical primitives, and sorting (Molin, 2021:40; Castro et al., 2023).

Pandas: This library is built on top of the NumPy library and is primarily used for data analysis and manipulation (Molin, 2021:40; Castro et al., 2023). Pandas provide two primary data structures, namely DataFrame and Series which assists in working with data (Molin, 2021:49; Castro et al., 2023). A frequently used data structure is the Data Frame (Castro et al., 2023). The Data Frame is a two-dimensional data structure that comprises rows and columns (Molin, 2021:56; Castro et al., 2023). The Series class provides a data structure for single-type arrays (Molin, 2021:53).

Scikit-learn/ sklearn: is a popular machine learning library that assists to build a model through implementing various learning algorithms and evaluating their performance (Fenner, 2019:20; Joshi, 2020:222; Castro et al., 2023; Molin, 2021:538). It has the ability to create pipelines that streamlines the preprocessing process and ensures that both the training and testing sets are treated consistently (Molin, 2021:570).

Matplotlib: This library is used to create a wide range of plots and visualisations for data analysis (Castro et al., 2023).

Table 5.2 lists the libraries and versions used to develop the machine learning models.

Table 5.2: Libraries used for the experiment

Libraries	Version
Python	3.9.0
sklearn	1.3.0
shap	0.46.0
Pandas	2.2.3
xGboost	1.7.4
matplotlib	3.5.1

5.2 Dataset

Two datasets were used in this study for predictions. One focused on plant-related predictions, and the other was used for water quality-related predictions.

I. **Plant dataset**

   The collected data were stored in an Excel sheet which contains 709 records. Details such as the number of leaves, plant diameter, plant height, ambient temperature, and ambient humidity were recorded.

II. **Water dataset**

   The collected data were stored in an Excel sheet into 526 records. These include details such as the pH, TDS, EC, and water temperature.

5.3 Data preparation

Data analysis helps to explore the hidden patterns, relations between variables and trends (Dangeti, 2017:11; Humayun et al., 2023). Data pre-processing prepares the raw data for analysis by addressing missing values, noisy data, and inconsistent formatting (Abdelaziz et al., 2025). Data preparation (wrangling) is part of data analysis (Molin, 2021:6). Data cleaning and data transformation are two tasks undertaken in this study as data preparation (Molin, 2021:119). During the data cleaning phase, redundant records were removed, and outliers were replaced with the median. After this process, the plant dataset contained 691 records, while the water dataset had 524 records.

The feature selection process is important in data preprocessing (Abdelaziz et al., 2025). The dependent and independent variables were identified from both datasets to predict plant diameter, plant height, water pH, and water TDS. Thereafter, the dataset was split into a standard percentage ratio of 80:20 as a training and test set, respectively (Géron, 2019:31; Kumar et al., 2023; Daniel et al., 2025).

The features in the dataset had different ranges, which can result in increased complexity and confusion. To improve model performance, it is necessary to scale all features to the same range through the feature engineering process (Molin, 2021:633; Keerthana et al., 2021; Abdelaziz et al., 2025). Hence, MinMaxScaler normalisation was applied separately to the feature and target variables in both the training and testing sets. This ensures that the values are scaled to fall within the range of 0 to 1 (Géron, 2019:72; Molin, 2021:562; Seegobin et al., 2024). After data cleaning, splitting, and scaling, the dataset was ready for model training (Obiora et al., 2023).

5.4 Model development

The supervised ML models Linear Regression (LR), Random Forest (RF), Deep Multilayer Perceptron (DML), eXtreme Gradient Boosting (XGBoost) and k-nearest neighbour (KNN) were selected for the prediction experiment. The plant and water quality datasets were used in the experiment (Islam et al., 2018; Keerthana et al., 2021; Ghandar et al., 2021; Kumar et al., 2023; Khandakar et al., 2024; Liu & Jiang, 2024). Hyperparameters were selected to structure the models, and this helps to enhance the model's performance (Elgeldawi et al., 2021). Hyperparameter tuning directly impacts the accuracy and generalisation capabilities of machine learning models (Dabool et al., 2024). Widely used hyperparameter tuning techniques, such as *Gridsearch* and *Randomsearch* were used to optimise the model's performance (Bischl et al., 2023; Dabool et al., 2024). *Gridsearch* performs an exhaustive search over a predefined set of hyperparameters, whereas *Randomsearch* performs a randomised search over hyperparameters to find optimal combinations for improving model performance (Yu & Zhu, 2020; Dabool et al., 2024). *Randomsearch* may perform better compared to *Gridsearch,* particularly when some hyperparameters are not uniformly distributed. Table 5.3 shows the various hyperparameters selected for the respective ML models. The best hyperparameter values were thus selected using this process (Dangeti, 2017,117; Molin, 2021,625; Ubayasena et al., 2023; Khandakar et al., 2024).

The *GridSearch*CV and *RandomSearch*CV classes from Scikit-learn were used in this study for hyperparameter tuning (Molin, 2021:627; Khandakar et al., 2024; Abdelaziz et al., 2025). To split the data into multiple combinations, 5-fold and 10-fold cross-validation were used in hyperparameter tuning, which are commonly used and recommended (Fenner, 2019: 129). 5-fold allows each model to be trained and tested, which reduces the chances of overfitting and provides a more accurate model performance assessment (Ozaga et al., 2024). 10-fold provides robust estimation because it uses 90% of the data for testing, whereas 10 % of the data is used for testing (Lantz, 2013;319, 322). The random splitting of data ensured a reduction in the chances of coincidental features getting more importance (Müller & Guido, 2016:252, 254; Joshi, 2020:166). Scikit-learn’s k-fold cross-validation was used for cross-validation (CV) (Molin, 2021:628).

Table 5.3: Hyperparameters selected for ML models

Models	Hyperparameters	Description	Reference
Random Forest	n_estimators	The number of trees in the random forest	(Molin, 2021:206; Olafadehan & Ahaotu: 2023)
	max_features	The maximum number of features that are evaluated for splitting at each node	(Molin, 2021:185; Olafadehan & Ahaotu: 2023)
	max_depth	The maximum depth of the tree. The Decision Tree stops right there.	(Molin, 2021:180; Olafadehan & Ahaotu:2023)
	min_samples_split	The minimum number of samples a node must have before it can be split.	(Molin, 2021:185; Olafadehan & Ahaotu:2023)
	min_samples_leaf	The minimum number of samples a leaf node must have	(Molin, 2021:185; Olafadehan & Ahaotu: 2023)
	forest__bootstrap	Whether bootstrap samples are used or not	(Molin, 2021:196)
KNN	n_neighbors	Number of neighbours	(Molin, 2021:663)
	weights	Each neighbour's impact on the prediction	(Fenner, 2019:363)
	Algorithm	The algorithm used to compute the nearest neighbours	(Giuseppe, 2018:289; Fenner, 2019:363)
	leaf_size	Leaf size of the tree-based algorithm	(Giuseppe, 2018:289; Fenner, 2019:363)
XGBoost	learning_rate	determines the contribution each tree will make to the final estimator	(Molin, 2021:655)
	n_estimators	Number of trees to control the ensemble training	(Géron, 2019:206)
	max_depth	Maximum depth of the tree	(Géron, 2019:182)
	Subsample	The fraction of training instances to be used for training each tree.	(Géron, 2019:209)
MLP	hidden_layer_size	Number of neurons in the hidden layer	(Dangeti, 2017:344)
	max_iter	Maximum number of iterations	(Dangeti, 2017:39)
	Activation	Activation function	(Dangeti, 2017:262)
	Solver	Optimiser for the reduction of errors	(Dangeti, 2017:262)
	Alpha	Regularisation strength to avoid overfitting	(Dangeti, 2017:262; Molin, 2021:669; Fenner, 2019:300)
	learning_rate	Used to control the rate of convergence of the algorithm	(Dangeti, 2017:287)
	batch_size	Number of observations considered at each iteration	(Géron, 2019:321; Dangeti, 2017:39)
	early_stopping	Stop training as soon as the validation error reaches a minimum	(Géron, 2019:142)
	learning_rate_init	Initial learning rate	(Géron, 2019:355)
Linear Regression	fit_intercept	Exploration into models with different biases	(Dangeti, 2017:38; Fenner, 2019:340; Olafadehan & Ahaotu, 2023; Khandakar et al., 2024)
	copy_x	Copy all variables in the dataset.	(Fenner, 2019:382; Khiem et al., 2022;

			Olafadehan & Ahaotu, 2023)

5.5 Model training

The selected supervised ML algorithms were used to train and predict the target variables. T*o* optimise the model performance, the hyperparameters were fine-tuned using *Gridsearch* and *Randomsearch* (Ubayasena et al., 2023; Khandakar et al., 2024). Hyperparameter fine-tuning was done for the different prediction cases (plant diameter, plant height, water pH, and water TDS). The optimal hyperparameters that produced the best model performance for plant diameter, plant height, water pH, and water TDS predictions are presented in Tables 5.4 – 5.7.

Table 5.4: Hyperparameters used in *Gridsearch* with 10-fold CV for plant diameter prediction

ML models	**Hyperparameters**
Linear Regression	fit_intercept: [True, False] copy_x: [True, False]
Random Forest	n_estimators: [50, 112, 175, 237, 300] max_features: ['sqrt', 'log2', 0.5] max_depth: [10, 20, 30, None] min_samples_split: [2, 5, 10] min_samples_leaf: [1, 2, 4] bootstrap: [True, False]
KNN	n_neighbors: np.arange(1, 10) weights: ['uniform', 'distance'] algorithm: ['auto', 'ball_tree', 'kd_tree', 'brute'] leaf_size: np.arange(20, 40, 5)
XGBoost	learning_rate: [0.01, 0.1, 0.2] n_estimators: [100, 200], max_depth: [3, 4, 5], subsample: [0.8, 0.9, 1.0],
MLP	hidden_layer_sizes: [(200, 150, 100, 50),(150, 100, 50), (120, 80, 40), (100, 50, 30), (50, 30)] max_iter: [10000, 50000] activation: ['relu'], solver: ['adam'], alpha: [0.0001, 0.001, 0.01] learning_rate: ['constant', 'adaptive'] batch_size: ['auto', 32, 64, 100,128] early_stopping': [True, False] learning_rate_init : [0.001, 0.01, 0.1]

Table 5.5: Hyperparameters used in *Gridsearch* with 5-fold CV for plant height prediction

ML models	Hyperparameters
Linear Regression	fit_intercept: [True, False] copy_x: [True, False]
Random Forest	n_estimators: [10, 50, 112, 175, 237, 300] max_features: [1.0] max_depth: [10, 20, 30, None] min_samples_split: [2, 5, 10] min_samples_leaf: [1, 2, 4] bootstrap: [True, False]
KNN	n_neighbors: np.arange(1, 5) weights: ['uniform', 'distance'] algorithm: ['auto', 'ball_tree', 'kd_tree', 'brute'] leaf_size: np.arange(10, 30, 60)
XGBoost	learning_rate: [0.01, 0.05, 0.1] n_estimators: [50, 100, 200] max_depth: [2, 3, 4] subsample: [0.8, 0.9, 1.0]
MLP	hidden_layer_sizes: [(200, 150, 100, 50), (150, 100, 50), (120, 80, 40), (100, 50, 30), (50, 30)] max_iter: [10000, 20000, 50000], activation: ['relu', 'tanh'] solver: ['adam'], alpha: [0.0001, 0.001, 0.01] learning_rate: ['constant', 'adaptive'] batch_size: ['auto', 32, 64, 100,128] early_stopping': [True, False] learning_rate_init : [0.001, 0.005, 0.01]

Table 5.6: Hyperparameters used in *Gridsearch* with 10-fold CV for water pH prediction

ML models	Hyperparameters
Linear Regression	fit_intercept: [True, False] copy_x: [True, False]
Random Forest	n_estimators: [50, 112, 175, 237, 300] max_features: ['sqrt', 'log2', 0.5] max_depth: [10, 20, 30, None] min_samples_split: [2, 5, 10] min_samples_leaf: [1, 2, 4] bootstrap: [True, False]
KNN	n_neighbors: np.arange(1, 5) weights: ['uniform', 'distance'] algorithm: ['auto', 'ball_tree', 'kd_tree', 'brute']

	leaf_size: np.arange(25, 30, 5)
XGBoost	learning_rate: [0.01, 0.1, 0.2] n_estimators: [100, 200] max_depth: [3, 4,5] subsample: [0.8, 0.9, 1.0]
MLP	hidden_layer_sizes: [(200, 150, 100, 50),(150, 100, 50), (120, 80, 40), (100, 50, 30), (50, 30)] max_iter: [10000, 50000], activation: ['relu'] solver: ['adam'], alpha: [0.0001, 0.001, 0.01] learning_rate: ['constant', 'adaptive'] batch_size: ['auto', 32, 64, 100,128] early_stopping': [True, False] learning_rate_init : [0.001, 0.01, 0.1]

Table 5.7: Hyperparameters used in *Gridsearch* with 10-fold CV for water TDS prediction

ML models	**Hyperparameters**
Linear Regression	fit_intercept: [True, False] copy_x: [True, False]
Random Forest	n_estimators: [50, 112, 175, 237, 300] max_features: ['sqrt', 'log2', 0.5] max_depth: [10, 20, 30, None] min_samples_split: [2, 5, 10] min_samples_leaf: [1, 2, 4] bootstrap: [True, False]
KNN	n_neighbors: np.arange(1, 10) weights: ['uniform', 'distance'] algorithm: ['auto', 'ball_tree', 'kd_tree', 'brute'] leaf_size: np.arange(20, 40, 5)
XGBoost	learning_rate: [0.01, 0.1, 0.2] n_estimators: [100, 200] max_depth: [3, 4,5] subsample: [0.8, 0.9, 1.0]
MLP	hidden_layer_sizes: [(200, 150, 100, 50),(150, 100, 50), (120, 80, 40), (100, 50, 30), (50, 30)] max_iter: [10000, 50000] activation: ['relu'] solver: ['adam'], alpha: [0.0001, 0.001, 0.01] learning_rate: ['constant', 'adaptive'] batch_size: ['auto', 32, 64, 100,128] early_stopping': [True, False] learning_rate_init : [0.001, 0.01, 0.1]

Linear Regression	fit_intercept: [True, False] copy_x: [True, False]

5.6 Model performance evaluation

The ML models were evaluated on the test dataset using the metrics Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R-squared (R^2), and Adjusted R-squared (Adjusted R^2) to assess their performance (Dangeti, 2017:29; Chicco et al., 2021). The evaluation metrics, purpose, and the best and worst value range are presented in Table 5.8.

Table 5.8: Evaluation metrics, purpose and value range

Evaluation metrics	Purpose	Value range	Reference
Mean Squared Error (MSE)	Squares of the difference between the predicted and actual value	best value = 0 and worst value = $+\infty$	(Chicco et al., 2021; Kumar et al., 2023; Priya, 2021; Sudheer et al., 2022)
Root Mean Squared Error (RMSE)	The square root of the Mean Squared error	best value = 0 and worst value = $+\infty$	(Chicco et al., 2021; Priya, 2021; Kumar et al., 2023)
Mean Absolute Error (MAE)	Difference between the predicted value and the actual value	best value = 0 and worst value = $+\infty$	(Chicco et al., 2021; Priya, 2021; Kumar et al., 2023)
R-squared (R^2)	Difference in variance with dependent variables	best value = +1 worst value = $-\infty$	(Dangeti, 2017:29; Chicco et al., 2021; Priya, 2021; Sudheer et al., 2022)
Adjusted R-squared (Adjusted R^2)	R-squared is adjusted for the number of independent variables in the model.	Less than or equal to R^2	(Sudheer et al., 2022)

The evaluation metrics and the corresponding scores of the best-performing models for different predictions, namely plant diameter, plant height, water pH, and water TDS, are presented in the subsequent sections. Both *Gridsearch* and *Randomsearch* were used for hyperparameter tuning with 5-fold and 10-fold cross-validation. The results of the best models are discussed here, while the remaining scores are provided in Appendix C for simplicity.

5.6.1 Plant diameter prediction

Table 5.9 presents the evaluation scores for plant diameter prediction experiments when *Gridsearch* 10-fold cross-validation was used.

Table 5.9: Plant diameter prediction using *Gridsearch* with 10-fold CV

ML models	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R-squared	Adjusted- R-squared
Linear Regression	0.02	0.15	0.11	0.54	0.53
Random Forest	0.00	0.05	0.03	0.94	0.94
KNN	0.01	0.09	0.06	0.82	0.82
XGBoost	0.00	0.05	0.03	0.94	0.94
MLP	0.00	0.07	0.04	0.86	0.86

Based on the evaluation metrics scores (Table 5.9) random forest (RF), and XGBoost produced the best performance for plant diameter prediction. This was followed by MLP, KNN, and Linear Regression. Both random forest (RF) and XGBoost achieved minimal error metrics, with MSE, RMSE, and MAE values of 0.00, 0.05, and 0.03, respectively. These models also demonstrated high predictive accuracy, as reflected in their R-squared and Adjusted R-squared scores of 0.94 (94%).

5.6.2 Plant height prediction

Table 5.10 depicts the performance evaluation scores for plant height prediction using *Gridsearch* with 5-fold cross-validation.

Table 5.10: Plant height prediction using *Gridsearch* with 5-fold CV

ML models	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R-squared	Adjusted-R-squared
Linear Regression	0.01	0.09	0.06	0.80	0.80
Random Forest	0.00	0.06	0.05	0.93	0.92
KNN	0.00	0.06	0.04	0.91	0.91
XGBoost	0.00	0.06	0.05	0.92	0.92
MLP	0.00	0.06	0.04	0.92	0.92

Gridsearch with 5-fold cross-validation provided the best overall performance. The Random Forest (RF) model, when tuned using *Gridsearch* with 5-fold cross-validation, achieved the highest performance (see Table 5.10). This was followed by MLP, XGBoost, KNN, and Linear Regression. RF achieved minimal MSE, RMSE, and MAE values of 0.00, 0.06, and 0.05, respectively, along with high R-squared and Adjusted R-squared scores of 93% and 92%. 5-fold cross-validation with *Gridsearch* for RF produced the most suitable models for plant height prediction, followed by MLP, XGBoost, KNN, and Linear Regression

5.6.3 Water pH prediction

The best-performed evaluation scores for water pH prediction are shown in Table 5.11.

Table 5.11: Water pH prediction using *Gridsearch* with 10-fold CV

ML models	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R-squared	Adjusted- R-squared
Linear Regression	0.03	0.19	0.15	0.55	0.54
Random Forest	0.02	0.13	0.09	0.78	0.77
KNN	0.02	0.13	0.09	0.78	0.77
XGBoost	0.02	0.13	0.09	0.79	0.79
MLP	0.03	0.18	0.15	0.60	0.59

In the water pH prediction experiment, 10-fold cross-validation using *Gridsearch* achieved the best overall performance (Table 5.11). XGBoost performed the best among all models in the 10-fold cross-validation using *Gridsearch* for water pH prediction. This was followed by RF, KNN, MLP, and Linear Regression. The XGBoost model achieved minimal MSE, RMSE, and MAE error values of 0.02, 0.13 and 0.09, along with high R-squared and Adjusted R-squared scores of 79%.

5.6.4 Water TDS prediction

The water TDS prediction performance evaluation, scored using *Gridsearch* 10-fold cross-validation experiment, is shown in Table 5.12.

Table 5.12: Water TDS prediction using *Gridsearch* with 10-fold CV

ML models	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R-squared	Adjusted- R-squared
Linear Regression	0.00	0.01	0.01	1.00	1.00
Random Forest	0.00	0.03	0.01	0.97	0.97
KNN	0.00	0.02	0.01	0.99	0.99
XGBoost	0.00	0.03	0.01	0.98	0.98
MLP	0.00	0.02	0.01	0.99	0.99

In the water TDS prediction experiments, *Gridsearch* with 10-fold cross-validation delivered the best overall performance (Table 5.12). The Linear Regression model particularly achieved minimal error values of 0.00 for MSE, 0.01 for RMSE, and 0.01 for MAE, along with R-squared and Adjusted R-squared scores of 100%. This was followed by KNN, MLP, XGBoost, and Linear Regression. Experiments on plant diameter, plant height, water pH, and water TDS prediction using Linear Regression, Random Forest, XGBoost, KNN, and MLP were conducted with *Gridsearch* and *Randomsearch* using 5-fold and 10-fold cross-validation. Overall, *Gridsearch* delivered better performance scores compared to *Randomsearch*.

5.7 Model explainability

Explainable AI aims to clarify and interpret machine learning models. In this study, SHapley Additive Explanations (SHAP) was used to present the mean absolute SHAP values through a bar graph and provide a global explanation of the selected models' predictions using a summary plot (Linardatos et al., 2021; Ekanayake et al., 2022). A deeper understanding of the features or parameters in an aquaponics system provides insight into their interdependencies and their combined impact on achieving optimal plant and fish production within the system.

Bar graph: The visualisation demonstrates how each feature influences the prediction. The bars are coloured red and blue. Red bars denote features that positively influence the prediction, while blue bars denote features that negatively influence it. The length/size of each bar signifies the strength of the feature's effect on the model's prediction, with longer bars indicating a stronger influence. The order of the bars in the graph reflects the importance of each feature in influencing the model's prediction, from the most influential feature to the least influential feature.

Summary plot: Visualises how each feature contributes to the model's predictions throughout the entire dataset. The red dot indicates high feature values, whereas the blue dot indicates low feature values. Points that are further from zero on the X-axis denote features with a higher or lower contribution to the prediction.

5.7.1 Plant diameter

The feature height has the greatest influence in predicting plant diameter, as shown in Figure 5.1, with a mean absolute SHAP value of +0.09, indicating that as plant height increases, the predicted diameter also increases, positively impacting the prediction. The second most influential feature is leaves, with a mean absolute SHAP value of +0.04, which also contributes positively to the diameter prediction. In comparison, humidity and temperature have relatively smaller positive effects, with mean absolute SHAP values of +0.03 and +0.02, respectively.

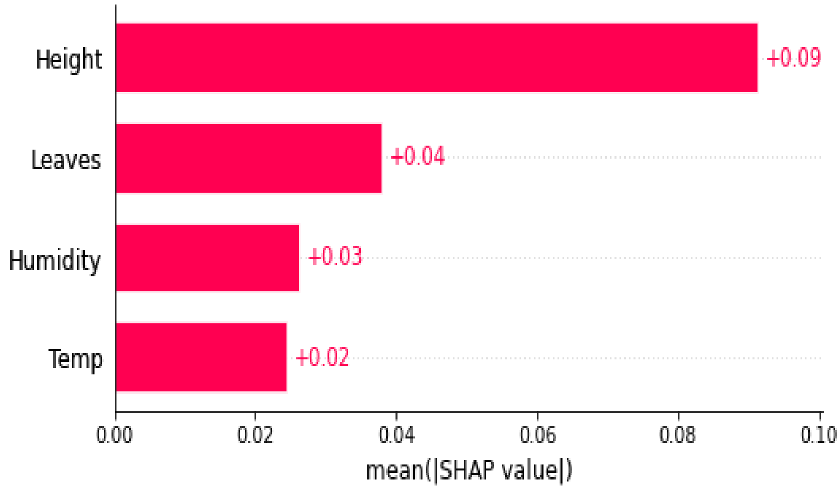


Figure 5.1: Mean absolute SHAP values of the random forest model for plant diameter prediction

The SHAP analysis for plant diameter prediction showed that plant height is the most influential factor, with the number of leaves, ambient humidity, and ambient temperature following in order of importance. Plants' morphological characters, such as plant height and number of leaves, served as key indicators in this study (Alshammari et al., 2024). Ambient humidity and temperature directly influence plant growth and plant development, as they are essential for transpiration and the photosynthetic processes (Chia & Lim, 2022). High humidity during transpiration can reduce air circulation, causing plants to halt transpiration and nutrient uptake from the growing medium. Long periods of such humidity saturation may lead to gradual rotting of the plants. Higher temperatures contribute to speeding up physiological processes with positive and negative effects. The increased temperatures promote faster growth and higher yield however, it also on the other hand removes the functional components from leaves due to high transpiration rates (Chowdhury et al., 2021). Fluctuations in ambient temperature affect atmospheric moisture levels, thereby causing changes in ambient humidity. Hence, it is essential to monitor and maintain ambient humidity and temperature for optimal plant growth.

According to Figure 5.2, higher values of plant height led to an increase in the predicted plant diameter, while a lower number of leaves resulted in a slight decrease in the prediction. Low humidity levels are associated with an increase in the predicted plant diameter, whereas high humidity levels have a slight decreasing effect. However, the temperature values are more evenly distributed, with both positive and negative impacts on the predicted diameter. High temperatures can either positively or negatively affect the prediction, depending on the specific data point.

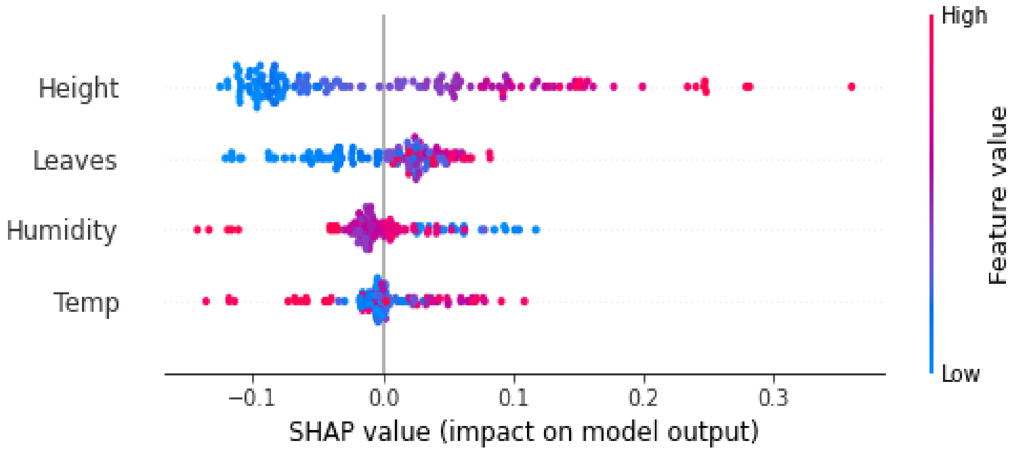


Figure 5.2: SHAP global explanation of the random forest model for plant diameter prediction

5.7.2 Plant height

The diameter feature exhibits the highest mean absolute SHAP value of +0.1 as shown in Figure 5.3, making it the most influential factor in the model's prediction of plant height. This indicates that as the diameter increases, the model predicts a higher plant height. The second most influential feature is leaves, with a mean absolute SHAP value of +0.06, which also positively contributes to the prediction. Among the environmental factors, temperature and humidity play a less significant role, with temperature showing a mean absolute SHAP value of +0.04 and humidity having the least positive contribution with a mean absolute SHAP value of +0.01.

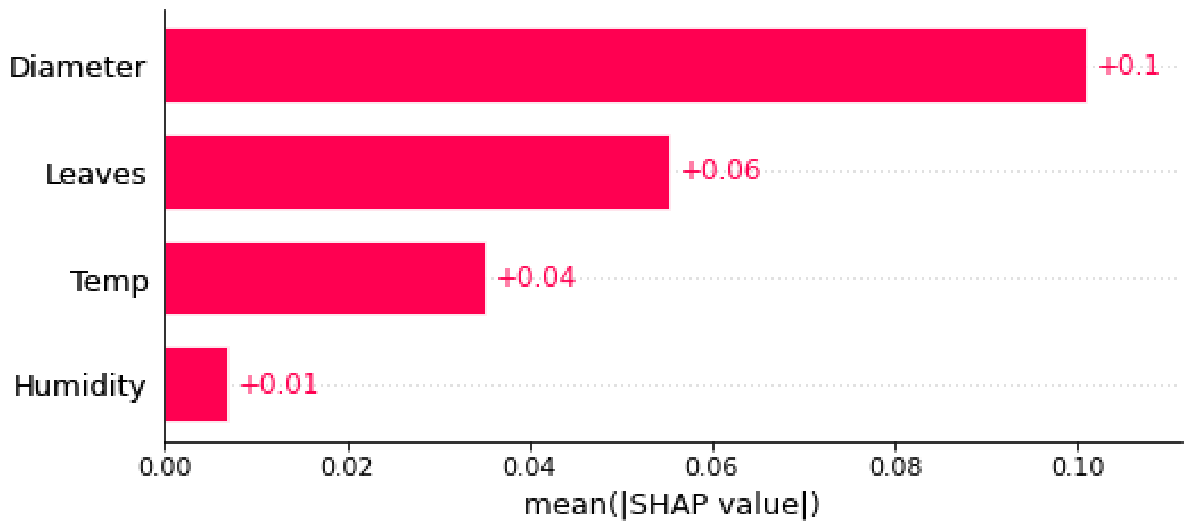


Figure 5.3: Mean absolute SHAP values of the random forest model for plant height prediction

In plant height prediction, SHAP analysis revealed that plant diameter had the most significant influence, followed by the number of leaves, ambient temperature and ambient humidity. Plant growth was determined by the width of the leaves and the number of leaves. Ambient temperature influences the speed of energy processing in plants. Furthermore, humidity has a direct bearing on the photosynthesis process and thus influences the growth and development of plants (Chia & Lim, 2022).

According to Figure 5.4, high values for the features' diameter, leaves, and temperature contribute to an increase in the predicted plant height. In contrast, humidity is more evenly distributed between positive and negative contributions. However, high humidity slightly increases the predicted plant height, while low humidity slightly decreases it.

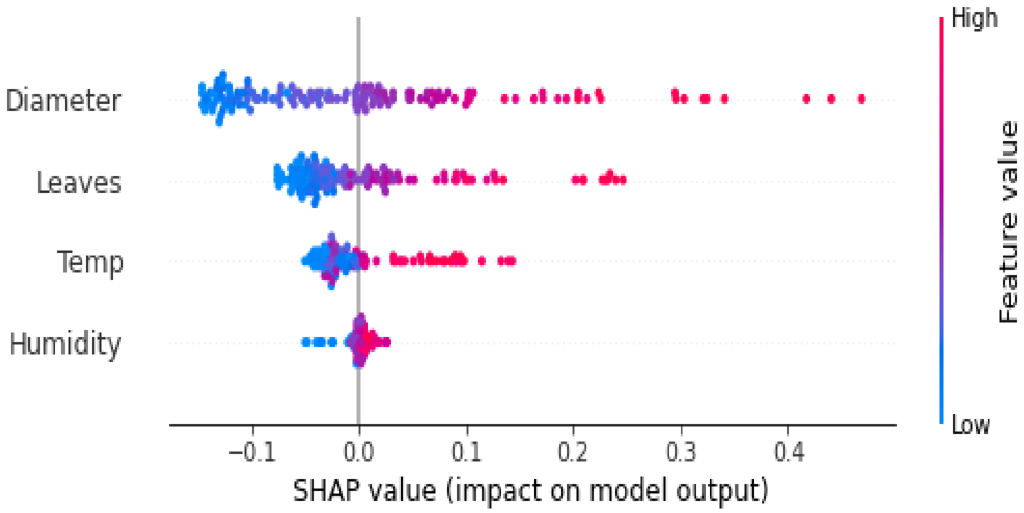


Figure 5.4: SHAP global explanation of the random forest model for plant height prediction

5.7.3 Water pH

The TDS feature demonstrated the highest mean absolute SHAP value of +0.19 as seen in Figure 5.5, making it the most influential factor in the pH model's prediction. This indicates that as the TDS increases, the model predicts a higher pH value. The second most influential feature is temperature, with a mean absolute SHAP value of +0.06, which also positively contributed to the prediction. EC has the least positive contribution with a mean absolute SHAP value of +0.03.

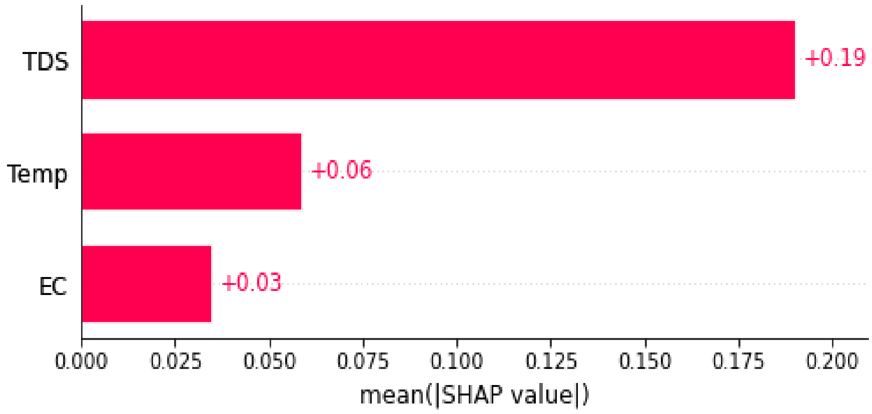


Figure 5.5: Mean absolute SHAP values of XGBoost for water pH prediction

In the pH prediction SHAP analysis, TDS indicated the most influential feature, followed by water temperature and EC. The pH level indicates the acidity or alkalinity of the water. This has a direct effect on how well the fish and other organisms survive in the water. There is a strong correlation between EC and TDS in the water. The relationship between EC and TDS was influenced by the temperature and pH of the water (Dewangan & Shrivastava, 2024). When the water temperature and pH increase, more toxic ammonia is produced (Maulini et al., 2022). Excess acid or alkali in the water can be toxic for many organisms as well, and thus, it is critical to monitor and maintain the pH level as much as possible (Kok et al., 2024)

According to Figure 5.6, low TDS values are associated with an increase in the predicted pH value. Also, it is noted that low-temperature values contributed to both increases and decreases in the predicted pH value. Additionally, low EC values slightly increase the predicted pH value, whereas high EC value decreases the predicted pH value.

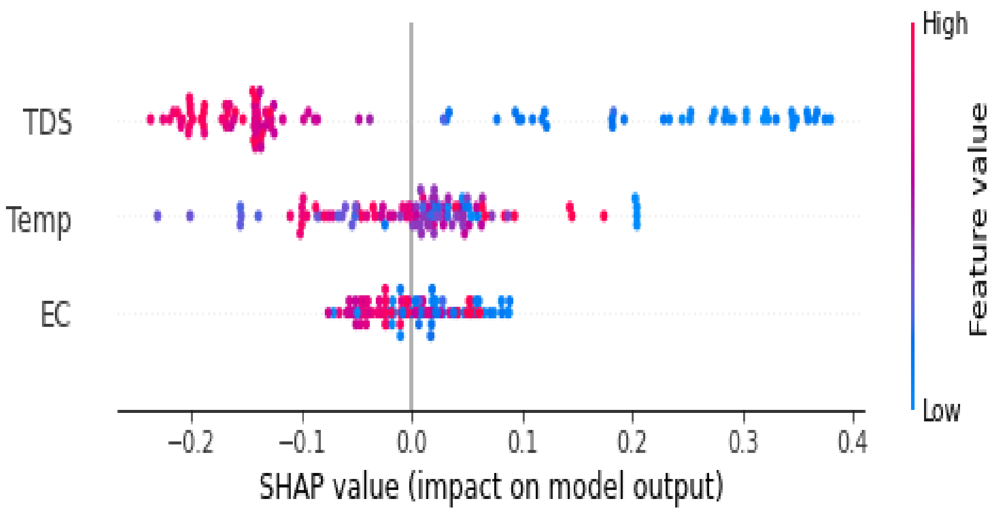


Figure 5.6: SHAP global explanation of the XGBoost model for water pH prediction

5.7.4 Water TDS

The EC feature exhibits the highest mean absolute SHAP value of +0.18, making it the most influential factor in the TDS model's prediction, as shown in Figure 5.7. This indicates that as EC increases, the model predicts a higher TDS value. The second most influential feature is pH, with a mean absolute SHAP value of +0.01, which also contributes positively to the prediction. However, the feature temperature, with a mean absolute SHAP value of +0, does not significantly contribute to the TDS prediction.

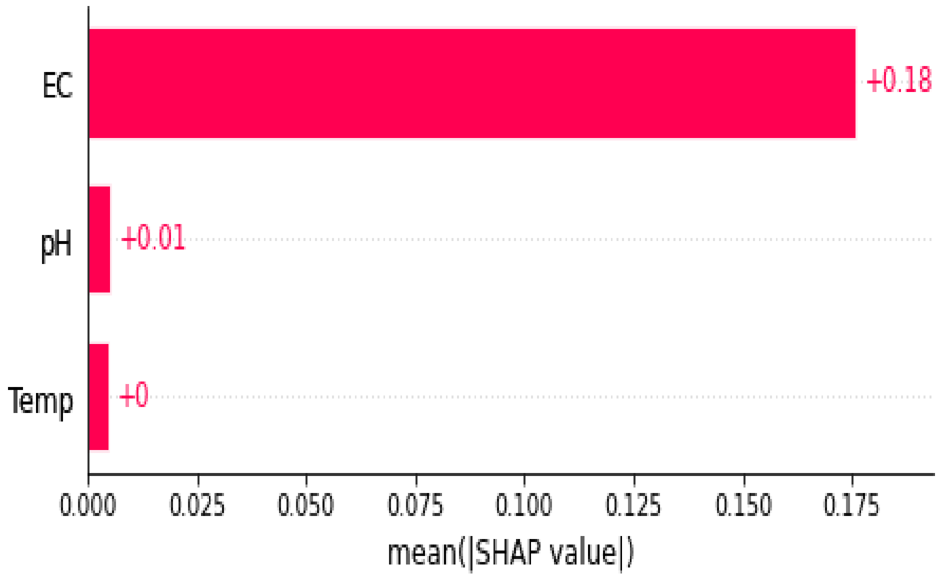


Figure 5.7: Mean absolute SHAP values of linear regression for water TDS prediction

Based on the SHAP analysis for TDS prediction, EC appeared as a strong influencing feature due to its high correlation with TDS. TDS reflects the amount of total nutrients, concentration of dissolved ions, salt and organic matter present in the water, whereas EC measures the ability of water to conduct electricity. Dissolved solids in water consist of ions, which are responsible for its ability to conduct electricity. This creates a strong correlation between EC and TDS, as an increase in the concentration of dissolved ions leads to higher EC values. Therefore, EC can be used as an indicator of TDS in water. However, the relationship between these two parameters is not always linear, as their behaviour can be influenced by various factors such as pH, water temperature, and the types of dissolved solids present in the water (Dewangan & Shrivastava, 2024). Temperature can affect both EC and TDS. Higher temperatures increase the electrical conductivity of water by enhancing ion mobility and also raise the solubility of salts and certain minerals, resulting in higher TDS levels. Therefore, it is important to measure water temperature along with EC and TDS (Dewangan et al., 2023).

According to Figure 5.8, high EC values increase the predicted TDS value, while low EC values decrease it. Similarly, low pH values slightly increase the prediction, whereas high pH values slightly decrease it. In addition, high-temperature values slightly increase the prediction, while low-temperature values slightly decrease it.

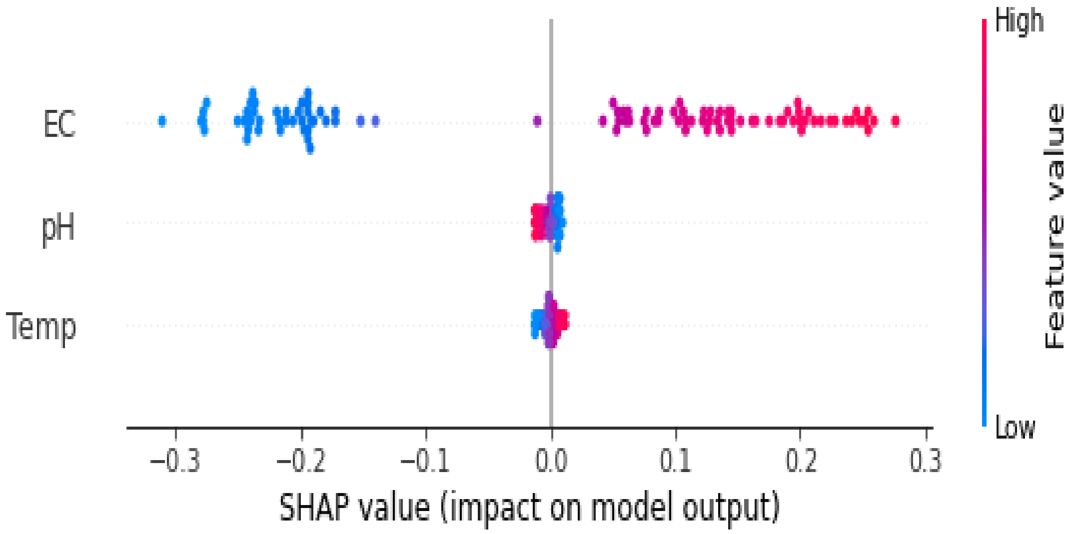


Figure 5.8: SHAP global explanation of the linear regression model for water TDS prediction

5.8 Chapter summary

In this chapter, the hardware and software specifications used to perform the machine learning experiment are described in detail. The experiment involved applying various models to make predictions, followed by an evaluation of their performance using regression metrics. Finally, the chapter explored feature importance, using SHAP (SHapley Additive Explanations) to understand the contribution of each feature to the model's predictions.

CHAPTER SIX
DECISION SUPPORT SYSTEM DEVELOPMENT AND EVALUATION

This chapter outlines the development and evaluation of the decision support system for aquaponics plant growth and water quality predictions. The system was developed using the Flask framework and deployed on PythonAnywhere. The System Usability Scale (SUS) was used to evaluate the usability of the developed system. The SUS is a reliable, free-of-cost instrument used for worldwide assessments of system usability applications (Brooke, 1996; Kortum & Bangor, 2013).

6.1 Requirements of the decision support system for aquaponics prediction

Aquaponics is a complex system that combines various disciplines. Various aquaponic and environmental parameters are thus crucial to the monitoring and control. However, deciding which parameters to be monitored and controlled can be tricky and challenging, as requirements vary. Additionally, changes in one parameter can influence others within the system.

The requirements of the aquaponics decision support system were identified based on research gaps in the literature and the key features of a DSS (Ghandar et al., 2021; Pechlivani et al., 2025). These include support for semi-structured or unstructured decision-making, provision of accurate predictions and actionable insights, and an interactive user interface (Darbi & Saleh, 2022; Pechlivani et al., 2025). Previous studies have primarily concentrated on developing predictive models and have not addressed the model explainability or the translation of model outputs to support decision-making (Ghandar et al., 2021; Amano et al., 2022; Debroy & Seban, 2022; Owusu et al., 2024; Liu et al., 2024; Khandakar et al., 2024; Liu & Jiang, 2024). To address these overlooked areas, the researcher defined requirements that would enable the DSS to predict key aquaponics parameters, such as plant growth, water quality, and to also rank the influencing factors in priority order. These functionalities will provide stakeholders with a clear direction on which parameters should be considered for aquaponics monitoring and control. This will improve both system performance and decision-making effectiveness.

The requirements of the proposed decision support system are the following:

i. It must be able to assist stakeholders in recognising the key parameters that require monitoring and controlling.
ii. The system must predict plant height, plant diameter, water pH, and water TDS based on user inputs.
iii. The system ranks the influencing parameters from highest to lowest.

iv. This system's ranking must provide guidance on which parameters need to be prioritised for monitoring and control to ensure optimal system performance.
v. The stakeholders must be able to provide feedback on the usability of the system.
vi. The system must be accessible from mobile devices and computers on any browser over the internet.

6.2 System design of the decision support system

This study aimed to develop a decision support system that assists stakeholders in the decision-making process. To achieve this, a data-driven decision support system was designed based on machine learning (ML). The purpose of the ML prediction was to identify the best algorithm for predicting plant diameter, plant height, water pH and water TDS (see Section 5.6). Thereafter, the best-performing algorithms were used to design a DSS using the Flask framework. The design allowed the stakeholders to provide input and receive predictions, accompanied by relevant explanations as a response from the DSS system. The system used ML-specific models to predict plant growth or water quality parameters. In addition, it presented the most influential parameters that contributed to the prediction, ranked from highest to lowest. Later, the system enabled users to evaluate its usability from their own perspective. The insights enabled stakeholders to monitor system performance more effectively and take corrective action on parameters according to their prioritisation, if necessary. The system has front-end and back-end components. The front-end handles the user interface, while the back-end is responsible for the business logic. The web-based architecture of the data-driven DSS system is shown in Figure 6.1.

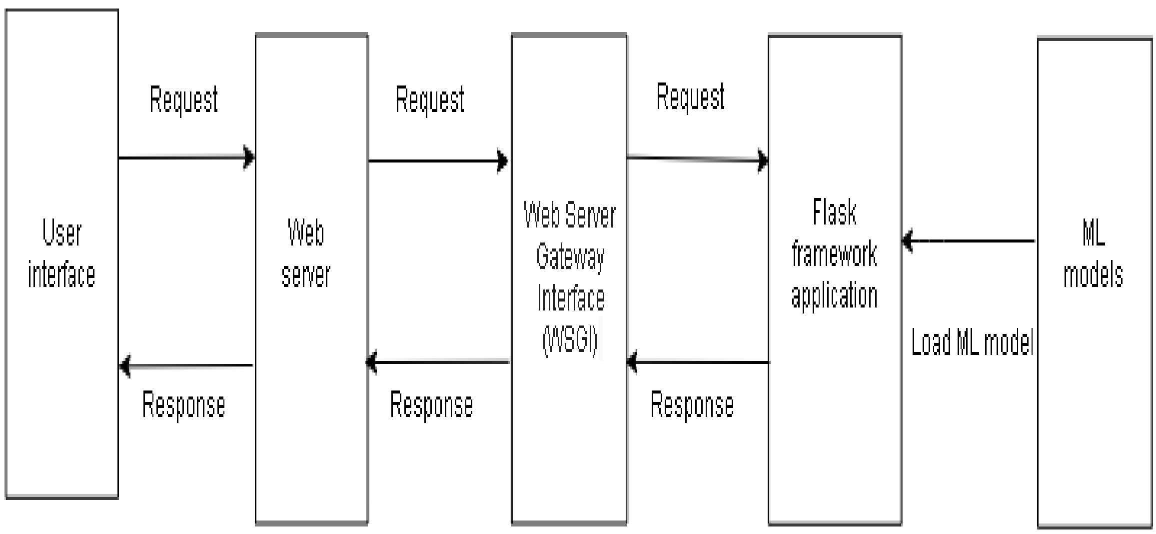


Figure 6.1: The web-based architecture of the data-driven DSS

Through the user interface (UI), users can access the system and send requests. These requests may involve loading web pages or making predictions. The request passes through the web server and the Web Server Gateway Interface (WSGI) to the Flask framework application. Based on the request, the system processes the data and sends a response back to the user.

The proposed DSS had a user interface that enabled users to interact effectively with the system. A user could navigate through the system, select the type of prediction to be performed, and then provide the required input data. After, the system will process the information provided and generate a prediction for the user. The process workflow of the decision support system is shown in Figure 6.2.

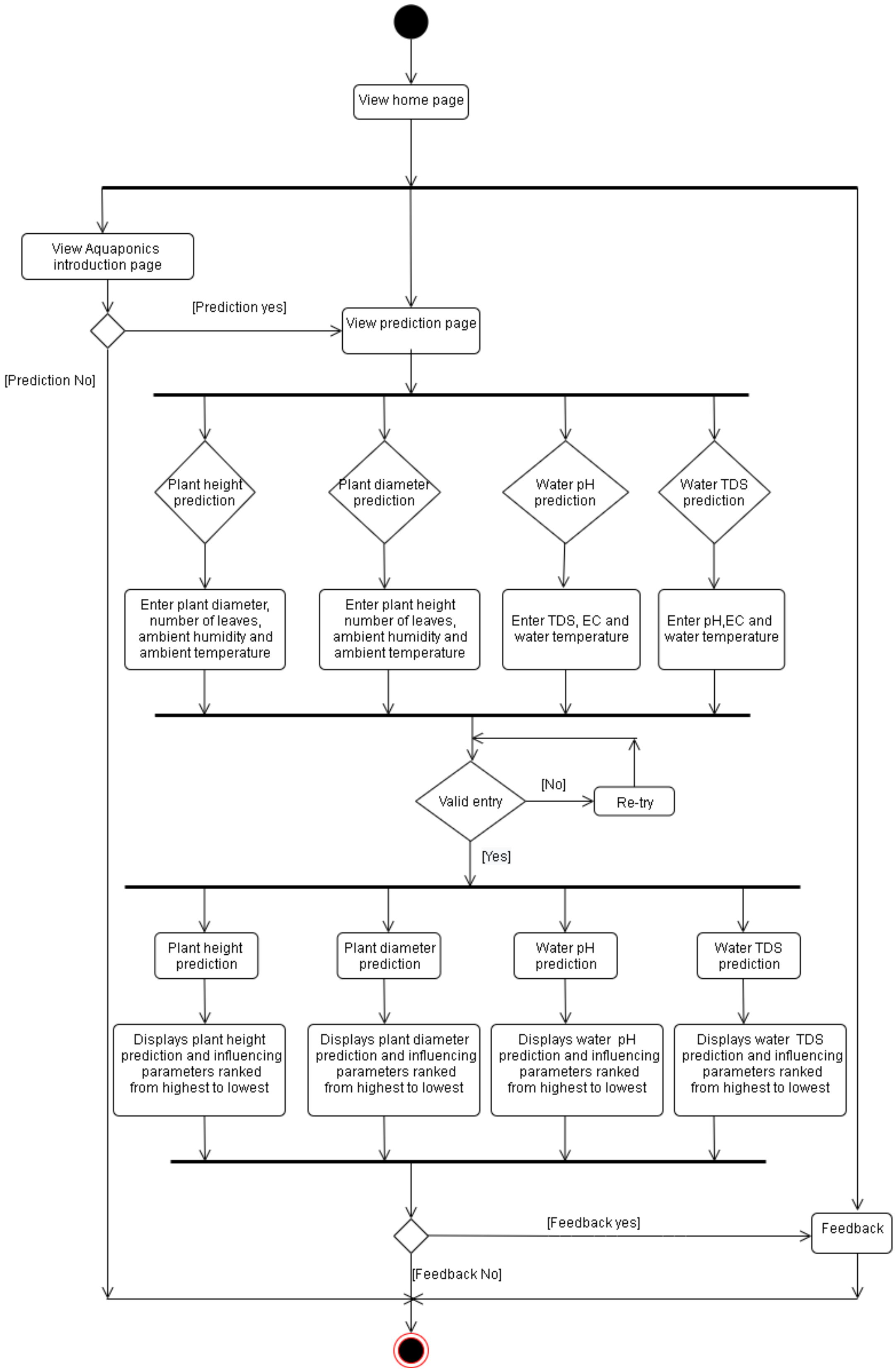


Figure 6.2: Process flow of the decision support system

The system loads at the home page, where a user is presented with a brief overview of the project's purpose. From there, a user can navigate through the system using the main menu options such as Aquaponics, Prediction, and Feedback. When a user selects the Aquaponics option, the user is provided with a brief introduction about aquaponics. Choosing the Prediction option allows a user to select one of the prediction options that are available, such as plant height, plant diameter, water pH, or water TDS. After making a choice by clicking a radio button, the user is prompted to input the relevant data into text fields and then required to click the Submit button to generate a prediction. The system first validates the input format. If the data entered is incorrect or incomplete, the user is prompted to re-enter the information. Once valid inputs are provided, the system processes the data in the backend and generates the prediction results along with a ranked list of the most influential parameters from the highest to lowest. If the user wishes to evaluate the usability of the system, it can be done by selecting the Feedback option, which allows the user to rate the usability of DSS based on the user's experience of the system.

6.3 Decision support system development

This section explains Python web application development using the Flask framework. It also expands on the web page layout and the deployment of the developed application on PythonAnywhere.

6.3.1 Flask framework

Flask is a lightweight micro framework for Python web development created by Armin Ronacher (Copperwaite & Leifer, 2015:1; Grinberg, 2018:3; Mufid et al., 2019). Flask has three main dependencies: routing, debugging and Web Server Gateway Interface (WSGI) subsystems, which come from Wekzeug; the template engine from the Jinja2 package; and command line integration from the Click package (Grinberg, 2018:3; Mufid et al., 2019). It provides developers with the libraries for handling web development tasks and allows them to integrate the extension based on the project requirements.

The development process utilised Flask version 3.0.3, Pandas version 2.2.3, Python version 3.9.13, and Bootstrap version 5.1.3. Cascading Style Sheets (CSS) were used to style HTML pages. The developed prototype is named *AquaGrowForecast*. The website's main menu (Navigation bar) and its objectives are summarised in Table 6.1.

Table 6.1: Website's main menu

Website Menu	Objective
Home Page	Explanation about the project.
Aquaponics	Introduction to aquaponics.
Prediction	Allows users to select options and make predictions, navigating to respective pages based on their choices.
Feedback	Enables users to provide feedback by completing a survey.

6.3.2 Layout of webpages

Webpage layout design is shown in figures 6.3 – 6.9.

a. A brief introduction about the project is given on the homepage. The layout is shown in Figure 6.3.

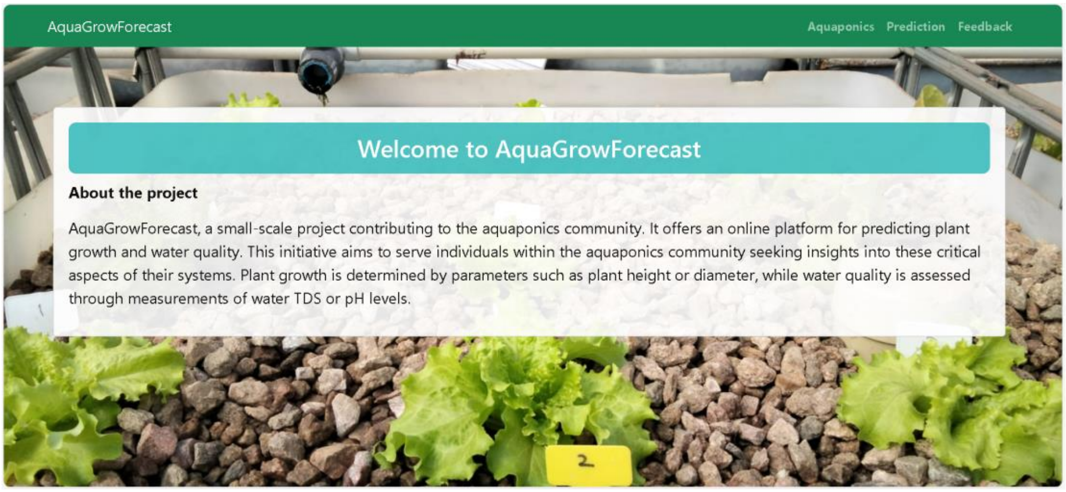


Figure 6.3: The home page layout

b. The aquaponics webpage provides a brief description of aquaponics to the user. The web page layout is shown in Figure 6.4.

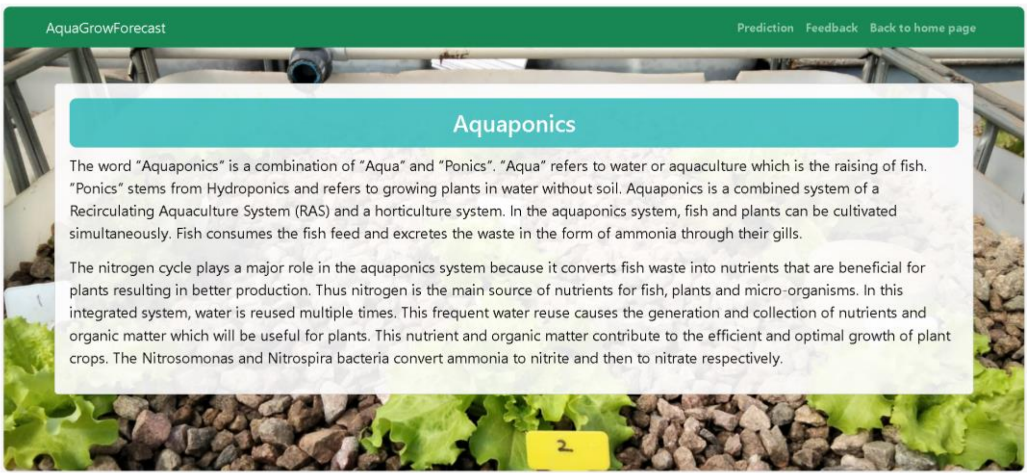


Figure 6.4: Aquaponics web page layout

c. In the prediction page, the user can select an option to proceed with the desired prediction. The page layout is shown in Figure 6.5.

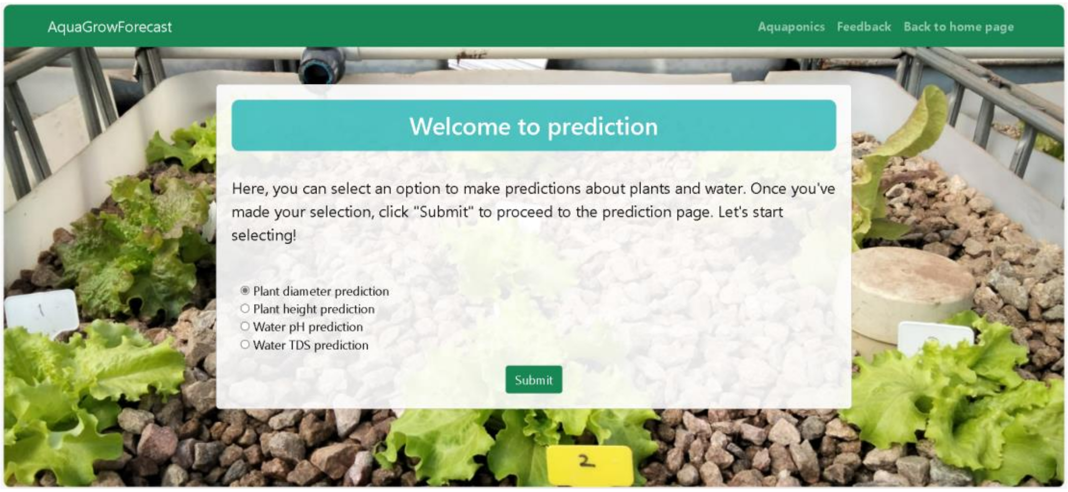


Figure 6.5: The prediction page layout

d. When a user selects the "Plant diameter prediction" option, the system navigates the user to the appropriate page where the user can input the required details for the output. The diameter prediction page layout is shown in Figure 6.6.

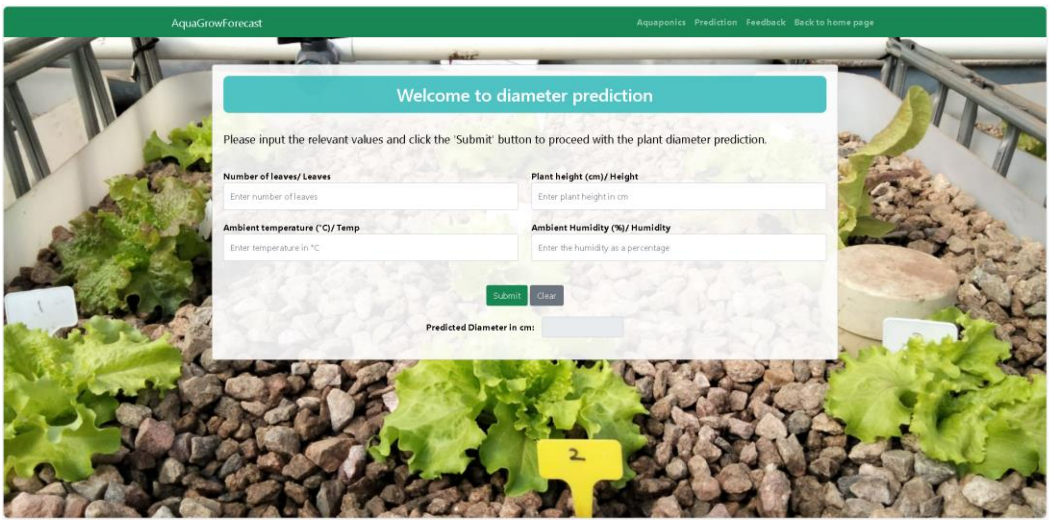


Figure 6.6: The diameter prediction page layout

e. When a user selects the "Plant height prediction" option, the system navigates the user to the height prediction page, where the user can input the required details for the output. The height prediction page layout is shown in Figure 6.7.

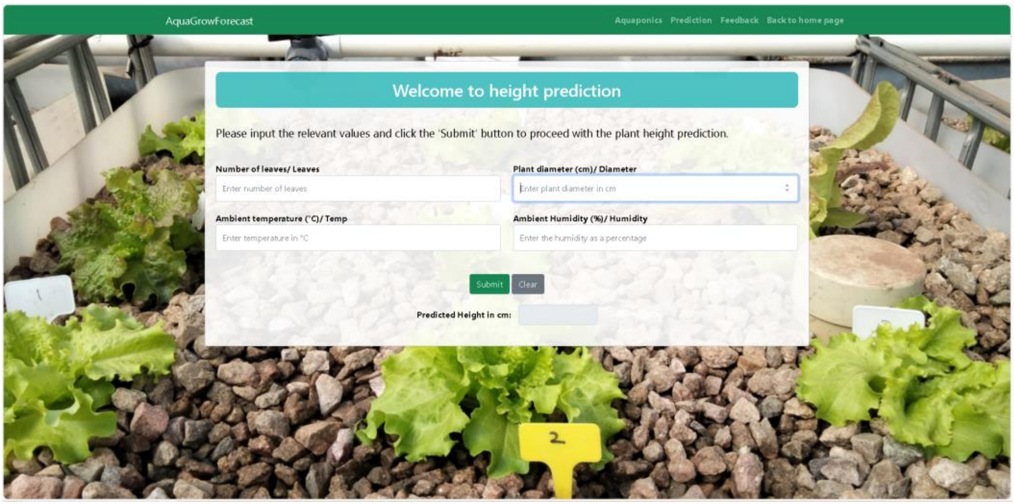


Figure 6.7: The height prediction page layout

f. When a user selects the "Water pH" option, the system will navigate to the pH prediction page, where the user can input the required details to generate the output. The water pH prediction page layout is shown in Figure 6.8.

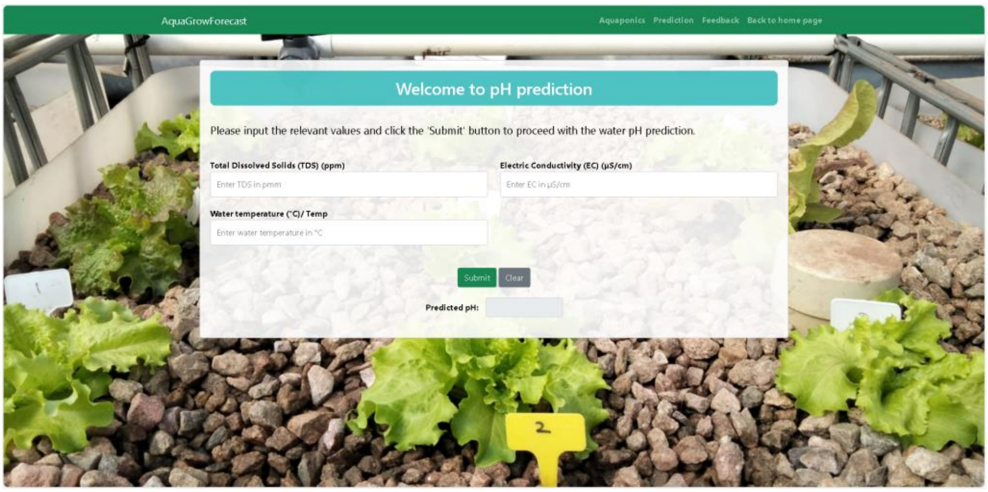


Figure 6.8: The water pH prediction page layout

g. When a user selects the "Water TDS prediction" option, the system navigates the user to the TDS prediction page, where the user can input the required details for the output. The TDS prediction page layout is shown in Figure 6.9.

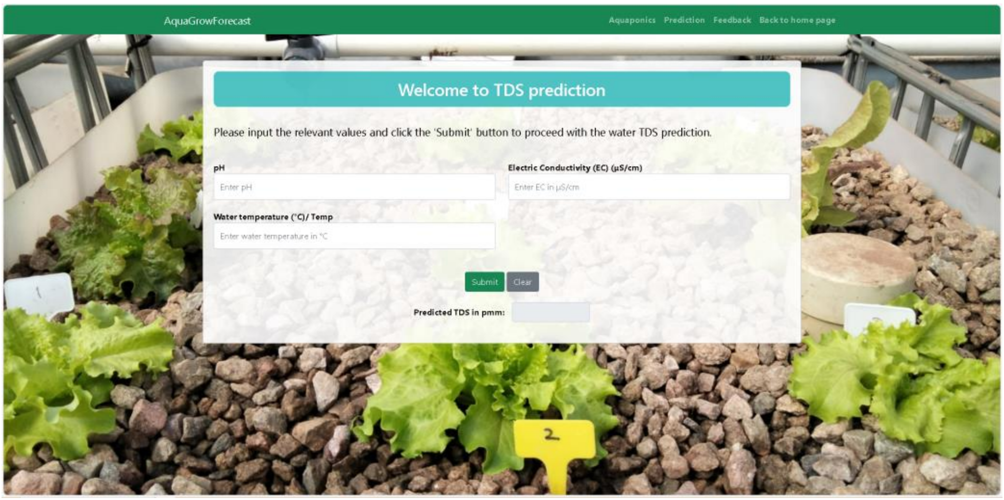
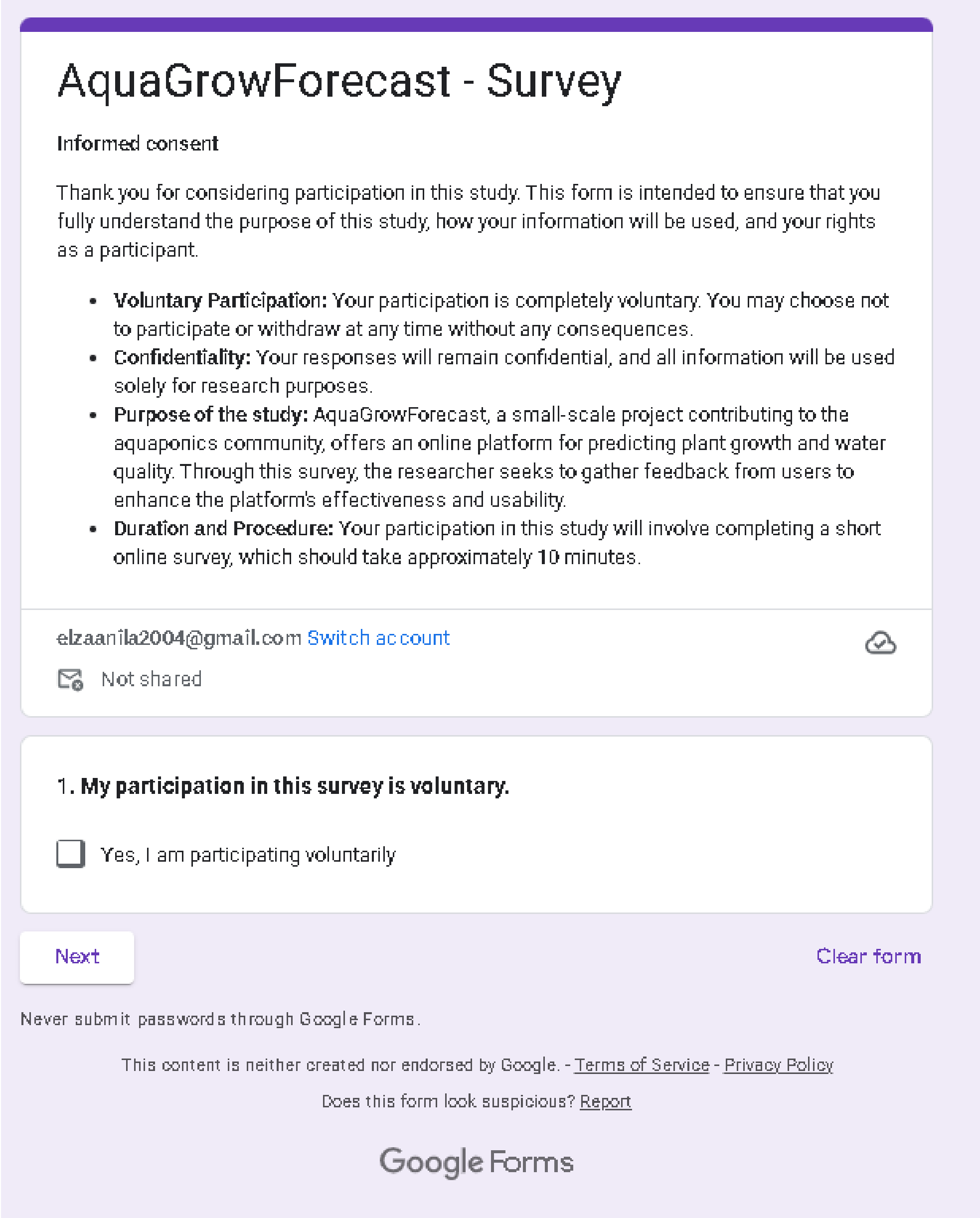


Figure 6.9: The TDS prediction page layout

The participants were presented with the opportunity to use the system, after which they provided feedback on its usability.

The feedback page consists of three sections:

1. **Informed Consent** – This section provides participants with the purpose of the feedback process and seeks their voluntary agreement to participate. Figure 6.10 below provides a structure of informed consent.



AquaGrowForecast - Survey

Informed consent

Thank you for considering participation in this study. This form is intended to ensure that you fully understand the purpose of this study, how your information will be used, and your rights as a participant.

- **Voluntary Participation:** Your participation is completely voluntary. You may choose not to participate or withdraw at any time without any consequences.
- **Confidentiality:** Your responses will remain confidential, and all information will be used solely for research purposes.
- **Purpose of the study:** AquaGrowForecast, a small-scale project contributing to the aquaponics community, offers an online platform for predicting plant growth and water quality. Through this survey, the researcher seeks to gather feedback from users to enhance the platform's effectiveness and usability.
- **Duration and Procedure:** Your participation in this study will involve completing a short online survey, which should take approximately 10 minutes.

elzaanila2004@gmail.com Switch account

Not shared

1. **My participation in this survey is voluntary.**

Yes, I am participating voluntarily

Next

Clear form

Never submit passwords through Google Forms.

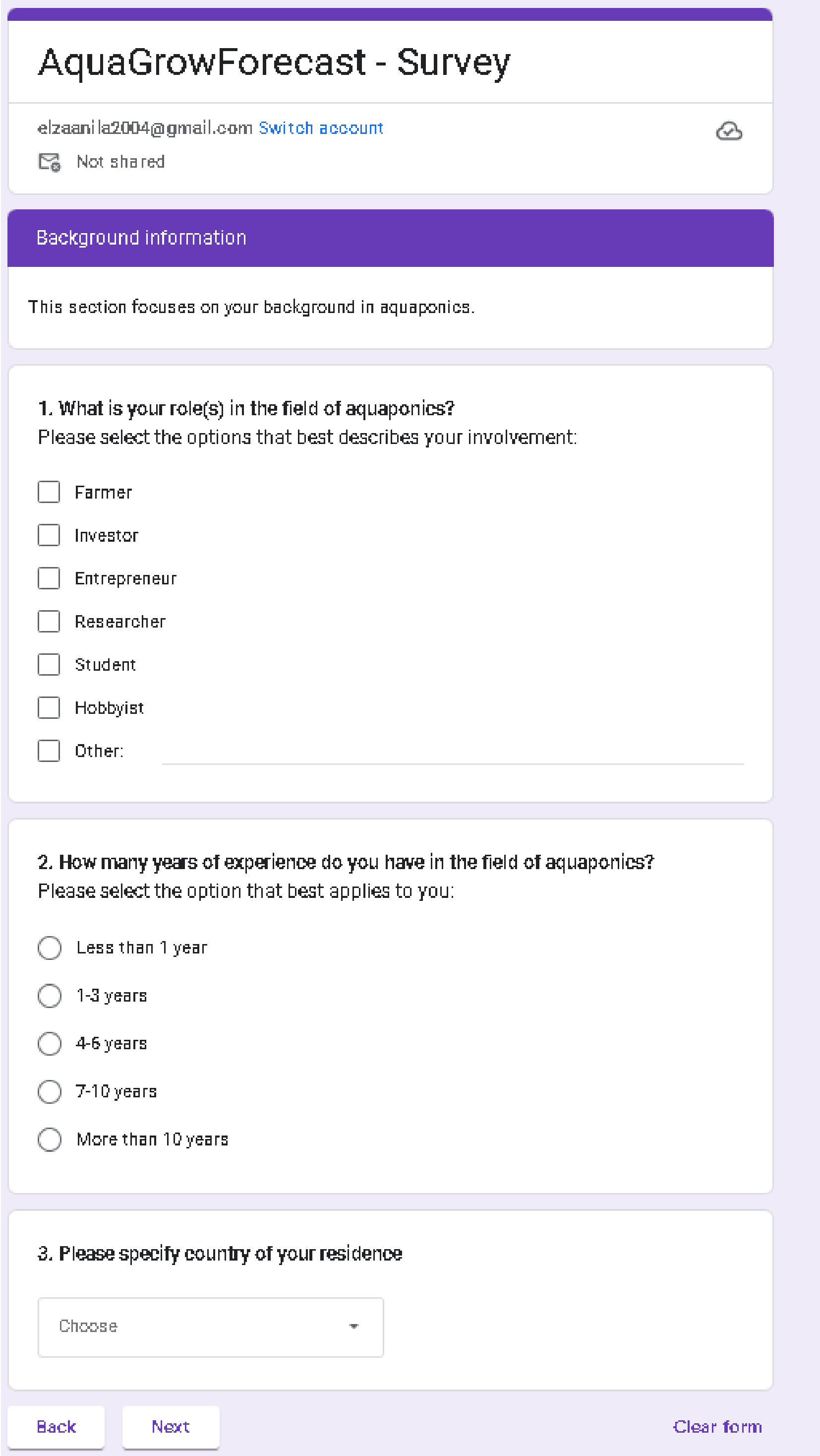
This content is neither created nor endorsed by Google. - Terms of Service - Privacy Policy

Does this form look suspicious? Report

Google Forms

Figure 6.10: Section 1- informed consent form

2. **Aquaponics Background** – This page was used to obtain information on the background, role and aquaponics experience level of the participant. The options are shown in Figure 6.11.



AquaGrowForecast - Survey

elzaanila2004@gmail.com Switch account

Not shared

Background information

This section focuses on your background in aquaponics.

1. What is your role(s) in the field of aquaponics?
Please select the options that best describes your involvement:

- [ ] Farmer
- [ ] Investor
- [ ] Entrepreneur
- [ ] Researcher
- [ ] Student
- [ ] Hobbyist
- [ ] Other: ____

2. How many years of experience do you have in the field of aquaponics?
Please select the option that best applies to you:

- ○ Less than 1 year
- ○ 1-3 years
- ○ 4-6 years
- ○ 7-10 years
- ○ More than 10 years

3. Please specify country of your residence

Choose ▾

Back　Next　Clear form

Figure 6.11: Section 2 - Aquaponics background form

6.3.3 Deploying Flask Apps: PythonAnywhere

PythonAnywhere is a cloud-based online Integrated Development Environment (IDE) (https://www.pythonanywhere.com/) based on the Python programming language (Visvizi et al., 2020; Suryawanshi, 2021; Sarala et al., 2021). PythonAnywhere was founded by Giles Thomas and Robert Smithson in 2012 (Suryawanshi, 2021). It provides web hosting services, which fall under the platform as a service (PaaS) model (Suryawanshi, 2021). PaaS is a service that enables web developers to host their websites on a platform that is managed and controlled by a third party (Stouffer, 2015:248). In the PythonAnywhere environment, users can deploy the Flask framework or Django framework applications, and it also allows users to write, edit and run the code directly (Visvizi et al., 2020; Suryawanshi, 2021). The developed Flask application is deployed using PythonAnywhere.

The PythonAnywhere link for predicting plant height, plant diameter, water pH and water TDS is available on annjiby.pythonanywhere.com. The link to the feedback survey is: *AquaGrowForecast* - Survey. Figures 6.12 and 6.13 shows typical instances (screenshots) of the system under use.

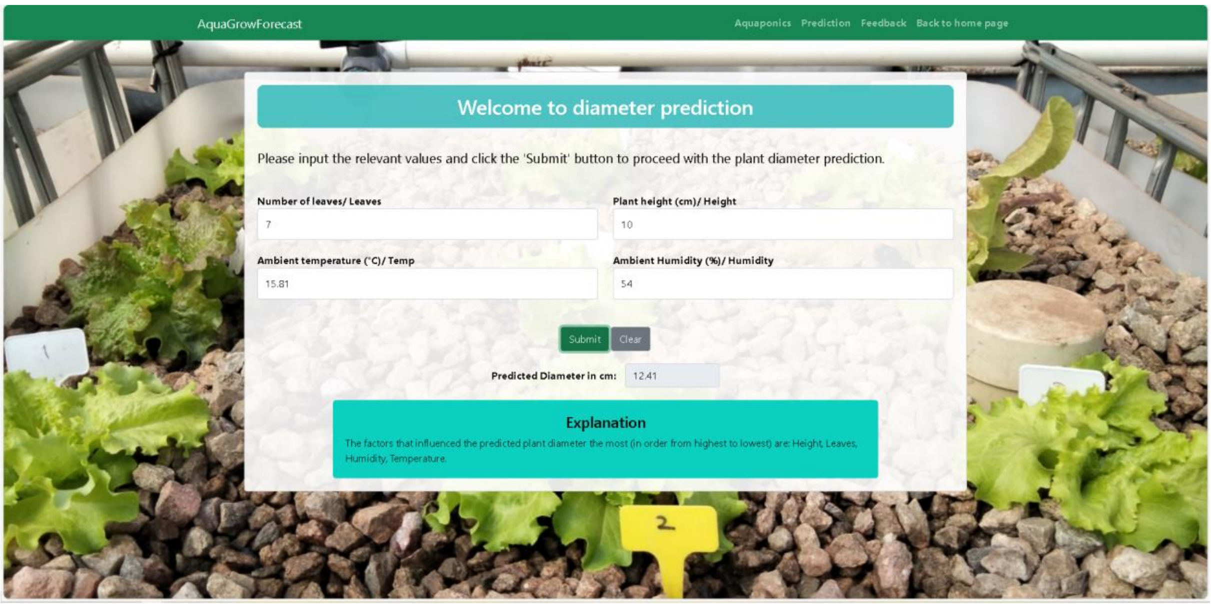


Figure 6.12: Screenshot of plant diameter prediction

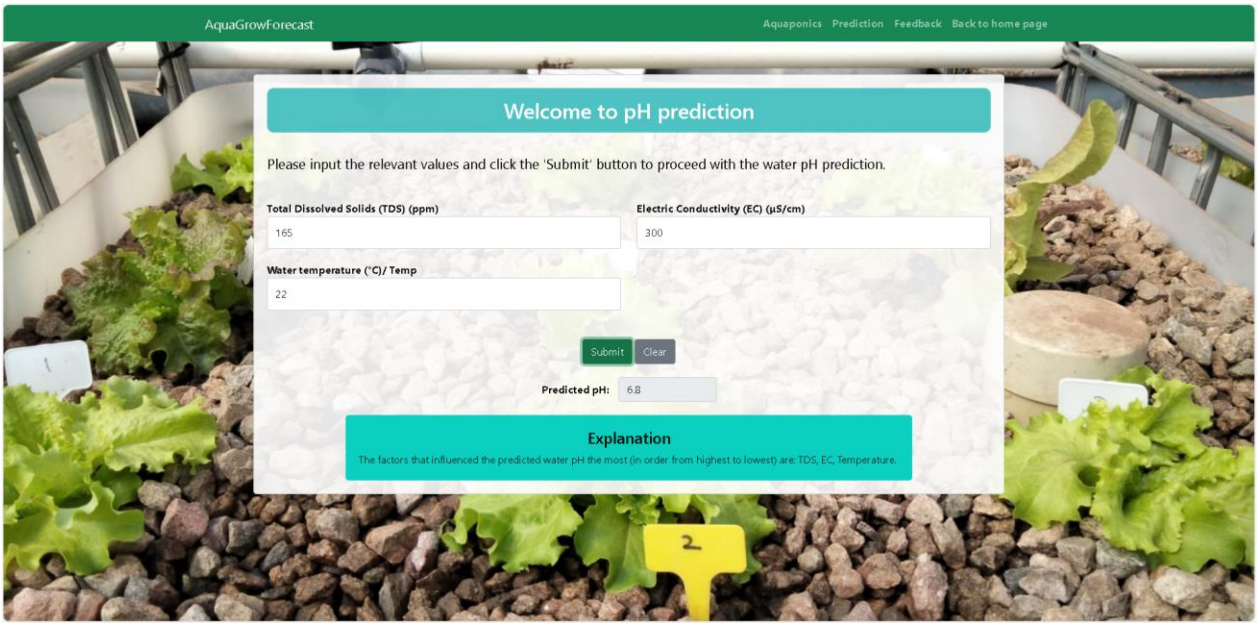


Figure 6.13: Screenshot of pH prediction

6.4 Usability evaluation using SUS

This section describes the procedure that was used to assess the usability of the developed DSS by using the System Usability Scale (SUS) questionnaire. The SUS is a survey instrument to measure the usability of the variability of products and services, including websites, which was developed by Brooke in 1986 (Kortum & Bangor, 2013; Setemen et al., 2019). The SUS is a five-point Likert scale consisting of 10 questions or survey items that users of the website will respond to (Setemen et al., 2019; Kortum & Bangor, 2013). The usability measurement assesses how well users can interact with the developed system. According to ISO 9241-11, usability measures should cover effectiveness, efficiency and satisfaction (Brooke, 1996).

i. *Effectiveness* measures the ability of users to complete tasks using the system and the quality of the output of the performed tasks (Brooke, 1996; Kortum & Bangor, 2013).
ii. *Efficiency* measures the resources consumed by the user to perform the tasks (Brooke, 1996; Kortum & Bangor, 2013).
iii. *Satisfaction* measures a user's assessment based on how well the developed system met his or her needs (Brooke, 1996; Kortum & Bangor, 2013).

It has been a trusted and reliable tool for assessing system usability, due to the speed and cost-effectiveness of implementation. A SUS template already exists, which is a tried and tested template. A basic tweaking of the SUS template would ensure that it is effective in the context of a particular study.

The participants rated each question on a scale from 1 to 5, with 1 indicating strong disagreement and 5 indicating strong agreement with the statement (Brooke, 1996).

The basic approach adopted is to let the users experience and work on the website. Later, they were requested to complete the survey.

The updated SUS questions/ items are presented below in Table 6.2.

Table 6.2: Updated SUS questions/items

No.	SUS items	Strongly disagree				Strongly agree
1.	I think that I would like to use *AquaGrowForecast system* frequently.	1	2	3	4	5
2.	I found *AquaGrowForecast system* unnecessarily complex.	1	2	3	4	5
3.	I thought *AquaGrowForecast system* was easy to use.	1	2	3	4	5
4.	I think that I would need the support of a technical person to be able to use *AquaGrowForecast system*.	1	2	3	4	5
5.	I found the various functions in *AquaGrowForecast system* were well integrated.	1	2	3	4	5
6.	I thought there was too much inconsistency in *AquaGrowForecast system*.	1	2	3	4	5
7.	I would imagine that most people would learn to use *AquaGrowForecast system* very quickly.	1	2	3	4	5
8.	I found *AquaGrowForecast system* very cumbersome to use.	1	2	3	4	5
9.	I felt very confident using *AquaGrowForecast system*.	1	2	3	4	5
10.	I needed to learn a lot of things before I could get going with *AquaGrowForecast system*.	1	2	3	4	5

6.5 Criteria for selecting participants to evaluate the developed system

Aquaponics practitioners, including researchers and members of the aquaponics community, were selected to evaluate the developed system through a survey that involved answering SUS (System Usability Scale) questions/ items. Requests were sent via email, WhatsApp, and Facebook. In total, 127 requests were sent. Table 6.3 shows the number of requests sent through each platform.

Table 6.3: Request-sent platforms and population

Request-sent platforms	**Population**
Email	90
Facebook	22
WhatsApp	15
Total	**127**

A total of 16 responses were received. However, one participant did not score an item, and another did not specify their role in the aquaponics field. Hence, these two responses were eliminated and the remaining 14 respondents who answered all three sections were used for the evaluation.

6.6 Evaluation results

This segment provides a summary of the surveyed information based on the 3 sections addressed in the feedback page.

Section 1:

All 14 respondents had voluntarily participated in the feedback survey.

Section 2:

This section provides a background summary of the participants in the aquaponics field. Table 6.4 depicts the participants' aquaponics background summary based on their roles, years of experience, and country of residence.

Table 6.4: Aquaponics background summary

Questions	Aquaponics background with the number of participants in parentheses)
What is your role(s) in the field of aquaponics?	Researcher (7), Hobbyist (4), Student (2) and Farmer & researcher (1)
How many years of experience do you have in the field of aquaponics?	1-3 years (7), Less than 1 year (5), 7-10 years (1) and 4-6 (1)
Country of residence	South Africa (5), India (4), United Kingdom (1), Germany (1), Zimbabwe (1), Australia (1) and Philippines (1)

Section 3:

This part summarises each item based on the scale provided by the participants. The participants answered the SUS items using a scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), exhibiting their level of agreement or disagreement with each statement. This scale helps to assess the usability of the developed system based on participants' views. The odd-numbered items have positive meanings, while the even-numbered items have negative meanings.

Item 1: I think that I would like to use the *AquaGrowForecast* system frequently.

This is to establish the practical assistance that this developed system would render to the participant.

Summary: The majority, 12 out of 14 (scale 4 and 5), participants provided positive ratings, where they would use the developed system frequently. Question 1 summary is depicted in Figure 6.14.

Figure 6.14: SUS item 1 responses

Item 2: **I found the *AquaGrowForecast* system unnecessarily complex.**

This question is to determine if the *AquaGrowForecast* system is unnecessarily complicated to use.

Summary: The majority of participants, 11 out of 14 (scale 1 and 2), confirmed that they don't find the system unnecessarily complex. Item 2 summary is depicted in Figure 6.15.

Figure 6.15: SUS item 2 responses

Item 3: I thought the *AquaGrowForecast* system was easy to use.

This question is to ensure consistency with the above question and validate if the system was easy to use.

Summary: A total of 11 out of 14 participants (scale 4 and 5) reported back that the system was easy to use. Item 3 summary is depicted in Figure 6.16.

Figure 6.16: SUS item 3 responses

Item 4: I think that I would need the support of a technical person to be able to use the *AquaGrowForecast* system.

This tries to ascertain if the system is straightforward and does not require any technical knowledge, etc., to use the system.

Summary: With 8 out of 14 (scale 1 and 2), there is a balanced overview of the participants expressing the need for a technical person to assist with the developed system. Item 4 summary is depicted in Figure 6.17.

Figure 6.17: SUS item 4 responses

Item 5: I found the various functions in the *AquaGrowForecast* system were well integrated. This ensures that the participant finds the system seamless and continuous.

Summary: 11 out of 14 feedback points (scale 4 and 5) expressed confidence that the system is well integrated. Item 5 summary is depicted in Figure 6.18.

Figure 6.18: SUS item 5 responses

Item 6: I thought there was too much inconsistency in the *AquaGrowForecast* system.

This is a follow-up question to the previous question to ensure validation of the above answers and establish more certainty in the feedback.

Summary: 10 out of 14 responses (scale 1 and 2) show that the participants did not feel there was an inconsistency in the system. The Item 6 summary is depicted in Figure 6.19.

Figure 6.19: SUS item 6 responses

Item 7: I would imagine that most people would learn to use the *AquaGrowForecast* system very quickly.

This is to establish if the participant sees it as a potentially easy system to use for the general public and other users.

Summary: 11 out of 14 responses (scale 4 and 5) showed confidence in people being able to learn to use the system easily. Item 7 summary is depicted in Figure 6.20.

Figure 6.20: SUS item 7 responses

Item 8: I found the *AquaGrowForecast* system very cumbersome to use.

This is to validate the above questions as well as get feedback on whether the system had any unnecessary complications. A need to streamline the system more, if required, is established.

Summary: 10 out of 14 responses (scale 1 and 2) showed that they don’t believe the system to be cumbersome to use. The item 8 summary is depicted in Figure 6.21.

Figure 6.21: SUS item 8 responses

Item 9: I felt very confident using the *AquaGrowForecast* system. This is to get an idea of whether the participant was comfortable and reassured of the system's operation and their use of it.

Summary: With 12 out of 14 responses (scale 4 and 5), it is clear that the majority of the respondents are confident in using the *AquaGrowForecast* system. The item 9 summary is depicted in Figure 6.22.

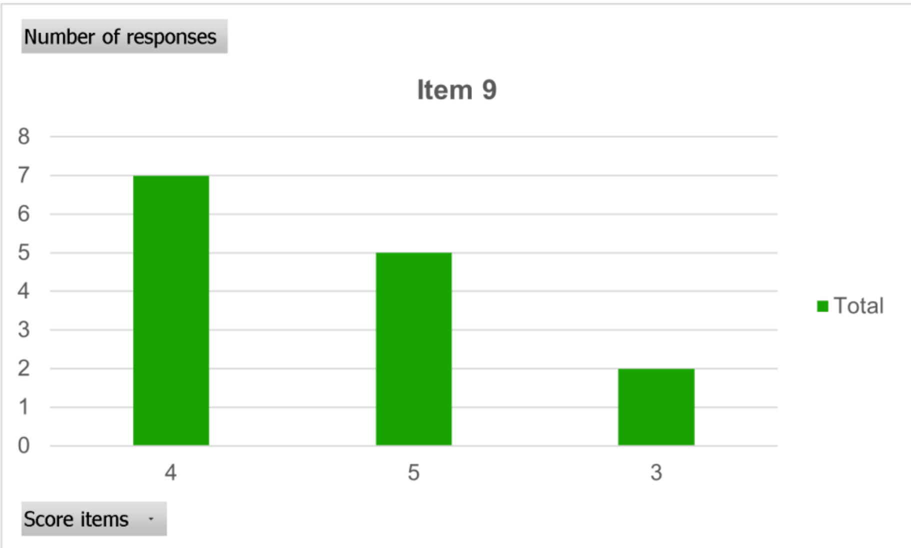


Figure 6.22: SUS item 9 responses

Item 10: **I needed to learn a lot of things before I could get going with the *AquaGrowForecast* system.**

This is to establish if the participant needed a lot of preparation or background knowledge before using the developed system.

Summary: 5 out of 14 participants (scale 4 and 5) need to learn a lot of things before using the developed *AquaGrowForecast* system. However, half (7) of the participants (item points 1 and 2) suggested they don't need to learn a lot of things to get going with the *AquaGrowForecast* system. Item 10 summary is depicted in Figure 6.23.

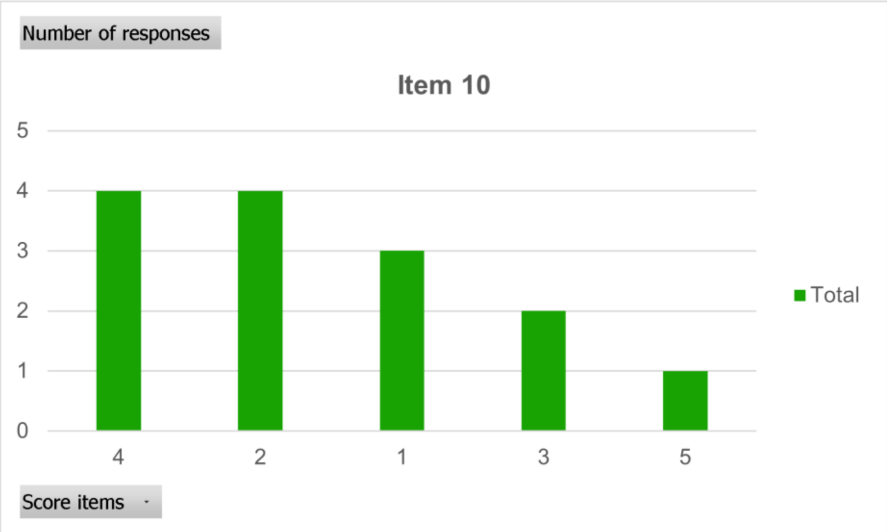


Figure 6.23: SUS item 10 responses

The received individual scores are meaningless in isolation. Hence, an SUS Score needs to be calculated to measure the overall usability of the system. The individual scores are shown in Table 6.5.

Table 6.5: Individual scores

Participants	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10
1.	4	3	3	4	3	3	5	2	4	4
2.	4	4	4	3	4	2	3	2	3	4
3.	2	2	3	3	5	2	4	3	4	3
4.	4	4	4	4	4	4	5	4	4	5
5.	4	1	5	4	5	2	4	2	4	4
6.	1	1	3	1	2	3	4	3	4	1
7.	4	2	4	2	4	2	4	2	3	2
8.	4	2	4	2	3	1	4	1	4	3
9.	4	1	4	2	5	1	3	1	5	1
10.	4	2	4	2	5	2	4	2	5	4
11.	5	1	5	1	5	1	5	1	5	1
12.	4	2	5	2	4	1	4	2	4	2
13.	4	1	4	2	5	3	4	3	5	2
14.	5	1	5	3	5	1	3	1	5	2

6.6.1 Calculating average system usability scores

The following method was applied to calculate the SUS score. The item score contributions from each question. Each item's/question's score contribution range is from 0 to 4. For odd-number items 1, 3, 5, 7, and 9, the score contribution is the scale position minus 1 (x-1). For even-number items 2, 4, 6, 8, and 10, the contribution is subtracting the scale position from 5 (5-x). To obtain the System Usability score, multiply the sum of the item scores by 2.5. This ranges from 0 (extremely poor usability) to 100 (excellent usability) (Brooke, 1996).

Total score = Sum of ((Score of each odd-numbered item − 1) + (5 − Score of each even-numbered item))

The calculated System Usability score is shown in Table 6.6.

Table 6.6: System Usability score

Participants	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	Total score	*2.5 (System Usability score)
1.	4	3	3	4	3	3	5	2	4	4	23	57.5
2.	4	4	4	3	4	2	3	2	3	4	23	57.5
3.	2	2	3	3	5	2	4	3	4	3	25	62.5
4.	4	4	4	4	4	4	5	4	4	5	20	50
5.	4	1	5	4	5	2	4	2	4	4	29	72.5
6.	1	1	3	1	2	3	4	3	4	1	25	62.5
7.	4	2	4	2	4	2	4	2	3	2	29	72.5
8.	4	2	4	2	3	1	4	1	4	3	30	75
9.	4	1	4	2	5	1	3	1	5	1	35	87.5
10.	4	2	4	2	5	2	4	2	5	4	30	75
11.	5	1	5	1	5	1	5	1	5	1	40	100
12.	4	2	5	2	4	1	4	2	4	2	32	80
13.	4	1	4	2	5	3	4	3	5	2	31	77.5
14.	5	1	5	3	5	1	3	1	5	2	35	87.5

The system Usability score, along with each participant's background, is shown in Table 6.7.

Table 6.7: Individual's System Usability score

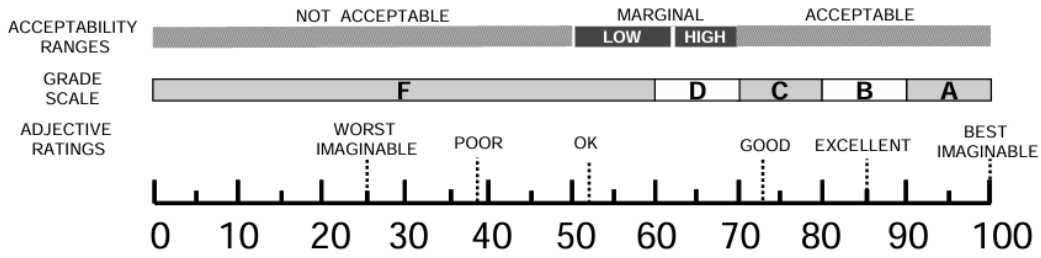
Participants	Role	Years of experience	Country	SU score
1.	Hobbyist	Less than 1 year	South Africa	57.5
2.	Farmer; Researcher	1-3 years	South Africa	57.5
3.	Researcher	Less than 1 year	India	62.5
4.	Researcher	Less than 1 year	Philippines	50
5.	Researcher	1-3 years	South Africa	72.5
6.	Researcher	7-10 years	Australia	62.5
7.	Researcher	1-3 years	South Africa	72.5
8.	Hobbyist	Less than 1 year	South Africa	75
9.	Researcher	1-3 years	India	87.5
10.	Student	1-3 years	India	75
11.	Hobbyist	4-6 years	United Kingdom	100
12.	Student	1-3 years	Germany	80
13.	Hobbyist	1-3 years	India	77.5
14.	Researcher	Less than 1 year	Zimbabwe	87.5

In this survey, individual participants' SUS scores ranged from 50 to 100, where 50 represented marginally acceptable usability and 100 indicated excellent usability. The highest score of 100 was achieved by a hobbyist with 4–6 years of experience, while the lowest score of 50 was reported by a researcher with less than 1 year of experience. When analysing the data based on each role:

i. **Researchers:** The usability scores varied widely, ranging from 50 to 87.5. This group demonstrated the broadest experience levels, spanning from less than 1 year to over 10 years in the aquaponics field. The variation in scores suggests that researchers' perception of usability may be influenced by their extensive and diverse expertise in the domain.
ii. **Hobbyists:** This group exhibited scores ranging from 57.5 to 100, with experience levels between less than 1 year and 6 years. The highest score of 100 was recorded in this category, indicating that hobbyists with moderate experience may find the system particularly intuitive and user-friendly.
iii. **Students:** Scores for students were relatively consistent, falling between 75 and 80. All participants in this group had 1–3 years of experience, suggesting a more uniform perception of system usability compared to the other groups.
iv. **Farmer and researcher:** With 1–3 years of experience, a score of 57.5 was recorded, classified as moderately acceptable. This suggests that the system's usability meets a basic standard but may require enhancements to better align with the needs and expectations of users in this category.

Finally, the converted mean score of the SUS is placed into the following categories: *acceptance level*, *grading scale*, and *adjective rating*. Bangor et al. (2009) developed the categories for the SUS scores. The Acceptability range is categorised into three sections: "*Not*

Acceptable”, “*Marginal*”, and “*Acceptable*”. The letter grade scale is classified as ‘A’, ‘B’, ‘C’, ‘D’, and ‘F’. This is an alternate way to understand the absolute meaning of an SUS score. The adjective ratings are split into seven: “*Worst Imaginable*”, “*Poor*”, “*OK*”, “*Good*”, “*Excellent*” and “*Best Imaginable*”. These provide a subjective label for an individual study’s mean SUS score (Bangor et al., 2009; Setemen et al., 2019). A System Usability Scale is shown in Figure 6.24.



.

Figure 6.24: System Usability Scale (Bangor et al., 2009)

The analysis emphasises system usability perceptions, influenced by the roles and experience levels of participants within the aquaponics field. The mean SUS score of the developed system after evaluation was 72.68, indicating that the system is acceptable in terms of the system usability scale shown in Figure 6.22. As per the Grade Scale, it falls under a rating of ‘C’ category, and the adjective rating is “Good”. The implication is that the system has good usability. This also gives the assurance that users find the system easy to use.

6.7 Chapter summary

This chapter explained how the decision support system was developed and evaluated. The results section presents the participants' responses and feedback, along with an overview of their basic background. The insights from the results highlight the usability of the developed system as acceptable.

CHAPTER SEVEN
CONCLUSION AND RECOMMENDATIONS

This chapter provides a summary of this study's research objectives. It also discusses the contributions of the study, the study's limitations, and concludes with recommendations and suggestions for future research.

7.1 Research summary

The study aimed to develop a decision support system capable of predicting plant growth, specifically in terms of plant diameter and height for the hydroponics component, as well as predicting the water quality for the aquaculture component of aquaponics. To achieve this, four research objectives were formulated as stated in Section 1.4.2. The thesis chapters were structured to address these research objectives.

Chapter One introduces the study's motivation, background, aim, objectives, and research questions, highlighting its significance and scope. This chapter is concluded with an overview of the thesis structure.

Chapter Two provides a theoretical overview, covering key concepts such as hydroponics, aquaculture, aquaponics, machine learning, decision-making, decision support systems, the Internet of Things (IoT), intelligent IoT, expert systems, and explainable AI. Additionally, the chapter identifies the gaps in existing studies and highlights areas that require further research exploration.

Chapter Three discusses the research philosophy, approach, methodological choices, and strategy. Experimental research design was outlined, as illustrated in Figure 3.1. The chapter also describes the data collection and analysis methods. Finally, the ethical considerations that guided the study were explained.

Chapter Four explains how aquaponics was set up for plant, water, and environment data collection and how data was collected using various methods, such as manual and IoT technologies. Thereafter, how the collected data was stored for experimentation is further explained.

Chapter Five presents the machine learning experimentation for aquaponics and the evaluation process. Experiments were performed using selected supervised machine learning algorithms, and the models' performances were evaluated using regression metrics. Finally, the most influential features for the aquaponics predictions were identified using SHAP — SHapley Additive exPlanations.

Chapter Six presents the development, deployment, and evaluation of the decision support system for aquaponics prediction. The usability and effectiveness of the developed system were assessed by using the System Usability Scale (SUS).

Chapter Seven provides a summary of the research objectives, contributions, and recommendations for future research.

The following explains how the related activities were carried out and how they contributed to achieving the study's objectives.

Objective 1: *To identify the key parameters used to measure plant growth and the monitored water quality parameters in aquaponics systems.*

A detailed review of the existing literature was conducted to identify the parameters commonly considered for assessing plant growth and water quality in aquaponics systems. The review found the parameters used to estimate plant growth to be ambient temperature, light intensity, plant height, stem diameter, and leaf area. The determination of water quality in aquaculture was not restricted to one parameter. The review showed that several parameters, such as pH, temperature, total dissolved solids (TDS), electrical conductivity (EC), ammonia, and dissolved oxygen, are used to assess the water quality. However, most studies emphasise pH and water temperature due to the significant impact they have on the maintenance of water quality and supporting fish growth in aquaponics systems.

Based on these findings, this study selected the following parameters, plant height, leaf count, plant diameter, ambient temperature and ambient humidity to determine the plant growth. For water quality analysis, pH, water temperature, TDS, and EC parameters were selected.

Objective 2: *To develop a prediction model that can be used to determine the optimal level of aquaponics systems by using machine learning (ML).*

This study explored four possibilities to predict plant growth and water quality in a tunnel-based media aquaponics system. By either estimating plant height or plant diameter, plant growth was established whereas water quality was determined by estimating either pH or TDS. For plant height estimation in plant growth, the following features, plant diameter, number of leaves, ambient temperature, and ambient humidity were considered. Features such as plant height, number of leaves, ambient temperature, and ambient humidity were considered for plant diameter estimation. For water quality pH estimation, the considered parameters were TDS, water temperature, and EC. Meanwhile, for water TDS estimation in water quality, the following features, pH, water temperature, and EC, were considered. Thereafter, five supervised machine learning algorithms, namely, Linear Regression (LR), Random Forest (RF), Multilayer Perceptron (MLP), eXtreme Gradient Boosting (XGBoost), and k-Nearest

Neighbor (KNN), were developed to realise aquaponics prediction covering the aspects of plant growth and water quality.

Objective 3: *To determine the performance of the different ML algorithms based on regression metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R-squared and Adjusted R-Squared.*

The developed ML models were evaluated and compared using various regression metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R-squared, and Adjusted R-squared. When evaluation was done, the conclusion derived was that the Random Forest algorithm outperformed others in predicting plant height and plant diameter. In the prediction of pH, eXtreme Gradient Boosting (XGBoost) performed the best. Linear Regression was found to be the most effective for predicting TDS. These results were based on the criteria to minimise errors and produce the most precise predictions for each parameter.

Objective 4. *To design and develop an ML-based decision support system for aquaponics to support decision-making.*

Post evaluation, the models with the best performance were selected to design and develop the decision support system (DSS) for aquaponics prediction using the Flask framework. After the user selects the various prediction options, the developed DSS could predict plant growth or water quality. The system then provided insights to users by ranking parameters based on their influence in the selected prediction, from high to low, using SHapley Additive exPlanations (SHAP) values.

Objective 5. *To assess the usability of a decision support system for aquaponics prediction from the perspective of aquaponics stakeholders.*

To evaluate the effectiveness of the developed DSS, a usability feedback survey was conducted by using the System Usability Scale (SUS). After the evaluation, the developed DSS obtained an overall score of 72.68%. This indicated the favourable acceptance of the system and highlighted the system's usability and potential value for aquaponics stakeholders and practitioners.

7.2 Contributions of the study

This study has made theoretical, methodological and practical contributions, which are discussed below.

7.2.1 Theoretical contribution

This study offers a theoretical contribution by providing a better understanding of how machine learning models can be applied to predict plant growth and water quality in the context of an aquaponics system. Another theoretical contribution is the implementation of explainable AI (XAI) in aquaponics farming to identify the parameters which have the highest influence on plant growth, and also to predict water quality based on data collected under South African weather conditions. Furthermore, the study advances theoretical knowledge in decision support systems by illustrating how machine learning predictions can enhance decision-making processes and improve operational outcomes in aquaponics. The incorporation of the System Usability Scale (SUS) evaluation method further contributes to the understanding of users' system acceptance.

7.2.2 Methodological contribution

This study investigated the application of intelligent Internet of Things, which includes the integration of IoT, machine learning, and AI (decision support system) for aquaponics prediction. Most aquaponics studies have applied machine learning for various aquaponics predictions, or IoT for data collection, or both. However, the integration of the Intelligent Internet of Things that combines ML, IoT, and AI has not been extensively explored. Also, this study addressed the problem of the lack of explainability of aquaponics solutions through the application of explainable AI.

7.2.3 Practical contribution

The practical contribution of this study lies in developing a decision support system (DSS) for aquaponics. The developed DSS will assist practitioners in making data-driven decisions, improving efficiency, and ensuring sustainability in aquaponics operations. The system will provide valuable insight based on the sizable data previously captured and processed by the trained ML models. Inexperienced newcomers and aquaponics hobbyists can be guided on the plant growth and water quality parameters expected at various intervals of the aquaponics cycle. Users of relatively bigger aquaponics setups can be assisted in validating their collected data against the values presented by this DSS for aquaponics prediction. The Aquaponics web application is easy to use, as per the SUS Survey conducted, and thus, data is presented in a readable and concise manner to the stakeholders. The most crucial parameters are provided so that the user can be aware of the parameters that have the most impact on overall plant growth and water quality.

7.3 Limitations of the study

The limitations that were observed during and after the study are outlined below:

i. **Limited parameters:** A limited number of parameters for prediction, which may not fully capture the complexity of aquaponics systems. The investigation did not incorporate elements such as extended ambient environmental conditions and more detailed water and nitrogen quality cycle parameters. Incorporating an extensive set of parameters would increase the scale of the study tremendously, thus, the most critical and impactful parameters were selected and used.

ii. **Different aquaponics settings**: The results could not be applied to different system configurations because the aquaponics environment was set in a grow-bed system. Other aquaponics setup options, such as floating raft systems, Nutrient Film Technique (NFT) or Deep Water Culture (DWC), were not used in this study.

iii. **Study duration**: Due to the short time frame during which the data was gathered, it might not accurately represent seasonal or long-term trends. An extended period of observation would provide more insights into the system's performance over time and enable a more accurate depiction of seasonal changes.

iv. **Restricted number of IoT devices:** The study's reliance on a small number of IoT devices might have limited the extent of system monitoring and data collection. Better system performance monitoring and analysis may be possible with more complete data from a larger range of IoT devices.

v. **Deployment:** XGBoost was initially selected for water pH prediction because it had the best performance. However, it could not be deployed on the PythonAnywhere platform due to library incompatibility. Therefore, random forest, which had a comparatively good performance and was compatible with the PythonAnywhere platform, was deployed.

vi. **Low farmer engagement:** Furthermore, the study observed a lower number of farmers' participation in the DSS evaluation, which might have restricted the amount of feedback gathered. It needs to be considered that aquaponics is still growing, hence, there are not many aquaponics farmers and practitioners in South Africa yet. Time constraints and insufficient use of technology for existing farmers are possibly other reasons.

7.4 Recommendations

Given the severe poverty and unemployment situation in South Africa, farming solutions such as aquaponics can be further investigated and possibly supported by the national or local government.

Aquaponics has the advantage of using the least amount of land or area while providing maximum fish and crop yield, which is highly beneficial twofold. Urban areas have limited space/land availability, whereas rural areas struggle with very limited water resources. Both these constraints are addressed by aquaponics through minimal land and water usage requirements. The government can look at training, providing funding and market accessibility for these potential users/stakeholders. Workshops and training sessions can help upskill inexperienced stakeholders.

Aquaponics setups can be scaled up by farmers as per the needs of their customers or the clients they sell their crops to. Local shops and restaurants can be provided with cost-effective, organic, fresh food produce from nearby areas instead of complicated logistics and expensive transportation.

The adoption of the developed decision support system for aquaponics can guide the users and the aquaponics workforce. The government can provide a platform for work seekers and aspirants who would like to get into the aquaponics field. Many fields connected within aquaponics have many opportunities for learning and specialising in its various fields and gaining valuable experience in the process. The government can even look at setting up potential organic food hubs and markets for communities and residents to benefit from such food production methods.

The infrastructure, small enterprise development initiatives, and funding must have a proper framework and must be implemented consistently. This will encourage interest from investor communities. This, in turn, will also encourage and broaden the usage of cutting-edge technologies such as artificial intelligence (AI), machine learning, and the Internet of Things (IoT) in aquaponics. The most cost-effective and reliable systems will easily gain adoption, leading to optimal aquaponics farming production.

7.5 Future work

Further research opportunities in aquaponics are outlined below:

i. **Expanding plant selection, fish selection and system architecture in aquaponics:** This study used leafy lettuce in hydroponics and Mozambique Tilapia in aquaculture. The scope of the study can be expanded to a wider variety of plants and fish, including various hydroponics system setups in aquaponics (Hao et al., 2020; Naputol et al., 2024; Liu & Jiang, 2024; Channa et al., 2024). Larger modular or scalable aquaponics unit setups can be investigated and further explored. The advantage of a scalable model is that it can be initially offered in a cost-effective, small setup for local communities. The local community can derive economic benefit with just a small investment, subscription or rent-to-own model. Moreover, and importantly, since unemployment is a dire situation in South Africa, local jobs in the community can be created after initial training is provided on the aquaponics units' setup process and maintenance. When the demand for fresh, organic produce increases, the units can easily be scaled up to a larger unit that produces more output.

ii. **Long-term, seasonal and climate-inclusive aquaponic research:** Access to publicly available aquaponics data remains limited and challenging (Channa et al., 2024). The creation and use of an efficient dataset within the aquaponic environment is one of the primary requirements for the aquaponic study (Taji et al., 2023). Studies thus far have focused on a much shorter data collection timeline. More detailed and long-term data collections, incorporating seasonal and various climatic conditions, can be done (Liu & Jiang, 2024; Anila & Daramola, 2024).

iii. **Explainable AI methods in aquaponics:** Despite the increasing use of IoT and machine learning technologies in aquaponics, research has not been able to incorporate ML interpretability techniques adequately enough to explain how predictions are made. ML interpretation will support the aquaponics community in advanced decision-making (Ekanayake et al., 2022). An explainable AI (XAI) method, namely SHAP, was utilised in this study to identify the most influential features. Expanding with more XAI methods could provide a comprehensive comparison with other explainability techniques (Das & Rad, 2020; Ekanayake et al., 2022; Anila & Daramola, 2024).

iv. **Prediction model:** Five prediction models were used in the study, which allowed for the investigation of the most successful of these prediction models. Finding additional machine learning and deep learning algorithms, particularly for complex systems, could be advantageous and yield better results (Liu & Jiang, 2024). Particular attention should be given to predictive analytics using deep learning in aquaponics, along with comparative evaluation against other machine learning models (Lauguico et al., 2020; Taji et al., 2023; Liu & Jiang, 2024).

v. **Smart aquaponics system for monitoring and control:** In the field of aquaponics research, the majority of IoT technologies are focused on monitoring. However, control mechanisms are becoming more essential to minimise human interaction and increase management effectiveness in maximising yield (Anila & Daramola, 2024). This can lead to an intelligent and self-regulating aquaponics system (Mahmoud et al., 2023).

vi. **Blockchain in aquaponics:** Blockchain technology makes aquaponics supply networks more transparent and traceable (Manju et al., 2024). It can be used to track every step of the aquaponics supply chain from seed to the actual final produce sold to the consumer. This will allow for complete visibility of the entire process to the customer. A comprehensive understanding of the aquaponics harvesting process among stakeholders and customers will contribute significantly to the holistic development of aquaponics systems.

vii. **Technology integration in aquaponics:** Smart aquaponics represents a growing field of investigation, where existing studies often integrate IoT, machine learning, or a combination of both technologies. Smart technology integration, including expert systems, blockchain, explainable AI (XAI), IoT, and machine learning, needs to be investigated (Anila & Daramola, 2024). Smart technology integration in aquaponics will lead to benefits such as lower labour demands, improved product quality, and more sustainability (Wang et al., 2020; Mahmoud et al., 2023).

viii. **Aquaponics predictions:** Real-time monitoring of the aquaponics system is aided by basic sensors, which are sourced cost-effectively. However, sensors for ammonia, nitrite, and nitrate are costly and difficult to obtain. Machine learning can be used to bridge the gap of having to purchase these expensive sensors by predicting the values needed (Channa et al., 2024). A growing world population requires food security for survival. Mounting concerns over food security have placed great emphasis on developing methods to accurately forecast anticipated crop yields (Muruganantham et al., 2022). This crop yield prediction requirement can be addressed by using relevant sensors and machine learning technologies, which enable better alignment between supply quantities and market demand.

ix. **Decision support system for smart aquaponics:** Future studies could incorporate a wider range of larger data sets thus enhancing the decision support system (DSS) by predicting crop and fish yields more accurately. Real-time data monitoring and collection from sensors or IoT devices could be integrated to continuously update predictions and make automatic adjustments based on constantly changing conditions. This can support an end-to-end smart aquaponics solution, which will provide the entire range of prediction, monitoring, controlling and decision support in aquaponics (Ramirez, 2024; Sridevi et al., 2024; Anila & Daramola, 2024). A future study could focus on building a fully automated aquaponics system that adjusts itself without any outside intervention. This system can regulate itself based on the external weather patterns predicted for that geographical area. This can ensure optimal adaptation to the outside weather and climatic conditions that have an effect on the productivity of the aquaponics system. The web or app interface can provide feedback, recommendations, and suggestions in response to queries from aquaponics practitioners (Ubayasena et al., 2023; Senapaty et al., 2024).

x. **Enhancing aquaponics evaluation through stakeholder participation:** Increasing stakeholder participation in future evaluation studies will help improve the breadth and relevance of the findings. As indicated by Anila and Daramola (2024), very few studies have thoroughly examined the validation of the suggested aquaponics systems, which does not permit effective evaluation of the proposed solution. A mobile application can be developed, thus enabling broader accessibility and increased adoption among stakeholders (Eneh et al., 2023).

xi. **Emerging evaluation methods and metrics in aquaponics for new technologies:** Comparison, observation, and expert feedback were used to evaluate the prototype, whereas performance evaluation metrics were used to assess the machine learning model (Anila & Daramola, 2024). As indicated by Anila and Daramola (2024), further evaluation methods need to be explored as technologies such as IoT, machine learning, explainable AI (XAI), and blockchain are increasingly integrated. The business and organisational requirements and objectives also need to be considered.

REFERENCES

Abbasi, R., Martinez, P. & Ahmad, R. 2022. An ontology model to represent aquaponics 4.0 system's knowledge. *Information Processing in Agriculture*, 9(4): 514–532. https://doi.org/10.1016/j.inpa.2021.12.001.

Abbink, W., Blanco Garcia, A., Roques, J.A.C., Partridge, G.J., Kloet, K. & Schneider, O. 2012. The effect of temperature and pH on the growth and physiological response of juvenile yellowtail kingfish Seriola lalandi in recirculating aquaculture systems. *Aquaculture*, 330: 130–135. https://doi.org/10.1016/j.aquaculture.2011.11.043.

Abdel-Fattah, F., Altamimi, F. & Farhan, K.A. 2021. Machine learning and data mining in cybersecurity. In the *2021 International Conference on Information Technology (ICIT)*. IEEE: 952–956.

Abdelaziz, M.T., Radwan, A., Mamdouh, H., Saad, A.S., Abuzaid, A.S., AbdElhakeem, A.A., Zakzouk, S., Moussa, K. & Darweesh, M.S. 2025. Enhancing network threat detection with random forest-based NIDS and permutation feature importance. *Journal of Network and Systems Management*, 33(2): 1–36. https://doi.org/10.1007/s10922-024-09874-0.

Abdolrasol, M.G.M., Suhail Hussain, S.M., Ustun, T.S., Sarker, M.R., Hannan, M.A., Mohamed, R., Ali, J.A., Mekhilef, S. & Milad, A. 2021. Artificial neural networks-based optimization techniques: A review. *Electronics*, 10: 1–43.

Abdullah, M.S.T. & Mazalan, L. 2022. Smart automation aquaponics monitoring system. *JOIV : International Journal on Informatics Visualization*, 6(1–2): 256–263. http://dx.doi.org/10.30630/joiv.6.1-2.925.

Abidin, Z., Putra, A.D., Zainuri, A. & Maulana, E. 2024. Smart aquaponics with automatic sensors cleaning for zero waste integrated farming system using Internet of Things. In the *2024 International Electronics Symposium (IES)*. IEEE: 120–125.

Abou Elassad, Z.E., Mousannif, H. & Al Moatassime, H. 2020. A proactive decision support system for predicting traffic crash events: a critical analysis of imbalanced class distribution. *Knowledge-Based Systems*, 205: 1–14. https://doi.org/10.1016/j.knosys.2020.106314.

Adadi, A. & Berrada, M. 2018. Peeking inside the Black-Box: a survey on Explainable Artificial Intelligence (XAI). IEEE, 6: 52138–52160.

Ahluwalia, V.K. 2022. Stereochemistry of biomolecules. In *Stereochemistry of Organic Compounds*. Cham,Switzerland: Springer, 477–502.

Alhnaity, B., Pearson, S., Leontidis, G. & Kollias, S. 2019. Using deep learning to predict plant growth and yield in greenhouse environments. In *International Symposium on Advanced Technologies and Management for Innovative Greenhouses: GreenSys2019*. 425–432.

Ali, M.R., Nipu, S.M.A. & Khan, S.A. 2023. A decision support system for classifying supplier selection criteria using machine learning and random forest approach. *Decision Analytics Journal*, 7: 1–14. https://doi.org/10.1016/j.dajour.2023.100238.

Alnuaimi, A.F.A.H. & Albaldawi, T.H.K. 2024. An overview of machine learning classification techniques. In *BIO Web of Conferences*. EDP Sciences: 1–24.

Alselek, M., Alcaraz-Calero, J.M., Segura-Garcia, J. & Wang, Q. 2022. Water IoT monitoring system for aquaponics health and fishery applications. *Sensors 2022*, 22(19): 1–20.

Alshammari, W.B., Alshammery, K., Lotfi, S., Altamimi, H., Alshammari, A., Al-Harbi, N.A., Jakovljević, D., Alharbi, M.H., Moustapha, M.E., Abd El-Moneim, D. & Abdelaal, K. 2024. Improvement of morphophysiological and anatomical attributes of plants under abiotic stress conditions using plant growth-promoting bacteria and safety treatments. PeerJ, 12: 1–25.

AlShrouf, A. 2017. Hydroponics, aeroponic and aquaponic as compared with conventional farming.

American Academic Scientific Research Journal for Engineering, Technology, and Sciences, 27(1): 247–255.

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M. & Farhan, L. 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8: 1–74. https://doi.org/10.1186/s40537-021-00444-8.

Alzubi, J., Nayyar, A. & Kumar, A. 2018. Machine learning from theory to algorithms: an overview. In *Journal of Physics: Conference Series*. IOP Publishing: 1–15.

Amano, J.R., Punongbayan, N., Caacbay, J., Agustin, E., Dela Vega, K.A., Soriano, A., Andaya, F., Mandayo, E. & Beano, M.G. 2022. A comparative analysis of machine learning algorithms for Bok Choy leaf disease identification in smart aquaponics. In *TENCON 2022 - 2022 IEEE Region 10 Conference (TENCON)*. IEEE: 1–6.

Ammayappan, S. & Smys, S. 2020. A Survey on Internet of Things (IoT)-based smart systems. *Journal of ISMAC*, 02(04): 181–189.

Anila, M. & Daramola, O. 2024. Applications, technologies, and evaluation methods in smart aquaponics: a systematic literature review. *Artificial Intelligence Review*, 58(1): 1–52.

Aouedi, O., Vu, T.H., Sacco, A., Nguyen, D.C., Piamrat, K., Marchetto, G. & Pham, Q.V. 2024. A survey on Intelligent Internet of Things: applications, security, privacy, and future directions. *IEEE Communications Surveys and Tutorials*: 1–56.

Arakkal Thaiparambil, N. & Radhakrishnan, V. 2022. Challenges in achieving an economically sustainable aquaponic system: a review. *Aquaculture International*, 30(6): 3035–3066.

Arigela, A.K., Banapuram, C. & Venu, N. 2024. Remote-based home automation with MQTT: ESP32 Nodes and Node-RED on Raspberry Pi. In *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*. IEEE: 313–318.

Arjune, S. & Kumar, S. V. 2022. Smart agriculture adoption based on farmers' perspective. In the *2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC)*. IEEE: 376–379.

Aslem, Y.I. & Abu-Naser, S.S. 2022. CLIPS-expert system to predict coriander diseases. *International Journal of Engineering and Information Systems (IJEAIS)*, 6(6): 89–95.

Awad, M. & Khanna, R. 2015. Deep neural networks. In *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*. Springer: 127–147.

Ayodele, T.O. 2010a. Introduction to machine learning. In Zhang, Y.(ed). *New Advances in Machine Learning*. Rijeka, Croatia: InTech.

Ayodele, T.O. 2010b. Machine learning overview. In Zhang, Y.(ed). *New Advances in Machine Learning*. Rijeka, Croatia: InTech.

Azad, Kamrun Naher, Salam, M.A. & Azad, Khairun Naher. 2016. Aquaponics in Bangladesh: current status and future prospects. *Journal of Bioscience and Agriculture Research*, 07(02): 669–677.

Badillo, S., Banfai, B., Birzele, F., Davydov, I.I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B. & Zhang, J.D. 2020. An Introduction to Machine Learning. Clinical Pharmacology and Therapeutics, 107(4): 871–885.

Bangor, A., Kortum, P. & Miller, J. 2009. Determining what individual SUS scores mean; adding an adjective rating. *Journal of usability studies*, 4(3): 114–123.

Barosa, R., Hassen, S.I.S. & Nagowah, L. 2019. Smart aquaponics with disease detection. In the *2019 Conference on Next Generation Computing Applications (NextComp)*. Mauritius: IEEE: 1–6.

Bhat, A.G., Mukherjee, S., Paradkar, V. & Balley, P. 2023. Unveiling the future of urban and semi-urban hydroponic farming with nutrient film technique. *Biotica Research Today*, 5(11): 780–782.

Bhatia, N. & Vandana. 2010. Survey of Nearest Neighbor techniques. *International Journal of Computer Science and Information Security (IJCSIS)*, 8(2): 302–305.

Bhatnagar, A. & Devi, P. 2013. Water quality guidelines for the management of pond fish culture. *International journal of environmental sciences*, 3(6): 1980–2009.

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A.L., Deng, D. & Lindauer, M. 2023. Hyperparameter optimization: foundations, algorithms, best practices, and open challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 13(2): 1–43.

Bishop, C.M. 1995. Neural networks for pattern recognition. *Clarendon Press google schola*, 2: 223–228.

Blancaflor, E., Banganay, K.N.U., Fernandez, K.E., Jamena, J.N.D., Rabanal, R.S.C. & Zamora, S.L.G. 2022. An IoT monitoring system designed for hydroponics plant cultivation. In *2022 5th International Conference on Computing and Big Data (ICCBD)*. IEEE: 102–106.

Bohanec, M., Bratko, I. & Rajkovic, V. 1990. Expert system for decision making. *Sistemica*, 1(1): 145–157. http://kt.ijs.si/MarkoBohanec/pub/IFIP1983.pdf [6 September 2022].

Breiman, L. 1996. Bagging predictors. *Machine Learning*, 24: 123–140.

Breiman, L. 2001. Random Forests. *Machine Learning*, 45: 5–32.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. 1984. *Classification and regression trees*. New York: CRC Press.

Brooke, J. 1996. SUS - a quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194): 4–7. https://www.researchgate.net/publication/319394819.

Butt, M.F.U., Yaqub, R., Hammad, M., Ahsen, M., Ansir, M. & Zamir, N. 2019. Implementation of aquaponics within IoT framework. In *2019 SoutheastCon*. Huntsville, AL, USA: IEEE: 1–6.

Bwapwa, J.K. 2019. Analysis on industrial and domestic wastewater in South Africa as a water-scarce country. *International Journal of Applied Engineering Research*, 14(7): 1474–1483.

Castro, O., Bruneau, P., Sottet, J.S. & Torregrossa, D. 2023. Landscape of high-performance Python to develop data science and machine learning applications. *ACM Computing Surveys*, 56(3): 1–30.

Chandramenon, P., Aggoun, A. & Tchuenbou-Magaia, F. 2024. Smart approaches to Aquaponics 4.0 with focus on water quality − Comprehensive review. *Computers and Electronics in Agriculture*, 225: 1–12. https://doi.org/10.1016/j.compag.2024.109256.

Channa, A.A., Munir, K., Hansen, M. & Tariq, M.F. 2024. Optimisation of small-scale aquaponics systems using Artificial Intelligence and the IoT: current status, challenges, and opportunities. *Encyclopedia*, 4: 313–336.

Chelghoum, A. 2024. Unveiling respondents' collaboration in online questionnaire responses. *ALTRALANG Journal*, 6(2): 377–388.

Chen, T. & Guestrin, C. 2016. XGBoost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM: 785–794. http://dx.doi.org/10.1145/2939672.2939785.

Chia, S.Y. & Lim, M.W. 2022. A critical review on the influence of humidity for plant growth forecasting. IOP Conference Series: Materials Science and Engineering, 1257: 1–6.

Chicco, D., Warrens, M.J. & Jurman, G. 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7: 1–24. https://doi.org/10.7717/peerj-cs.623.

Chitralekha, G. & Roogi, J.M. 2021. A quick review of ML algorithms. In *2021 6th International Conference on Communication and Electronics Systems (ICCES)*. Coimbatre, India: IEEE: 1–5.

Choi, S.Y. & Kim, A.M. 2019. Development of indoor aquaponics control system using a computational thinking-based convergence instructional model. *Universal Journal of Educational Research*, 7(5A): 68–79.

Chowdhury, H. & Asiabanpour, B. 2024. Influencing factors for the plant growth patterns in hydroponic and aquaponic environments: A subgroup analysis for sustainable agricultural production. *Green Technologies and Sustainability*, 2(2): 1–15.

Cooper, H. 1998. *Synthesizing research: a guide for literature reviews*. 3rd ed. USA: Sage.

Copperwaite, M. & Leifer, C. 2015. *Learning Flask framework*. UK: Packt Publishing Ltd.

Corcoran, P. 2016. The Internet of Things: why now, and what's next? *IEEE Consumer Electronics Magazine*, 5(1): 63–68.

Cresswell, J.W. 2014. *Research design: qualitative, quantitative and mixed methods approaches*. 4th ed. Thousand Oaks, CA: Sage Publications.

Crotty, M. 2003. *The foundations of social research: Meaning and perspective in the research process*. London: Sage Publications Ltd.

Dabool, H., Alashwal, H., Alnuaimi, H., Alhouqani, A., Alkaabi, S. & Al Ahbabi, A. 2024. Comparative analysis of hyperparameter tuning methods in classification models for ensemble learning. In 2024 7th International Conference on Algorithms, Computing and Artificial Intelligence *(ACAI)*. IEEE: 1–5.

Dai, W. & Ji, W. 2014. A MapReduce implementation of C4.5 Decision Tree algorithm. *International Journal of Database Theory and Application*, 7(1): 49–60.

Daiane, de S.L.W., Daniele, F. de O., Hudson, de O.R., Guilherme, F.F., Luiz, A.A.G., Erica, C.A.L. & Gustavo, C. 2021. Nitrate concentration and nitrate/ammonium ratio on lettuce grown in hydroponics in Southern Amazon. *African Journal of Agricultural Research*, 17(6): 862–868.

Dangeti, P. 2017. *Statistics for Machine Learning*. Birmingham, UK: Packt Publishing Ltd.

Daniel, M.A., Chong, S., Chong, L. & Wee, K. 2025. Optimising phishing detection : a comparative analysis of machine learning methods with feature selection. *Journal of Informatics and Web Engineering*, 4(1): 1–13.

Daramola, O., Kavu, T.D., Kotze, M.J., Marnewick, J.L., Sarumi, O.A., Kabaso, B., Moser, T., Stroetmann, K., Fwemba, I., Daramola, F., Nyirenda, M., van Rensburg, S.J. & Nyasulu, P.S. 2025. Predictive modelling and identification of critical variables of mortality risk in COVID-19 patients. Scientific Reports, 15(1): 1–20. https://doi.org/10.1038/s41598-023-46712-w.

Darbi, H.A. & Saleh, E.M. 2022. Decision support system: analysis and design methodology. In *2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)*. IEEE: 260–266.

Das, A. & Rad, P. 2020. Opportunities and challenges in Explainable Artificial Intelligence (XAI): a survey. : 1–24. http://arxiv.org/pdf/2006.11371.pdf.

Debroy, P., Majumder, P., Majumdar, P., Das, A. & Seban, L. 2025. Analysis of opportunities and challenges of smart aquaponic system: a summary of research trends and future research avenues. *Sustainable Environment Research*, 35(1): 1–19.

Debroy, P. & Seban, L. 2022a. A fish biomass prediction model for aquaponics system using machine learning algorithms. In Chen, J. I.-Z., Wang, H., Du, K.-L., & Suma, V.(eds.). *Machine learning and autonomous systems. Smart innovation, systems and technologies*. Singapore: Springer: 383–397. https://doi.org/10.1007/978-981-16-7996-4_27.

Debroy, P. & Seban, L. 2022b. A tomato fruit biomass prediction model for aquaponics system using machine learning algorithms. *IFAC-PapersOnLine*, 55(1): 709–714. https://doi.org/10.1016/j.ifacol.2022.04.116.

Defa, R.P., Ramdhani, M., Priramadhi, R.A. & Aprillia, B.S. 2019. Automatic controlling system and IoT-based monitoring for pH rate on the aquaponics system. In *Journal of Physics: Conference Series*. IOP Publishing: 1–7.

Deng, C., Gao, Y., Gu, J., Miao, X. & Li, S. 2010. Research on the growth model of aquaculture organisms based on neural network expert system. In the *2010 Sixth International Conference on Natural Computation*. Yantai, China: IEEE: 1812–1815.

Desdhanty, V.S. & Rustam, Z. 2021. Liver cancer classification using random forest and extreme gradient boosting (xgboost) with genetic algorithm as feature selection. In *2021 International Conference on Decision Aid Sciences and Application (DASA)*. Sakheer, Bahrain: IEEE: 716–719.

Deshpande, A., Deshpande, P. & Padmini, T.N. 2024. Smart aquaponics: water quality monitoring and agriculture and environment. In *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)*. Vellore, India: IEEE: 1–6.

Dewangan, S.K. & Shrivastava, S.K. 2024. A correlation Between Ec and TDS in Water: a Review. International Journal of Progressive Research in Engineering Management and Science (IJPREMS), 05(4): 180–186. www.ijprems.com.

Dewangan, S.K., Shrivastava, S.K., Kadri, M.A., Saruta, S., Yadav, S. & Minj, N. 2023. Temperature effect on electrical conductivity (EC) & total dissolved solids (TDS) of water: A review. International Journal of Research and Analytical Reviews, 10(2): 514–520. https://www.researchgate.net/publication/371539432.

Dharshan, A., Kumar, P., Ravimaran, S. & Srinivasulu Reddy, U. 2024. LSTM-based soft-sensor for estimating nitrate concentration in aquaponics pond. In *2024 3rd International Conference for Innovation in Technology (INOCON)*. Bangalore, India: IEEE: 1–8.

Dubey, S., Pandey, R.K. & Gautam, S.S. 2013. Literature review on Fuzzy Expert system in agriculture. *International Journal of Soft Computing and Engineering*, 2(6): 289–291.

Dupont, C., Cousin, P. & Dupont, S. 2018. IoT for aquaculture 4.0 smart and easy-to-deploy real-time water monitoring with IoT. In the *2018 Global Internet of Things Summit (GIoTS)*. Bilbao, Spain: IEEE: 1–5.

Dutta, A., Dahal, P., Tamang, P., Saban Kumar, E. & Prajapati, R. 2018. IoT-based aquaponics monitoring system. In *1st KEC Conference Proceedings*. Lalitpur: 75–80. https://www.researchgate.net/publication/327953706 [22 October 2021].

Ekanayake, D., De Alwis, P., Harshana, P., Munasinghe, D., Jayakody, A. & Gamage, M.N. 2022. A smart aquaponic system for enhancing the revenue of farmers in Sri Lanka. In the *2022 International Conference on ICT for Smart Society (ICISS)*. IEEE: 1–6.

Ekanayake, I.U., Meddage, D.P.P. & Rathnayake, U. 2022. A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using SHapley additive explanations (SHAP). *Case Studies in Construction Materials*, 16: 1–20. https://doi.org/10.1016/j.cscm.2022.e01059.

Elgeldawi, E., Sayed, A., Galal, A.R. & Zaki, A.M. 2021. Hyperparameter tuning for machine learning algorithms used for Arabic sentiment analysis. *Informatics*, 8(4): 1–21. https://doi.org/10.3390/informatics8040079.

Elguea-Aguinaco, Í., Serrano-Muñoz, A., Chrysostomou, D., Inziarte-Hidalgo, I., Bøgh, S. & Arana-Arexolaleiba, N. 2023. A review on reinforcement learning for contact-rich robotic manipulation tasks. *Robotics and Computer-Integrated Manufacturing*, 81: 1–24.

Emmanuel, G., Meyliana, Hidayanto, A.N. & Prabowo, H. 2022. A literature review: drug supply chain and neural networks. In *Proceeding - 6th International Conference on Information Technology, Information Systems and Electrical Engineering: Applying Data Sciences and Artificial Intelligence Technologies for Environmental Sustainability (ICITISEE)*. IEEE: 559–564.

Eneh, A.H., Udanor, C.N., Ossai, N.I., Aneke, S.O., Ugwoke, P.O., Obayi, A.A., Ugwuishiwu, C.H. & Okereke, G.E. 2023. Towards an improved Internet of Things sensors data quality for a smart aquaponics system yield prediction. *MethodsX*, 11: 1–8. https://doi.org/10.1016/j.mex.2023.102436.

Ergün, S. 2023. Explaining XGboost predictions with Shap value: a comprehensive guide to interpreting Decision Tree-based models. *New Trends in Computer Sciences*, 1(1): 19–31.

Espinal, C.A. & Matulić, D. 2019. Recirculating aquaculture technologies. In Goddek, S., Joyce, A., Kotzen, B., & Burnell, G. M.(eds.). *Aquaponics food production systems*. Cham, Switzerland: Springer International Publishing, 35–79.

Eze, E. & Ajmal, T. 2020. Dissolved oxygen forecasting in aquaculture: a hybrid model approach. *Applied Sciences*, 10(20): 1–14.

Febriandirza, A., Kurniawan, M.R. & Nastiti, A.D. 2023. Web-based decision support system of employee admission using multi-attribute utility theory. In the *2023 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*. IEEE: 120–124.

Fenner, M. 2019. *Machine learning with Python for everyone*. Boston: Addison-Wesley Professional.

Firouzi, F., Farahani, B., Weinberger, M., DePace, G. & Aliee, F.S. 2020. *IoT fundamentals: definitions, architectures, challenges, and promises*. Firouzi, F., Krishnendu, C., & Nassif, S.(eds.). Cham: Springer. https://doi.org/10.1007/978-3-030-30367-9_1.

Fontoura, L.C.M.M., De Castro Lins, H.W., Bertuleza, A.S., D'assunção, A.G. & Neto, A.G. 2021. Synthesis of multiband frequency selective surfaces using machine learning with the decision tree algorithm. *IEEE Access*, 9: 85785–85794.

Ford, F.N. 1985. Decision support systems and expert systems: a comparison. *Information & Management*, 8(1): 21–26. https://doi.org/10.1016/0378-7206(85)90066-7.

Forgionne, G.A. 2003. An Architecture for the integration of decision-making support functionalities. In Mora, M., Forgionne, G. A., & Gupta, J. N. D.(eds.). *Decision-Making Support Systems : Achievements, Trends, and Challenges for the New Decade*. Hershey, PA: Idea Group Inc.

Francis-Floyd, R., Watson, C., Petty, D. & Pouder, D.B. 2009. Ammonia in aquatic systems. *University of Florida (UF)/Institute of Food and Agricultural Sciences (IFAS), FA*, 16: 1–5.

Frasetya, B., Harisman, K. & Ramdaniah, N.A.H. 2021. The effect of hydroponics systems on the growth of lettuce. In *IOP Conference Series: Materials Science and Engineering*. IOP Publishing: 1–6.

French, S. & Turoff, M. 2007. Decision support systems. *Communications of the ACM*, 50(3): 39–40. https://doi.org/10.1145/1226736.1226762.

Freund, Y. & Schapire, R.E. 1999. A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*, 14(5): 771–780.

Friedman, J.H. 2001. Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232.

Friuli, M., Masciullo, A., Blasi, F.S., Mita, M., Corbari, L. & Surano, I. 2021. A 4.0 Sustainable aquaponic system based on the combined use of superabsorbing natural hydrogels and innovative sensing technologies for the optimization of water use. In the *2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI)*. Naples, Italy: IEEE: 429–434.

Fruscella, L., Kotzen, B. & Milliken, S. 2021. Organic aquaponics in the European Union: towards sustainable farming practices in the framework of the new EU regulation. *Reviews in Aquaculture*, 13(3): 1661–1682.

Fukase, E. & Martin, W. 2020. Economic growth, convergence, and world food demand and supply. *World Development*, 132: 1–12. https://doi.org/10.1016/j.worlddev.2020.104954.

Gaftandzhieva, S., Hussain, S., Hilčenko, S., Doneva, R. & Boykova, K. 2023. Data-driven decision-making in Higher Education Institutions: State-of-play. International Journal of Advanced Computer Science and Applications (IJACSA) International, 14(6): 397–405.

Gan, L. 2022. XGBoost-based e-commerce customer loss prediction. *Computational Intelligence and Neuroscience*, 2022: 1–10. https://doi.org/10.1155/2022/1858300.

Gao, J., Li, M. & Zhuang, W. 2021. Industrial Internet of Things: smart factory. Wireless Networks (United Kingdom), 8(9): 11–61.

Ganesha, H.R. & Aithal, P.S. 2022. How to choose an appropriate research data collection method and method choice among various research data collection methods and method choices during PhD. program in India? *International Journal of Management, Technology, and Social Sciences(IJMTS)*, 7(2): 455–489.

Garbev, A. 2022. Possibilities for visual programming in robotics via Node-RED. In the *2022 International Conference Automatics and Informatics (ICAI)*. Varna, Bulgaria: IEEE: 367–371.

Garg, K., Garg, S., Dixit, H. & Pandey, K. 2024. Machine learning-driven analysis of mental health indicators in social media posts. In *2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC)*. IEEE: 1056–1062.

Géron, A. 2019. *Hands-on machine learning with Scikit-Learn, Keras & TensorFlow*. 2nd ed. Tache, N.(eds.). Sebastopol: O'Reilly Media, Inc.

Ghandar, A., Ahmed, A., Zulfiqar, S., Hua, Z., Hanai, M. & Theodoropoulos, G. 2021. A decision support system for urban agriculture using digital twin: a case study with aquaponics. *IEEE Access*, 9: 35691–35708.

Giuseppe, B. 2018. *Machine learning algorithms : popular algorithms for data science and machine learning*. 2nd ed. Birmingham: Packt Publishing. https://research.ebsco.com/linkprocessor/plink?id=11708ffd-8f29-3156-aec6-7901ec31fa93. [24 March 2024].

Gnanasagar, V. & Vivek, M. 2020. Design and implementation of a controller for a recirculating aquaponics system using IoT. *International Research Journal of Engineering and Technology (IRJET)*, 7(5): 6347–6351. https://www.researchgate.net/publication/341817455 [28 October 2021].

Goddek, S., Delaide, B., Mankasingh, U., Ragnarsdottir, K.V., Jijakli, H. & Thorarinsdottir, R. 2015. Challenges of sustainable and commercial aquaponics. *Sustainability*, 7(4): 4199–4224.

Goddek, S., Joyce, A., Kotzen, B. & Burnell, G.M. eds. 2019. *Aquaponics food production systems*. Cham, Switzerland: Springer International Publishing.

Goddek, S., Joyce, A., Wuertz, S., Körner, O., Bläser, I., Reuter, M. & Keesman, K.J. 2019. Decoupled aquaponics systems. In *Aquaponics food production systems*. Cham, Switzerland: Springer International Publishing, 201–229.

Grinberg, M. 2018. *Flask web development: developing web applications with Python*. 2nd ed. Sebastopol, CA: O'Reilly Media, Inc.

Gubbi, J., Buyya, R., Marusic, S. & Palaniswami, M. 2013. Internet of Things (IoT): a vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7): 1645–1660.

Gupta, I. & Nagpal, G. 2020. *Artificial Intelligence and Expert Systems*. Dulles, VA: Mercury Learning & Information.

Habibi Aghdam, H., Jahani Heravi, E. & AG, S.I.P. 2018. *Guide to convolutional neural networks: a practical application to traffic-sign detection and classification*. Germany: Springer International Publishing.

Hadi, M.S., Sihombing, Y.P., Mustika, S.N., Mizar, M.A., Lestari, D. & Izhar, C.A.A. 2022. Aquaponic plant Control and Monitoring System Using IoT-based decision tree logic. In the *2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*. Yogyakarta, Indonesia: IEEE: 705–710.

Hak, F., Guimaraes, T. & Santos, M. 2022. Towards effective clinical decision support systems: a systematic review. *PLoS ONE*, 17(8): 1–19. https://doi.org/10.1371/journal.pone.0272846.

Hao, Y., Ding, K., Xu, Y., Tang, Y., Liu, D. & Li, G. 2020. States, trends, and future of aquaponics research. *Sustainability (Switzerland)*, 12(18): 1–14.

Haruo, Y., Yamamoto, H., Arakawa, M. & Naka, I. 2020. Development and evaluation of environmental/growth observation sensor network system for aquaponics. In the *2020 IEEE International Conference on Consumer Electronics (ICCE)*. Las Vegas, NV, USA: IEEE: 1–6.

Haykin, S. 1994. *Neural networks: a comprehensive foundation*. Upper Saddle River, NJ: Prentice Hall.

He, F. 2023. IPhone sales prediction based on multilinear regression model: evidence from Statista. In *2023 IEEE International Conference on Sensors, Electronics and Computer Engineering (ICSECE)*. IEEE: 341–346.

Héder, M. 2023. Explainable AI: a brief history of the concept. *Ercim News*, (134): 9–10.

Hızıroğlu, A., Pişirgen, A., Özca, M. & A, İ. 2022. Artificial Intelligence in healthcare industry: A transformation from model-driven to knowledge-driven DSS. Artificial Intelligence Theory and Applications, 2(1): 41–58. https://dergipark.org.tr/en/pub/aita/issue/70443/1136223.

Hong, Y., Lee, J., Park, S., Kim, J. & Jang, K.J. 2024. Next-generation nitrate, ammonium, phosphate, and potassium ion monitoring system in closed hydroponics: review on state-of-the-art sensors and their applications. *AgriEngineering*, 6(4): 4786–4811.

Hosny, K.M., Magdi, A., Salah, A., El-Komy, O. & Lashin, N.A. 2023. Internet of Things applications using Raspberry-Pi: a survey. *International Journal of Electrical and Computer Engineering(IJECE)*, 13(1): 902–910.

Hsiao, S.J. & Sung, W.T. 2020. Building a fish-vegetable coexistence system based on a wireless sensor network. *IEEE Access*, 8: 192119–192131.

Humayun, K.M., Sobur, A., Ruhul Amin, M. & Student, G. 2023. Walmart data analysis using machine learning. *International Journal of Creative Research Thoughts*, 11(7): 1–5.

Imandoust, S.B. & Bolandraftar, M. 2013. Application of K-Nearest Neighbor (KNN) approach for predicting economic events : theoretical background. *Int. Journal of Engineering Research and Applications*, 3(5): 605–610.

Islam, T., Chisty, T.A. & Chakrabarty, A. 2018. A deep neural network approach for crop selection and yield prediction in Bangladesh. In the *2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*. IEEE: 1–6.

Ismaili, S., Idrizi, F., Rustemi, A., Ibraimi, M. & Idrizi, H. 2024. IoT-based irrigation system for smart agriculture. In the *2024 XXXIII International Scientific Conference Electronics (ET)*. IEEE: 1–

6.

Jamali, M.A.J., Heidari, A., Allahverdizadeh, P., Norouzi, F. & Bahrami, B. 2020. The IoT landscape. In *Towards the Internet of Things. EAI/Springer Innovations in Communication and Computing*. Cham: Springer, 1–8.

Janiesch, C., Zschech, P. & Heinrich, K. 2021. Machine learning and deep learning. *Electronic Markets*, 31(3): 685–695.

Janjanam, D., Ganesh, B. & Manjunatha, L. 2021. Design of an expert system architecture: an overview. In *Journal of Physics*. IOP Publishing: 1–7.

Jerry, J.A. 2020. Automated aquaponics system: smart farming for the changing climate. https://iit-techambit.in/automated-aquaponics-system/ [31 October 2021].

Jia, J. & Wang, W. 2020. Review of reinforcement learning research. In *Proceedings - 2020 35th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. IEEE: 186–191.

Jiang, W. & Liu, J. 2024. NeuroAqua: developing an optimized artificial intelligence and Internet of Things-based aquaponics system. In *2024 IEEE 3rd International Conference on Computing and Machine Intelligence, ICMI 2024 - Proceedings*. IEEE: 1–5.

John, J. & Mahalingam, P.M.P.R. 2021. Automated fish feed detection in IoT-based aquaponics system. In *2021 8th International Conference on Smart Computing and Communications: Artificial Intelligence, AI-Driven Applications for a Smart World, ICSCC 2021*. Kochi, Kerala, India: Institute of Electrical and Electronics Engineers Inc.: 286–290.

Jones Jr., J.B. 2005. *Hydroponics : a practical guide for the soilless grower*. 2nd ed. Boca Raton: CRC Press.

Joshi, A. V. 2020. *Machine learning and artificial intelligence*. Cham: Springer.

Joy, D. & Sreekumar, K. 2014. A survey on expert system in agriculture. *International Jounal of Computer Science and Information Technology*, 5(6): 7861–7864.

Kadam, S., Gohokar, V. & Kute, R. 2025. Machine learning and Explainable AI for Thai Basil Growth Prediction in Hydroponics. IEEE Access, 13: 99479–99489.

Kaelbling, L.P., Littman, M.L. & Moore, A.W. 1996. Reinforcement learning: a survey. *Journal of Artificial Intelligence Research*, 4: 237–285. https://www.jair.org/index.php/jair/article/view/10166 [14 June 2022].

Kalasampath, K., Kn, S., Sajeev, S., Kuppa, S.S., Ajay, K. & Angulakshmi, M. 2025. A Literature Review on Applications of Explainable Artificial Intelligence (XAI). IEEE Access, 13: 41111–41140.

Kamil, M., Effendi, R., Kassim, M., Sulaiman, A. & Shahbudin, S. 2020. IoT smart agriculture for aquaponics and maintaining goat stall system. *International Journal of Integrated Engineering*, 12(8): 240–250.

Kannan, M., Elavarasan, G., Balamurugan, A., Dhanusiya, B. & Freedon, D. 2022. Hydroponic farming – A state of art for the future agriculture. *Materials Today: Proceedings*, 68: 2163–2166. https://doi.org/10.1016/j.matpr.2022.08.416.

Karimanzira, D. & Rauschenbach, T. 2021. An intelligent management system for aquaponics. *Automatisierungstechnik*, 69(4): 345–350. https://www.degruyter.com/document/doi/10.1515/auto-2020-0036/html [27 October 2021].

Kassim, M.R.M. 2020. IoT applications in smart agriculture: issues and challenges. In the *2020 IEEE Conference on Open Systems, ICOS 2020*. IEEE: 19–24.

Kathuria, S., Rawat, P., Kumar, G.R., Singh, B. & Taneja, A. 2024. Amalgamation of technologies for smart monitoring of sustainable aquaculture. In the *2023 International Conference on Smart Devices (ICSD)*. IEEE: 1–5.

Keerthana, M., Meghana, K.J.M., Pravallika, S. & Kavitha, M. 2021. An ensemble algorithm for crop yield prediction. In *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*. IEEE: 963–970.

Keleko, A.T., Kamsu-Foguem, B., Ngouna, R.H. & Tongne, A. 2023. Health condition monitoring of a complex hydraulic system using Deep Neural Network and DeepSHAP explainable XAI. Advances in Engineering Software, 175: 1–24.

Khan, A.A., Chaudhari, O. & Chandra, R. 2024. A review of ensemble learning and data augmentation models for class imbalanced problems: combination, implementation and evaluation. *Expert Systems with Applications*, 244: 1–29.

Khandakar, A., Elzein, I.M., Nahiduzzaman, M., Ayari, M.A., Ashraf, A.I., Korah, L., Zyoud, A., Ali, H. & Badawi, A. 2024. Smart aquaponics: an innovative machine learning framework for fish farming optimization. *Computers and Electrical Engineering*, 119: 1–20. https://doi.org/10.1016/j.compeleceng.2024.109590.

Khaoula, T., Abdelouahid, R.A., Ezzahoui, I. & Marzak, A. 2021. Architecture design of monitoring and controlling of IoT-based aquaponics system powered by solar energy. *Procedia Computer Science*, 191: 493–498. https://doi.org/10.1016/j.procs.2021.07.063.

Khatri, K.K. 2020. Research paradigm: a philosophy of educational research. *International Journal of English Literature and Social Sciences*, 5(5): 1435–1440.

Khiem, N.M., Takahashi, Y., Yasuma, H., Dong, K.T.P., Hai, T.N. & Kimura, N. 2022. A novel machine learning approach to predict the export price of seafood products based on competitive information: the case of the export of Vietnamese shrimp to the US market. *PLoS ONE*, 17(9): 1–20. http://dx.doi.org/10.1371/journal.pone.0275290.

Kim, J., Yu, B. & O'Hara, S. 2022. LSTM filter for smart agriculture. *Procedia Computer Science*, 210: 289–294.

Kitchenham, B. & Charters, S.M. 2007. *Guidelines for performing systematic literature reviews in software engineering*. UK. https://www.researchgate.net/publication/302924724_Guidelines_for_performing_Systematic_Literature_Reviews_in_Software_Engineering [20 January 2023].

Kjellby, R.A., Cenkeramaddi, L.R., Froytlog, A., Lozano, B.B., Soumya, J. & Bhange, M. 2019. Long-range Self-powered IoT devices for agriculture aquaponics based on multi-hop topology. In the *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*. Limerick, Ireland: *IEEE*: 545–549.

Kledal, P.R., König, B. & Matulić, D. 2019. Aquaponics: the ugly duckling in organic regulation. In Goddek, S., Joyce, A., Kotzen, B., & M, B. G.(eds.). *Aquaponics Food Production Systems*. Springer Open: 487–500.

van Klompenburg, T., Kassahun, A. & Catal, C. 2020. Crop yield prediction using machine learning: a systematic literature review. *Computers and Electronics in Agriculture*, 177: 1–18.

Kok, C.L., Kusuma, I.M.B.P., Koh, Y.Y., Tang, H. & Lim, A.B. 2024. Smart aquaponics: an automated water quality management system for sustainable urban agriculture. *Electronics*, 13(5): 1–19.

König, B., Janker, J., Reinhardt, T., Villarroel, M. & Junge, R. 2018. Analysis of aquaponics as an emerging technological innovation system. *Journal of Cleaner Production*, 180: 232–243.

Kortum, P.T. & Bangor, A. 2013. Usability ratings for everyday products measured with the system usability scale. *International Journal of Human-Computer Interaction*, 29(2): 67–76.

Krishna, R.D., Siva, P.S., Sai Kumar, A., Kunduru, J. & Batta, N. 2023. IoT-based water quality monitoring for smart aquaculture. In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE: 1–6.

Kumar, A. & Savaridassan, P. 2023. Monitoring and accelerating plant growth using IoT and hydroponics. In the *2023 International Conference on Computer Communication and Informatics (ICCCI)*. Coimbatore, India: IEEE: 1–6.

Kumar, H.N., Baskaran, S., Hariraj, S. & Krishnan, V. 2016. An autonomous aquaponics system using 6LoWPAN-based WSN. In *2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*. Vienna, Austria: IEEE: 125–132.

Kumar, P., Tiwari, P. & Srinivasulu, R.U. 2023. Estimating fish weight growth in aquaponic farming through machine learning techniques. In *2023 3rd International Conference on Intelligent Technologies (CONIT)*. IEEE: 1–7.

Kumar, R. & Sharma, B. 2023. Comparative analysis of smart cities-based architecture, applications, technologies, & challenges in Internet of Things. In *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*. IEEE: 1–5.

Kurian, O.A., Saji, A., Joseph, S. & Kuriakose, B.P. 2019. Automated water quality monitoring system for aquaponics. *International Research Journal of Engineering and Technology (IRJET)*, 06(05): 7832–7841.

Kusuma, C., Meela, N., Saridena, S., Kanuparthi, Y.H., Kaushik, M.N. & Sumathi, D. 2023. Automated monitoring system for healthier aquaculture farming. In *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*. Chennai, India: IEEE: 1–11.

Kyaw, T.Y. & Ng, A.K. 2017. Smart aquaponics system for urban farming. *Energy Procedia*, 143: 342–347.

Lai, Q., Zhou, C., Shi, M. & Gao, K. 2022. Identification and analysis of IoT devices based on multilayer perceptron. In *2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE)*. IEEE: 69–73.

Lantz, B. 2013. *Machine Learning with R*. First. Birmingham: Packt Publishing Ltd.

Lantz, B. 2015. Machine Learning with R. Second. Birmingham: Packt Publishing Ltd.

Lauguico, S.C., Concepcion, R.S., Alejandrino, J.D., Tobias, R.R., Macasaet, D.D. & Dadios, E.P. 2020. A comparative analysis of machine learning algorithms modeled from machine vision-based lettuce growth stage classification in smart aquaponics. *International Journal of Environmental Science and Development*, 11(9): 442–449.

Lebied, M. 2018. A guide to the methods, benefits & problems of the interpretation of data. https://www.datapine.com/blog/data-interpretation-methods-benefits-problems/ [27 October 2021].

LeCun, Y., Bengio, Y. & Hinton, G. 2015. Deep learning. *Nature*, 521(7553): 436–444. https://www.nature.com/articles/nature14539 [27 July 2022].

Lee, C.H. & Jhang, J.H. 2019. System design for Internet of Things-assisted urban aquaponics farming. In the *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*. Osaka, Japan: Institute of Electrical and Electronics Engineers Inc.: 986–987.

Lee, Y.G., Oh, J.Y., Kim, D. & Kim, G. 2023. SHAP value-based feature importance analysis for short-term load forecasting. *Journal of Electrical Engineering and Technology*, 18(1): 579–588. https://doi.org/10.1007/s42835-022-01161-9.

Lekić, M. & Gardašević, G. 2018. IoT sensor integration to Node-RED platform. In the *2018 17th International Symposium Infoteh-Jahorina (Infoteh)*. East Sarajevo, Bosnia and Herzegovina: IEEE: 1–5.

Lennard, W. & Goddek, S. 2019. Aquaponics: the basics. In Goddek, S., Joyce, A., Kotzen, B., & Burnell, G. M.(eds.). *Aquaponics Food Production Systems*. 113–145.

Li, H., Cao, Y., Li, S., Zhao, J. & Sun, Y. 2020. XGBoost model and its application to personal credit evaluation. *IEEE Intelligent Systems*, 35(3): 52–61.

Li, Zhiqiang, Li, S., Li, Zhilong, Hu, Y. & Gao, H. 2021. Application of XGBoost in P2P default prediction. In *Journal of Physics: Conference Series*. IOP Publishing: 1–9.

Liao, S.H. 2005. Expert system methodologies and applications—a decade review from 1995 to 2004. *Expert Systems with Applications*, 28(1): 93–103.

Lin, N. & Shi, W. 2014. The research on Internet of Things application architecture based on web. *In 2014 IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA)*: 184–187.

Linardatos, P., Papastefanopoulos, V. & Kotsiantis, S. 2021. Explainable AI: a review of machine learning interpretability methods. *Entropy*, 23(1): 1–45.

Lindholm-Lehto, P. 2023. Water quality monitoring in recirculating aquaculture systems. *Aquaculture, Fish and Fisheries*, 3(2): 113–131.

Lipton, Z.C., Berkowitz, J. & Elkan, C. 2015. A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*: 1–38.

Liu, J. & Jiang, W. 2024. Optimization of aquaponics system efficiency based on artificial intelligence approach. In *2024 IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI)*. IEEE: 1–6.

Liu, J., Zhang, C., An, D. & Wei, Y. 2024. Development and application of an innovative dissolved oxygen prediction fusion model. *Computers and Electronics in Agriculture*, 227: 1–12. https://doi.org/10.1016/j.compag.2024.109496.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y. & Alsaadi, F.E. 2017. A survey of deep neural network architectures and their applications. *Neurocomputing*, 234: 11–26.

Liu, Y., Dong, Y., Jiang, Z. & Chen, X. 2024. Interpretable predictive model for inclusions in electroslag remelting based on XGBoost and SHAP analysis. *Metallurgical and Materials Transactions B: Process Metallurgy and Materials Processing Science*, 55(3): 1428–1441.

Lorenzo, A.R., Dula, A.Y., Valeroso, N.A.C., David, J., Supang, B.N.G., Victoria Padilla, M., Madrigal, G.A.M., Amado, T.M. & Tolentino, L.K.S. 2019. Dissolved oxygen (DO) meter hydrological modelling using predictive algorithms. In the *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*. IEEE: 1–6.

Lu, J., Zhang, G., Ruan, D. & Wu, F. 2007. *Multi-Objective Group Decision Making*. Chen, W. K.(ed). London: Imperial College Press.

Lundberg, S.M. & Lee, S.I. 2017. A unified approach to interpreting model predictions. In *Neural Information Processing Systems*. Long Beach, CA, USA: 1–10.

Luta, D.A. & Siregar, M. 2023. The effectiveness of giving various planting media to the floating raft aquaponics system on the growth of shallots. *The International Conference on Education, Social Sciences and Technology (ICESST)*, 2(1): 60–65.

Lynda, D., Brahim, F., Hamid, S. & Hamadoun, C. 2023. Towards a semantic structure for classifying IoT agriculture sensor datasets : an approach based on machine learning and web semantic technologies. *Journal of King Saud University-Computer and Information Sciences*, 35(8): 1–23.

Machlev, R., Heistrene, L., Perl, M., Levy, K.Y., Belikov, J., Mannor, S. & Levron, Y. 2022. Explainable Artificial Intelligence (XAI) techniques for energy and power systems: review, challenges and opportunities. *Energy and AI*, 9: 1–13. https://doi.org/10.1016/j.egyai.2022.100169.

Mahajan, P., Uddin, S., Hajati, F. & Moni, M.A. 2023. Ensemble learning for disease prediction: a review. *Healthcare*, 11: 1–21.

Mahesh, B. 2020. Machine learning algorithms: a review. *International Journal of Science and Research (IJSR)*, 9(1): 381–386.

Mahkeswaran, R. & Ng, A.K. 2020. Smart and sustainable home aquaponics system with feature-rich Internet of Things mobile application. In the *2020 6th International Conference on Control, Automation and Robotics (ICCAR)*. Institute of Electrical and Electronics Engineers Inc.: 607–611.

Mahmoud, M.M.M., Darwish, R. & Bassiuny, A.M. 2023. Development of an economic smart aquaponic system based on IoT. *Journal of Engineering Research*, 12(4): 886–894. https://doi.org/10.1016/j.jer.2023.08.024.

Maity, P., Sur, K. & Muhopadhyay, I. 2023. The future of the Internet of Things (IoT) and IoT authentication. In the *2023 11th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON)*. IEEE: 1–6.

Mandap, J.P., Sze, D., Reyes, G.N., Dumlao, S.M., Reyes, R. & Chung, W.Y.D. 2018. Aquaponics pH level, temperature, and dissolved oxygen monitoring and control system using Raspberry Pi as network backbone. In *TENCON 2018 - 2018 IEEE Region 10 Annual International Conference*. IEEE: 1381–1386.

Manju, M., Indira Priya, P., Shivaraj, S., Rissi Kumar, P. & Shivesh, P.R. 2024. Smart fields: enhancing agriculture with machine learning. In *2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications, Theme: Healthcare and Internet of Things (AIMLA)*. IEEE: 1–5.

Manju, M., Karthik, V., Hariharan, S. & Sreekar, B. 2017. Real-time monitoring of the environmental parameters of an aquaponic system based on Internet of Things. In the *2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM)*. Chennai, India: IEEE: 943–948.

Mason, B.J., Morgan, L., Abdul, P.J. & Beerman, M. 2018. *Nutrients for hydroponics and tissue culture*. Nerang, Queensland: ACS Distance Education.

Mathew, A., Amudha, P. & Sivakumari, S. 2021. Deep learning techniques: an overview. In Hassanien, A., Bhatnagar, R., & Darwish, A.(eds.). *Advanced machine learning technologies and applications. AMLTA 2020. Advances in Intelligent Systems and Computing*. Singapore: Springer: 599–608.

Matzavela, V. & Alepis, E. 2021. Decision tree learning through a predictive model for student academic performance in intelligent M-Learning environments. *Computers and Education: Artificial Intelligence*, 2: 1–12.

Maucieri, C., Nicoletto, C., Van Os, E., Anseeuw, D., Van Havermaet, R. & Junge, R. 2019. Hydroponic Technologies. In Goddek, S., Joyce, A., Kotzen, B., & Burnell, G. M.(eds.). *Aquaponics Food Production Systems*. 77–110.

Maulini, R., Sahlinal, D. & Arifin, O. 2022. Monitoring of pH, ammonia (NH3) and temperature parameters aquaponic water in the 4.0 revolution era. In *IOP Conference Series: Earth and Environmental Science*. IOP Publishing: 1–10. https://iopscience.iop.org/article/10.1088/1755-1315/1012/1/012087 [2 June 2023].

Mchunu, N., Lagerwall, G. & Senzanje, A. 2018. Aquaponics in South Africa: results of a national survey. *Aquaculture Reports*, 12: 12–19.

Mchunu, N., Lagerwall, G. & Senzanje, A. 2019. Aquaponics model specific to South African conditions. *South African Journal of Agricultural Extension (SAJAE)*, 47(1): 73–91.

Megdad, M.M.M., Ayyad, M.N., Al-Qadi, M.H., El-Habibi, M.F., Alqatrawi, M.J.A., Sababa, R.Z. & Abu-Naser, S.S. 2022. Mint expert system diagnosis and treatment. *International Journal of Academic Information Systems Research*, 6(5): 22–28.

Mekonnen, Y., Namuduri, S., Burton, L., Sarwat, A. & Bhansali, S. 2020. Review—machine learning techniques in wireless sensor network based precision agriculture. *Journal of The Electrochemical Society*, 167(3): 1–10.

Menon, P.C. 2020. IoT-enabled aquaponics with wireless sensor smart monitoring. In *Proceedings of the 4th International Conference on IoT in Social, Mobile, Analytics and Cloud, ISMAC 2020*. Institute of Electrical and Electronics Engineers Inc.: 171–176.

Mienye, I.D. & Jere, N. 2024. A survey of Decision Trees: concepts, algorithms, and applications. *IEEE Access*, 12: 86716–86727.

Milutinovíc, G., Seipel, S. & Ahonen-Jonnarth, U. 2021. Geospatial decision-making framework based on the concept of satisficing. *ISPRS International Journal of Geo-Information*, 10(5): 1–20.

Misra, T. & Dinker, A.G. 2025. DIDM: Deepfake image detection using deep learning. In 2025 International Conference on Networks and Cryptology (NETCRYPT). IEEE: 1–6.

Mitchell, R. & Frank, E. 2017. Accelerating the XGBoost algorithm using GPU computing. *PeerJ Computer Science*, 3(7): 1–37. https://peerj.com/articles/cs-127 [15 August 2022].

Mohamed Ramli, N., Cruz, C.R. de & Prapti, D.R. 2024. The importance of ammonia removal process in aquaculture systems — towards improving Malaysian aquaculture production. *Advances in Agricultural and Food Research Journal*, 5(1): 1–26.

Mohammed, A.A., Ambak, K., Mosa, A.M. & Syamsunur, D. 2019. Expert system in engineering transportation: A Review. *Journal of Engineering Science and Technology*, 14(1): 229–252.

Mohd Ali, M.F., Asrul Ibrahim, A. & Mohd Zaman, M.H. 2021. Optimal sizing of solar panel and battery storage for a smart aquaponic system. In the *2021 IEEE 19th Student Conference on Research and Development (SCOReD)*. Malaysia: IEEE: 186–191.

Mohseni, S., Zarei, N. & Ragan, E.D. 2021. A multidisciplinary survey and framework for design and evaluation of Explainable AI systems. *ACM Transactions on Interactive Intelligent Systems*, 11(3–4): 1–45.

Mokhtar, A., El-Ssawy, W., He, H., Al-Anasari, N., Sammen, S.S., Gyasi-Agyei, Y. & Abuarab, M. 2022. Using machine learning models to predict hydroponically grown lettuce yield. *Frontiers in Plant Science*, 13: 1–10.

Molin, S. 2021. *Hands-on data analysis with Pandas: A Python data science handbook for data collection, wrangling, analysis, and visualization*. 2nd ed. Birmingham, UK: Packt Publishing.

Morchid, A., El Alami, R., Raezah, A.A. & Sabbar, Y. 2024. Applications of Internet of Things (IoT) and sensors technology to increase food security and agricultural sustainability: benefits and challenges. *Ain Shams Engineering Journal*, 15(3): 1–15. https://doi.org/10.1016/j.asej.2023.102509.

Mori, J., Erickson, K. & Smith, R.L. 2021. Predictive modeling of pH in an aquaponics system using Bayesian and Non-Bayesian linear regression to inform system maintenance. *ACS Agricultural Science and Technology*, 1(4): 400–406.

Mouha, R.A. 2021. Internet of Things (IoT). *Journal of Data Analysis and Information Processing*, 9(2): 77–101.

Mufid, M.R., Basofi, A., Al Rasyid, M.U.H., Rochimansyah, I.F. & Rokhim, A. 2019. Design an MVC model using Python for Flask framework development. *International Electronics Symposium*

(IES): 214–219.

Müller, A.C. & Guido, S. 2016. *Introduction to machine learning with Python: a guide for data scientists*. Sebastopol, CA. O'Reilly Media, Inc.

Murdan, A.P. & Joyram, A. 2021. An IoT-based solar powered aquaponics system. In *2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*. IEEE: 1–6.

Muruganantham, P., Wibowo, S., Grandhi, S., Samrat, N.H. & Islam, N. 2022. A systematic literature review on crop yield prediction with deep learning and remote sensing. *Remote Sensing*, 14(9): 1–21.

Naeem, S., Ali, A., Anam, S. & Ahmed, M.M. 2023. An unsupervised machine learning algorithms: comprehensive review. *International Journal of Computing and Digital Systems*, 13(1): 911–921.

Nagori, V. & Trivedi, B. 2014. Types of expert system: comparative study. *Asian Journal of Computer and Information System*, 02(02). www.ajouronline.com [7 September 2022].

Naimi, A.I. & Balzer, L.B. 2018. Stacked generalization: an introduction to super learning. *European Journal of Epidemiology*, 33(5): 459–464. https://link.springer.com/article/10.1007/s10654-018-0390-z [15 July 2022].

Nair, C.S., Manoharan, R., Nishanth, D., Subramanian, R., Neumann, E. & Jaleel, A. 2025. Recent advancements in aquaponics with special emphasis on its sustainability. *Journal of the World Aquaculture Society*, 56(1): 1–39.

Nandy, A. & Biswas, M. 2018. *Reinforcement learning: with Open AI, Tensorflow and Keras using Python*. Berkeley, CA. Apress. http://link.springer.com/10.1007/978-1-4842-3285-9 [14 June 2022].

Naputol, A.T., Ituriaga, J.I., Yu Jeco-Espaldon, B.M., Guanzon, G., Co, K. & Giner, M.G. 2024. Development of an IoT-based aquaponics monitoring system with automated feeder. In the *2024 IEEE Symposium on Industrial Electronics and Applications (ISIEA)*. IEEE: 1–6.

Naresh, R., Jadav, S.K., Singh, M., Patel, A., Singh, B., Beese, S. & Pandey, S.K. 2024. Role of hydroponics in improving water-use efficiency and food security. *International Journal of Environment and Climate Change*, 14(2): 608–633.

Naser, B.A.A.-Z., Saleem, A.L., Ali, A.H., Alabassi, S. & Al-Baghdadi, M.A.R.S. 2019. Design and construction of smart IoT-based aquaponics powered by PV cells. *International Journal of Energy and Environment(IJEE)*, 10(3): 127–134. www.IJEE.IEEFoundation.org [22 October 2021].

Ni'mah, S.J., Murjazin, M., Nurhuda, A., Lathif, N.M. & Al Fajri, M. 2024. Ontology, epistemology, and axiology of islamic educational philosophy: an introduction. *Matan : Journal of Islam and Muslim Society*, 6(1): 33–43.

Nichani, A., Saha, S., Upadhyay, T., Ramya, A. & Tolia, M. 2018. Data acquisition and actuation for aquaponics using IoT. In the *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT)*. Bangalore, India: IEEE: 46–51.

Nielsen, M.A. 2015. *Neural networks and deep learning*. San Francisco, CA, USA: Determination Press.

Ntulo, M.P., Owolawi, P.A., Mapayi, T., Malele, V., Aiyetoro, G. & Ojo, J.S. 2021. IoT-based smart aquaponics system using Arduino Uno. In the *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*. IEEE: 1–6.

Obaido, G., Mienye, I.D., Egbelowo, O.F., Emmanuel, I.D., Ogunleye, A., Ogbuokiri, B., Mienye, P. & Aruleba, K. 2024. Supervised machine learning in drug discovery and development:

algorithms, applications, challenges, and prospects. *Machine Learning with Applications*, 17: 1–20.

Obiora, C.N., Ali, A. & Hasan, A.N. 2023. Using the Multilayer Perceptron (MLP) model in predicting the patterns of solar irradiance at several time intervals. In *Proceedings of the 31st Southern African Universities Power Engineering Conference (SAUPEC)*. IEEE: 1–6.

Obirikorang, K.A., Sekey, W., Gyampoh, B.A. & Ashiagbor, G. 2021. Aquaponics for improved food security in Africa: a review. *Frontiers in Sustainable Food Systems*, 5: 1–10. www.frontiersin.org [28 October 2021].

Ozaga, A., Nugraheni, M. & Widodo. 2024. Comparative study of logistic regression and K-Nearest Neighbor for heart disease prediction. In 7th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI). IEEE: 679–683.

Ogidan, E., Dimililer, K. & Ever, Y.K. 2019. Machine learning for expert systems in data analysis. In the *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*. IEEE: 1–5.

Olafadehan, O.A. & Ahaotu, I.D. 2023. Comparative analysis of machine learning algorithms in predicting rate of penetration during drilling. *Journal of Petroleum & Chemical*, 1(1): 1–16.

Ong, Z.J., Ng, A.K. & Kyaw, T.Y. 2019. Intelligent outdoor aquaponics with automated grow lights and Internet of Things. In the *2019 IEEE International Conference on Mechatronics and Automation (ICMA)*. Tianjin, China: IEEE: 1778–1783.

Oscar, C., Bruneau, P., Sottet, J.S. & Torregrossa, D. 2023. Landscape of high-performance Python to develop data science and machine learning applications. ACM Computing Surveys, 56(3): 1–30.

Owusu, R., Mayoko, J.C. & Lee, J.Y. 2024. Prediction of water quality temperature in the growth pattern of fish (Nile tilapia) and plant (lettuce) in a prototype-safe, automated aquaponics environment using deep learning. *Open Access Library Journal*, 11(9): 1–22.

Palm, H.W., Knaus, U., Appelbaum, S., Strauch, S.M. & Kotzen, B. 2019. Coupled aquaponics systems. In Goddek, S., Joyce, A., Kotzen, B., & Burnell, G. M.(eds). *Aquaponics Food Production Systems*. Cham, Switzerland: Springer International Publishing,163–200.

Pan, B. 2018. Application of XGBoost algorithm in hourly PM2.5 concentration prediction. In *IOP Conference Series: Earth and Environmental Science*. Bristol, UK: Institute of Physics Publishing, 1–7.

Pandey, R., Paul, V., Das, M., Meena, M. & Meena, R.C. 2017. Plant growth analysis. *Manual of ICAR Sponsored Training Programme on "Physiological Techniques to Analyze the Impact of Climate Change on Crop Plants"*: 103–107.

Panigrahi, B., Kathala, K.C.R. & Sujatha, M. 2023. A machine learning-based comparative approach to predict the crop yield using supervised learning with regression rodels. Procedia Computer Science, 218: 2684–2693.

Papazoglou, D., Devrelis, S., Loufakis, M., Symeonidis, P., Ioannidis, D., Tzovaras, D., Oikonomou, G. & Kourmpetis, I. 2024. Web-based decision support system for remote weather radar maintenance: design and implementation. In the *2024 7th International Conference on Artificial Intelligence and Big Data (ICAIBD)*. IEEE: 34–41.

Pappu, S., Vudatha, P., Niharika, A.V., Sankaranarayanan, S. & Karthick, T. 2017. Intelligent IoT-based water quality monitoring system. *International Journal of Applied Engineering Research*, 12: 5447–5454. http://www.ripublication.com [26 October 2021].

Parsa, A.B., Movahedi, A., Taghipour, H., Derrible, S. & Mohammadian, A.K. 2020. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accident Analysis & Prevention*, 136: 1–17.

Pasha, A.K., Mulyana, E., Hidayat, C., Ramdhani, M.A., Kurahman, O.T. & Adhipradana, M. 2018. System design of controlling and monitoring on aquaponic based on Internet of Things. In *2018 4th International Conference on Wireless and Telematics (ICWT)*. Nusa Dua, Bali, Indonesia: IEEE: 1–5.

Pechlivani, E.M., Gkogkos, G., Christakakis, P., Kapetas, D., Hadjigeorgiou, I., Ioannidis, D. & Tzovaras, D. 2025. Aquaponic Farming with Advanced Decision Support System: Practical Implementation and Evaluation. In *2025 IEEE 6th International Conference on Image Processing, Applications and Systems (IPAS)*. Lyon, France: IEEE: 1–7.

Perumal, B., Shyam, S.N., Sakthi Viswanath, S., Mohamed Esa, M., Abdul Fatthah, P.M. & Rajesh, V. 2024. Smart aquaponics oversight: IoT sensors and mobile app empowered monitoring. In *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*. IEEE: 1–4.

Petrea, Ștefan M., Simionov, I.A., Antache, A., Nica, A., Oprica, L., Miron, A., Zamfir, C.G., Neculiță, M., Dima, M.F. & Cristea, D.S. 2023. An analytical framework on utilizing various integrated multi-trophic scenarios for basil production. *Plants*, 12(3): 1–50.

Petrea, Ş.M., Cristea, V., Dediu, L., Contoman, M. & Antache, A. 2013. A study of nitrogen cycle in an integrated aquaponic system with different plant densities. *Bulletin UASVM Animal Science and Biotechnologies*, 70(1): 55–64.

Phillips-Wren, G. 2017. Intelligent systems to support human decision making. In *Artificial Intelligence: Concepts, Methodologies, Tools, and Applications*. USA: IGI Global, 3023–3036.

Phillips, P.J., Hahn, C.A., Yates, A.N., Greene, KristenFontana, P.C., Broniatowski, D.A. & Przybocki, M.A. 2020. Four principles of Explainable Artificial Intelligence. *National Institute of Standards and Technology Interagency or Internal Report*: 1–36. https://doi.org/10.6028/NIST.IR.8312-draft.

Pillay, T.V.R. 2004. *Aquaculture and the environment*. 2nd ed. Oxford, United Kingdom: Blackwell Publishing Ltd.

Power, D.J. 2007. A brief history of decision support systems. *DSSResources.COM*. https://www.dssresources.com/history/dsshistory.html [29 October 2021].

Power, D.J. 2002. Categorizing decision support systems: a multidimensional approach. In Mora, M., Forgionne, G. A., & Gupta, J. N. D.(eds.). *Decision making support systems: achievements, trends, and challenges for the new decade*. United States: IGI Global.

Power, D.J. 2008. Decision support systems: a historical overview. In *Handbook on decision support systems 1*. Berlin, Heidelberg. Springer: 121–140. https://link.springer.com/chapter/10.1007/978-3-540-48713-5_7 [12 January 2022].

Prabha, R., Saranish, R.S., Sowndharya, S., Santhosh, A.C., Varsha, R. & Sumathi, K. 2020. IoT-controlled aquaponic system. In the *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. Coimbatore, India: Institute of Electrical and Electronics Engineers Inc.: 376–379.

Prajwala, T.R. 2015. A comparative study on Decision Tree and Random Forest using R Tool. *International Journal of Advanced Research in Computer and Communication Engineering*, 4(1): 196–199.

Prasetya, N.W., Anto, Y., Azis, B. & Budi, M.H.S. 2024. Adaptive IoT-based monitoring system in Edge Network for aquaponics ecosystem. In the *2024 3rd International Conference on Digital Transformation and Applications (ICDXA)*. IEEE: 1–5.

Prathanrat, P. & Polprasert, C. 2018. Performance prediction of Jupyter Notebook in JupyterHub using machine learning. In the *2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*. IEEE: 157–162.

Prianka, D., Shilpa, G. V, Parvin, H., Prerana, B. & Gauda, R. 2024. Wick-based hydroponics system with mobile application. In the *2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)*. IEEE: 1–5.

Prince, B. & Prince, D. 2018. Introduction to the Intelligent Internet of Things. In *Memories for the Intelligent Internet of Things*. Texas, USA: John Wiley & Sons, Ltd: 1–2.

Priya, K.S. 2021. Linear regression algorithm in machine learning through MATLAB. *International Journal for Research in Applied Science and Engineering Technology*, 9(12): 989–995.

Pruneski, J.A., Williams, R.J., Nwachukwu, B.U., Ramkumar, P.N., Kiapour, A.M., Martin, R.K., Karlsson, J. & Pareek, A. 2022. The development and deployment of machine learning models. Knee Surgery, Sports Traumatology, Arthroscopy, 30(12): 3917–3923. https://doi.org/10.1007/s00167-022-07155-4.

Pu’ad, M.F.M., Azami Sidek, K. & Mel, M. 2020. IoT-based water quality monitoring system for aquaponics. In *Journal of Physics: Conference Series*. Melaka, Malaysia: IOP Publishing: 1–8. https://iopscience.iop.org/article/10.1088/1742-6596/1502/1/012020 [2 June 2023].

Qadeer, A., Butt, S.J., Asam, H.M., Mehmood, T., Nawaz, M.K. & Haidree, S.R. 2020. Hydroponics as an innovative technique for lettuce production in greenhouse environment. *Pure and Applied Biology*, 9(1): 20–26.

Quinlan, J.R. 1996. Learning Decision Tree classifiers. *ACM Computing Surveys (CSUR), 28(1), pp.71-72.*, 28(1): 71–72.

Rahayu, L.P., Kindhi, B. Al, Pradika, C.D., Adhim, F.I., Priananda, C.W., Musthofa, A., Indasyah, E. & Istiqomah, F. 2021. Design of pH control system and water recirculation in aquaponic cultivation using Mamdani fuzzy logic control. In the *2021 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation(ICAMIMIA)*. Surabaya, Indonesia: IEEE: 287–292.

Rajabi, M., Hossani, S. & Dehghani, F. 2019. A literature review on current approaches and applications of fuzzy expert systems. *arXiv preprint arXiv:1909.08794*: 1–13. https://arxiv.org/abs/1909.08794v1 [20 September 2022].

Rajaseger, G., Chan, K.L., Tan, K.Y., Ramasamy, S., Khin, M.C., Amaladoss, A. & Haribhai, P.K. 2023. Hydroponics: current trends in sustainable crop production. *Bioinformation*, 19(9): 925–938.

Rajendran, S., Domalachenpa, T., Arora, H., Li, P., Sharma, A. & Rajauria, G. 2024. Hydroponics: exploring innovative sustainable technologies and applications across crop production, with emphasis on potato mini-tuber cultivation. *Heliyon*, 10(5): 1–14. https://doi.org/10.1016/j.heliyon.2024.e26823.

Rakocy, J.E., Masser, M.P. & Thomas, M.L. 2006. Recirculating aquaculture tank production systems: aquaponics-integrating fish and plant culture. *SRAC Publication*, 454: 1–16.

Raman, R. & Vasmatkar, A. 2024. IoT-enabled aquaponics and hydroponics for efficient indoor farming systems with cloud computing. In the *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*. IEEE: 1–5.

Ramirez, C.R. 2024. Integrating IoT for optimized water quality in aquaponics_ innovations in aquaculture management. *Applied Physics, Material Sciences, and Engineering Journal*, 12(3): 1–6.

Ramraj, S., Uzir, N., Raman, S. & Banerjee, S. 2016. Experimenting XGBoost algorithm for prediction and classification of different datasets. *International Journal of Control Theory and Applications*, 9(40): 651–662. https://www.researchgate.net/publication/318132203 [16 August 2022].

Rao, N.J., Vineetha, N., Dinesh, D. & Sree, D.K. 2024. Multilingual web-based agricultural decision support system: integrating machine learning for accurate vegetable price prediction. In the *2024 International Conference on Inventive Computation Technologies (ICICT)*. IEEE: 1001–1007.

Rasamoelina, A.D., Adjailia, F. & Sinčák, P. 2020. A review of activation function for artificial neural network. In the *2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*. IEEE: 281–286.

Ray, S. 2019. A quick review of machine learning algorithms. In *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Perspectives and Prospects (COMITCon)*. IEEE: 35–39.

Reddy, K.S.P., Roopa, Y.M., Kovvada Rajeev, L.N. & Nandan, N.S. 2020. IoT-based smart agriculture using machine learning. *Proceedings of the 2nd International Conference on Inventive Research in Computing Applications, ICIRCA 2020*: 130–134.

Reddy, P., Sohel, M. & Naik, S.M. 2024. An IoT-based smart hydroponics - plant automation bot. In the *2024 ASU International Conference on Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)*. IEEE: 1203–1207.

Ren, Q., Zhang, L., Wei, Y. & Li, D. 2018. A method for predicting dissolved oxygen in aquaculture water in an aquaponics system. *Computers and Electronics in Agriculture*, 151: 384–391.

Ren, Z., Wang, S. & Zhang, Y. 2023. Weakly supervised machine learning. *CAAI Transactions on Intelligence Technology*, 8(3): 549–580.

Resh, H.M. 2013. *Hydroponic food production: a definitive guidebook for the advanced home gardener and the commercial hydroponic grower*. 7th ed. Boca Raton: CRC Press.

Reyes-Yanes, A., Abbasi, R., Martinez, P. & Ahmad, R. 2022. Digital twinning of hydroponic grow beds in intelligent aquaponic systems. *Sensors*, 22(19): 1–18.

Reyes-Yanes, A., Martinez, P. & Ahmad, R. 2020. Real-time growth rate and fresh weight estimation for little gem romaine lettuce in aquaponic grow beds. *Computers and Electronics in Agriculture*, 179: 1–10.

Riansyah, A., Mardiati, R., Effendi, M.R. & Ismail, N. 2020. Fish feeding automation and aquaponics monitoring system based on IoT. In the *2020 6th International Conference on Wireless and Telematics (ICWT)*. Yogyakarta, Indonesia: IEEE: 1–4.

Richardson, C., Shah, S. & Yuan, X. 2022. Semi-supervised machine learning for analyzing COVID-19-related Twitter data for asian hate speech. In the *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE: 1643–1648.

Roberto, K. 2003. *How-to-hydroponics*. 4th ed. Farmingdale, New York: Future Garden, Inc.

Rokach, L. & Maimon, O. 2005. Decision Trees. In Rokach, L. & Maimon, O.(eds.). *Data mining and knowledge discovery handbook*. Boston, MA: Springer, 165–192.

Rothman, D. 2020. *Hands-on Explainable AI (XAI) with Python : interpret, visualize, explain, and integrate reliable AI for fair, secure, and trustworthy AI apps*. Birmingham, UK: Packt Publishing Ltd.

Roustaei, N. 2024. Application and interpretation of linear-regression analysis. *Medical Hypothesis, Discovery, and Innovation in Ophthalmology*, 13(3): 151–159.

Rozie, F., Syarif, I. & Al Rasyid, M.U.H. 2020. Design and implementation of intelligent aquaponics monitoring system based on IoT. In the *2020 International Electronics Symposium (IES)*. IEEE: 534–540.

Ru, D., Liu, J., Hu, Z., Zou, Y., Jiang, L., Cheng, X. & Lv, Z. 2017. Improvement of aquaponic performance through micro- and macro-nutrient addition. *Environmental Science and*

Pollution Research, 24(19): 16328–16335.

Russell, R. 2018. *Machine learning step-by-step guide to implement machine learning algorithms with Python*. Scotts Valley, CA, USA: CreateSpace Independent Publishing Platform.

Saaid, M.F., Yahya, N.A.M., Noor, M.Z.H. & Ali, M.S.A.M. 2013. A development of an automatic microcontroller system for Deep Water Culture (DWC). In the *2013 IEEE 9th International Colloquium on Signal Processing and Its Applications*. Kuala Lumpur, Malaysia: IEEE: 328–332.

Saha, S., Rajib, R.H. & Kabir, S. 2018. IoT-based automated fish farm aquaculture monitoring system. In the *2018 International Conference on Innovations in Science, Engineering and Technology (ICISET)*. Chittagong, Bangladesh: IEEE: 201–206.

Saini, M.K. & Saini, R.K. 2020. Agriculture monitoring and prediction using Internet of Things (IoT). In the *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*. IEEE: 53–56.

Sallenave, R. 2016. *Important water quality parameters in aquaponics systems*. Las Cruces, NM, USA: College of Agricultural, Consumer and Environmental Sciences: 1-8.

Sarala, B., Rithika, A.C., Praneetha, K., Nivetha, S., Jeevithae, K. & Ilangovanf, K. 2021. Building social bot with Python and Tweepy. *Turkish Online Journal of Qualitative Inquiry (TOJQI)*, 12(9): 7039–7046.

Saranya, A. & Subhashini, R. 2023. A systematic review of Explainable Artificial Intelligence models and applications: recent developments and future trends. *Decision Analytics Journal*, 7: 1–14. https://doi.org/10.1016/j.dajour.2023.100230.

Sarker, I.H. 2021. Machine learning: algorithms, real-world applications and research directions. *SN Computer Science*, 2(3): 1–21. https://doi.org/10.1007/s42979-021-00592-x.

Saunders, M., Lewis, P.T. & Thornhill, A. 2019. *Research methods for business students*. 8th ed. Harlow, UK: Pearson Education.

Schapire, R.E. 1999. A brief introduction to boosting. In the *Sixteenth International Joint Conference on Artificial Intelligence.* 1401–1406. www.research.att.com/schapire [11 August 2022].

Schapire, R.E. 2003. The boosting approach to machine learning: an overview. In Denison, D. D., Hansen, M. H., Holmes, C. C., Mallick, B., & Yu, B.(eds.). *Nonlinear Estimation and Classification. Lecture Notes in Statistics*. Springer, New York, NY: 149–171. https://link.springer.com/chapter/10.1007/978-0-387-21579-2_9 [11 August 2022].

Schatsky, D., Kumar, N. & Bumb, S. 2017. Signals for strategists Intelligent IoT: bringing the power of AI to the Internet of Things: 1–8.

Schwarz, M. 1995. *Soilless culture management*. Berlin, Heidelberg: Springer.

Seber, G.A. & Lee, A.J. 2012. *Linear regression analysis*. 2nd ed. Hoboken, New Jersey: John Wiley & Sons.

Seegobin, N., Abdalla, Y., Li, G., Murdan, S., Shorthouse, D. & Basit, A.W. 2024. Optimising the production of PLGA nanoparticles by combining design of experiment and machine learning. *International Journal of Pharmaceutics*, 667: 1–8.

Senapaty, M.K., Ray, A. & Padhy, N. 2024. A decision support system for crop recommendation using machine learning classification algorithms. *Agriculture*, 14(8): 1–40.

Setemen, K., Erawati Dewi, L.J. & Purnamawan, I.K. 2019. PAON usability testing using system usability scale. In *Journal of Physics: Conference Series*. IOP Publishing: 1–5.

Seyghaly, R., Garcia, J., Masip-Bruin, X. & Kuljanin, J. 2024. SBNNR: Small-Size Bat-Optimized KNN Regression. Future Internet, 16(11): 1–20.

Shafeena, T. 2016. Smart Aquaponics system: challenges and opportunities. *European Journal of Advances in Engineering and Technology*, 3(2): 52–55. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1058.803&rep=rep1&type=pdf [28 October 2021].

Shahzadi, R., Ferzund, J., Tausif, M. & Suryani, M.A. 2016. Internet of Things-based expert system for smart agriculture. *International Journal of Advanced Computer Science and Applications*, 7(9): 341–350. www.ijacsa.thesai.org [7 September 2022].

Shaout, A. & Scott, S.G. 2017. IoT Fuzzy Logic aquaponics monitoring and control hardware real-time system. In *Proceedings of the International Arab Conference on Information Technology*. Yassmine Hammamet, Tunisia: 22–24. https://acit2k.org/ACIT/images/stories/year2014/month1/ACIT2017_Proceeding/139.pdf [29 October 2021].

Shapley, L.S. 1953. A value for n-person games. In *Games, Contributions to the Theory of 2*. 1–12.

Sharma, H. & Kumar, S. 2016. A survey on decision tree algorithms of classification in data mining. *Article in International Journal of Science and Research(IJSR)*, 5(4): 2094–2097.

Sharma, N., Acharya, S., Kumar, K., Singh, N. & Chaurasia, O.P. 2018. Hydroponics as an advanced technique for vegetable production: an overview. *Journal of Soil and Water Conservation*, 17(4): 364–371.

Shrestha, A. & Dunn, B. 2010. Hydroponics. *Oklahoma Cooperative Extension Service*: 1–4.

Shrestha, A. & Mahmood, A. 2019. Review of deep learning algorithms and architectures. *IEEE Access*, 7: 53040–53065.

Silalahi, A.O., Sinambela, A., Pardosi, J.T.N. & Panggabean, H.M. 2022. Automated water quality monitoring system for aquaponic pond using LoRa TTGO SX1276 and Cayenne Platform. In the *2022 IEEE International Conference on Computer Science and Information Technology (ICOSNIKOM)*. Laguboti, North Sumatra, Indonesia: IEEE: 1–6.

Singh, J., Narayan, A., Rafique, I.M., Kumar, T., Sarma, K., Ahirwal, S. & Bharati, V. 2021. Aquaponics: Toward a sustainable multi-trophic production of fish and vegetables. *Food and Scientific Reports*, 2(6): 46–49.

Sinha, A., Pokhriyal, K. & Kumar, S. 2024. A decision support system for edge computing intelligence-driven insights on natural disasters. In the *2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT)*. IEEE: 1–6.

Smith, B. 2012. Ontology. In Hurtado. G. & Nudler. O. (eds.). *The furniture of the world: Essays in ontology and metaphysics.* Amsterdam: Brill, 47–68.

Snyder, H. 2019. Literature review as a research methodology: an overview and guidelines. *Journal of Business Research*, 104: 333–339.

Somerville, C., Cohen, M., Pantanella, E., Stankus, A. & Lovatelli, A. 2014. Small-scale aquaponic food production: integrated fish and plant farming. *FAO Fisheries and Aquaculture Technical Paper*, (589): 1–262.

Song, Y.Y. & Lu, Y. 2015. Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27(2): 130–135.

Souha, A., Benaddi, L., Ouaddi, C., Bouziane, E.M. & Jakimi, A. 2024. Towards a decision support system for tourism sector in Daraa-Tafilalet. In the *2024 4th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*. FEZ, Morocco: IEEE: 01–05.

Sprague Jr, R.H. 1980. A framework for the development of decision support systems. *MIS quarterly*, 4(4): 1–26.

Sridevi, A., Devadharshini, K., Deepika, B., Elakkiya, K. V. & Bharathi, G. 2024. Smart aquaponics and hydroponics monitoring using IoT. In the *2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications, Theme: Healthcare and Internet of Things (AIMLA)*. IEEE: 1–6.

Statistics South Africa. 2017. *Poverty on the rise in South Africa.* http://www.statssa.gov.za/?p=10334 [26 October 2021].

Stouffer, J. 2015. *Mastering Flask : Gain expertise in Flask to create dynamic and powerful web applications*. Birmingham, UK: Packt Publishing Ltd.

Subakti, M.R., Nurhayati, N. & Rahayu, M.S. 2022. The effect of concentration of AB mix and ZPT solutions on the growth and production of mustard plants (Brassica juncea L.) in hydroponic wick systems. In *E3S Web of Conferences*. EDP Sciences: 1–7. https://doi.org/10.1051/e3sconf/202233901010.

Sudheer, A., Baranwal, I. & Gadhoke, B. 2022. KNN-Based Hybrid models for geolocation-based problems. In the *2022 2nd International Conference on Intelligent Technologies (CONIT)*. IEEE: 1–6.

Sunardi, A., Suud, F.I., Woro Agus, N. & Gunawan, I. 2021. IoT application on aquaponics system energy optimization. In *Journal of Physics: Conference Series*. IOP Publishing: 1–9. https://iopscience.iop.org/article/10.1088/1742-6596/1772/1/012046 [2 June 2023].

Suryawanshi, A. 2021. Smart factory using industrial Internet of Things. *Journal of Emerging Technologies and innovative Research(JETIR)*, 8(9): 338–341.

Sutton, R.S. & Barto, A.G. 2018. *Reinforcement learning: an introduction*. 2nd ed. Cambridge, MA: The MIT Press. https://books.google.co.za/books?hl=en&lr=&id=uWV0DwAAQBAJ&oi=fnd&pg=PR7&dq=reinforcement+learning&ots=misJo5Z3p4&sig=jrS39qidcOjcvtj1nI1upLnc_bM#v=onepage&q=reinforcement learning&f=false [14 June 2022].

Taha, M.F., Abdalla, A., ElMasry, G., Gouda, M., Zhou, L., Zhao, N., Liang, N., Niu, Z., Hassanein, A., Al-Rejaie, S., He, Y. & Qui, Z. 2022. Using deep convolutional neural network for image-based diagnosis of nutrient deficiencies in plants grown in aquaponics. *Chemosensors*, 10(2): 1–23.

Taha, M.F., ElManawy, A.I., Alshallash, K.S., ElMasry, G., Alharbi, K., Zhou, L., Liang, N. & Qiu, Z. 2022. Using machine learning for nutrient content detection of aquaponics-grown plants based on spectral data. *Sustainability*, 14(19): 1–17.

Taherdoost, H. 2021. Data collection methods and tools for research: a step-by-step guide to choose data collection technique for academic and business research projects. *International Journal of Academic Research in Management (IJARM)*, 10(1): 10–38. https://hal.science/hal-03741847.

Tai, M.C.-T. 2020. The impact of artificial intelligence on human society and bioethics. *Tzu-Chi Medical Journal*, 32(4): 339–343. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7605294/ [26 October 2021].

Taji, K., Sohail, A., Ghanimi, F., Ghanimi, I., ilyas, S. & Ahmad, Y.T. 2023. A systematic literature review of computational studies in aquaponic system literature review of computational studies in aquaponic system. *International Journal of Advanced Computer Science and Applications*, 14(9): 333–343.

Takami, G., Tokuoka, M., Goto, H. & Nozaka, Y. 2016. Machine learning applied to sensor data analysis. *Yokogawa Technical Report English Edition*, 59(1): 27–30.

Tan, X.J., Cheor, W.L., Yeo, K.S. & Leow, W.Z. 2022. Expert systems in oil palm precision agriculture: a decade systematic review. *Journal of King Saud University - Computer and Information Sciences*, 34(4): 1569–1594.

Taud, H. & Mas, J.F. 2018. Multilayer Perceptron (MLP). In Camacho Olmedo, M., Paegelow, M., Mas, J., & Escobar, F.(eds.). *Geomatic approaches for modeling land change scenarios*. Cham: Springer, 451–455.

Taunk, K., De, S., Verma, S. & Swetapadma, A. 2019. A brief review of nearest neighbor algorithm for learning and classification. In the *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*. IEEE: 1255–1260.

Thakur, K., Kuthiala, T., Singh, G., Arya, S.K., Iwai, C.B., Ravindran, B., Khoo, K.S., Chang, S.W. & Awasthi, M.K. 2023. An alternative approach towards nitrification and bioremediation of wastewater from aquaponics using biofilm-based bioreactors: a review. *Chemosphere*, 316: 1–14. https://linkinghub.elsevier.com/retrieve/pii/S0045653523001157 [17 April 2023].

Thakur, P., Wadhwa, H., Kaushal, S. & Shubham. 2023. Nutrient dynamics for hydroponic production system. *International Journal of Plant & Soil Science*, 35(21): 982–993.

Thomas, R.N. & Gupta, R. 2020. Ensemble learning techniques and its efficiency in machine learning: a survey. In the *2nd International Conference on Data, Engineering and Applications (IDEA)*. IEEE: 1–6.

Thorarinsdottir, R.I., Kledal, P.R., Skar, S.L.G., Sustaeta, F., Mankasingh, U., Pantanella, E., Ven, R. van de & Shultz, C. 2015. *Aquaponics guidelines*. Thorarinsdottir, R. I.(ed.). Reykjavik, Iceland: Haskolaprent.

Timmons, M.B. & Ebeling, J.M. 2010. *Recirculating aquaculture*. 2nd ed. Itacha, USA: Cayuga Aqua Ventures.

Tolentino, L.K.S., Fernandez, E.O., Jorda, R.L., Amora, S.N.D., Bartolata, D.K.T., Sarucam, J.R. V, Sobrepena, J.C.L. & Sombol, K.Y.P. 2019. Development of an IoT-based aquaponics monitoring and correction system with temperature-controlled greenhouse. In the *2019 International SoC Design Conference (ISOCC)*. Jeju, Korea (South): IEEE: 261–262.

Traister, J. 2018. Bad effects of chemicals in our food. https://healthfully.com/bad-effects-of-chemicals-in-our-food-8069859.html [26 October 2021].

Trejo-Téllez, L.I. & Gómez-Merino, F.C. 2012. Nutrient solutions for hydroponic systems. In Asao, T.(ed.). *Hydroponics: a standard methodology for plant biological researches*. Rijeka, Croatia: InTech, 1–12.

Tripathi, K.P. 2011. Decision Support System is a tool for making better decisions in the organization. *Indian Journal of Computer Science and Engineering (IJCSE)*, 2: 112–117.

Turnsek, M., Joly, A., Thorarinsdottir, R. & Junge, R. 2020. Challenges of commercial aquaponics in Europe: beyond the hype. *Water*, 12(1): 1–18.

Ubayasena, W.P.H., Hemantha, N.S.C., Wijayakoon, W.M.T.B., Kavindya, R.M.N., Kalapuge, V. & Samarasekara, P. 2023. Smart aquaponics system. *International Research Journal of Innovations in Engineering and Technology (IRJIET)*, 07(11): 693–707.

Udanor, C.N., Ossai, N.I., Nweke, E.O., Ogbuokiri, B.O., Eneh, A.H., Ugwuishiwu, C.H., Aneke, S.O., Ezuwgu, A.O., Ugwoke, P.O. & Christiana, A. 2022. An Internet of Things-labelled dataset for aquaponics fish pond water quality monitoring system. *Data in Brief*, 43: 1–6.

Ugwu, C.I., Ekere, J.N. & Onoh, C. 2021. Research paradigms and methodological choices in the research process. *Journal of Applied Information Science and Technology*, 14(2): 116–124.

Valiente, F.L., Garcia, R.G., Domingo, E.J.A., Estante, S.M.T., Ochaves, E.J.L., Villanueva, J.C.C. & Balbin, J.R. 2018. Internet of Things (IoT)-based mobile application for monitoring of automated aquaponics system. In the *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*. Baguio: IEEE: 1–6.

Varkey, K. Ben, Ebenezer, V., Bijolin Edwin, E., Jeyaraj, K.A., Varghese, M. & Roshni Thanka, M.

2021. Smart aquaponics: challenges and opportunities. *Annals of the Romanian Society for Cell Biology*, 25: 9715–9722.

Vashi, S., Ram, J., Modi, J., Verma, S. & Prakash, C. 2017. Internet of Things (IoT): a vision, architectural elements, and security issues. In the *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*. IEEE: 492–496.

Veazie, P., Pandey, P., Young, S., Ballance, M.S., Hicks, K. & Whipker, B. 2022. Impact of macronutrient fertility on mineral uptake and growth of Lactuca Sativa 'Salanova Green' in a hydroponic system. *Horticulturae*, 8(11): 1–13.

Velazquez-Gonzalez, R.S., Garcia-Garcia, A.L., Ventura-Zapata, E., Barceinas-Sanchez, J.D.O. & Sosa-Savedra, J.C. 2022. A review on hydroponics and the technologies associated for medium and small-scale operations. *Agriculture (Switzerland)*, 12(5): 1–21.

Verma, D.K., Satyaveer, N.K.M., Kumar, P. & Jayaswa, R. 2022. Important water quality parameters in aquaculture: an overview. *Agriculture and Environment*, 3(3): 24–29.

Vernandhes, W., Salahuddin, N.S., Kowanda, A. & Sari, S.P. 2017. Smart aquaponic with monitoring and control system based on IoT. In the *2017 Second International Conference on Informatics and Computing (ICIC)*. IEEE: 1–6.

Villanueva, J.C.C., Ballado, A.H. & Estember, R.D. 2022. Effects of pH level, temperature in fish tank, and temperature in grow bed on the growth of Nile tilapia and romaine lettuce in an automated aquaponics system. In *Proceedings of the 3rd Asia Pacific International Conference on Industrial Engineering and Operations Management*. 4594–4605.

Visvizi, A., Jussila, J., Lytras, M.D. & Ijäs, M. 2020. Tweeting and mining OECD-related microcontent in the post-truth era: a cloud-based app. *Computers in Human Behavior*, 107: 1–7. https://doi.org/10.1016/j.chb.2019.03.022.

Vui, C.S., Soon, G.K., On, C.K., Alfred, R. & Anthony, P. 2013. A review of stock market prediction with artificial neural network (ANN). In the *2013 IEEE International Conference on Control Systems, Computing and Engineering*. IEEE: 477–482.

Walczak, S. 2018. Artificial neural networks. In *Encyclopedia of Information Science and Technology*. IGI Global Scientific Publishing: 120–131. https://www.igi-global.com/chapter/artificial-neural-networks/183727 [23 July 2022].

Wan, S., Zhao, K., Lu, Z., Li, J., Lu, T. & Wang, H. 2022. A modularized IoT monitoring system with edge computing for aquaponics. *Sensors*, 22(23): 1–18.

Wang, H. & Raj, B. 2017. On the origin of deep learning. *arXiv preprint arXiv:1702.07800*: 1–72.

Wang, J., Li, L. & Zeller, A. 2021. Restoring execution environments of Jupyter notebooks. In the *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*. Madrid, ES: IEEE: 1622–1633.

Wang, S.C., Lin, W.L. & Hsieh, C.H. 2020. To improve the production of agriculture using IoT-based aquaponics system. *International Journal of Applied Science and Engineering*, 17(2): 207–222.

Wei, L. 2023. Genetic Algorithm Optimization of Concrete Frame Structure Based on Improved Random Forest. In 2023 International Conference on Electronics and Devices, Computational Science (ICEDCS). IEEE: 249–253.

Wen, H., Liu, B., Di, M., Li, J. & Zhou, X. 2024. A SHAP-enhanced XGBoost model for interpretable prediction of coseismic landslides. Advances in Space Research, 74(8): 3826–3854. https://doi.org/10.1016/j.asr.2024.07.013.

Wibowo, R.R.D.I., Ramdhani, M., Priramadhi, R.A. & Aprillia, B.S. 2019. IoT-based automatic monitoring system for water nutrition on aquaponics system. In *Journal of Physics: Conference Series*. IOP Publishing: 1–10. https://iopscience.iop.org/article/10.1088/1742-

6596/1367/1/012071 [2 June 2023].

Wijayanto, A., Wardhana, K. & Aziz, A. 2021. Implementation of Internet of Things (IoT) for aquaponic system automation. In *Proceedings of the 2021 International Conference on Computer, Control, Informatics and Its Applications*. 176–181.

Wilson, A., Sukumar, R. & Hemalatha, N. 2021. Machine learning model for rice yield prediction using KNN regression. *agriRxiv*: 1–6

Wongkiew, S., Hu, Z., Chandran, K., Lee, J.W. & Khanal, S.K. 2017. Nitrogen transformations in aquaponic systems: a review. *Aquacultural Engineering*, 76: 9–19. https://doi.org/10.1016/j.aquaeng.2017.01.004.

World Health Organization. 2021. *Pandemic year marked by spike in world hunger*. UN Report. https://www.who.int/news/item/12-07-2021-un-report-pandemic-year-marked-by-spike-in-world-hunger [16 April 2022].

Yanes, A.R., Martinez, P. & Ahmad, R. 2020. Towards automated aquaponics: a review on monitoring, IoT, and smart systems. *Journal of Cleaner Production*, 263: 1–15. https://doi.org/10.1016/j.jclepro.2020.121571.

Yang, J., Guo, Y., Chen, T., Qiao, L. & Wang, Y. 2023. Data-driven prediction of greenhouse aquaponics air temperature based on adaptive time pattern network. *Environmental Science and Pollution Research*, 30(16): 48546–48558.

Yang, L. & Shami, A. 2020. On hyperparameter optimization of machine learning algorithms: theory and practice. *Neurocomputing*, 415: 295–316. https://doi.org/10.1016/j.neucom.2020.07.061.

Yelapure, S.J. & Kulkarni, R. V. 2012. Literature review on expert system in agriculture. *International Journal of Computer Science and Information Technologies*, 3(5): 5086–5089.

Yildiz, H.Y., Radosavljevic, V., Parisi, G. & Cvetkovikj, A. 2019. Insight into risks in aquatic animal health in aquaponics. In Goddek, S., Joyce, A., Kotzen, B., & Burnell, G. M.(eds.). *Aquaponics Food Production Systems*. Cham, Switzerland: Springer International Publishing, 435–452.

Yildiz, H.Y., Robaina, L., Pirhonen, J., Mente, E., Domínguez, D. & Parisi, G. 2017. Fish welfare in aquaponic systems: its relation to water quality with an emphasis on feed and faeces- a review. *Water*, 9(1): 1–17. https://doi.org/10.3390/w9010013.

Yu, T. & Zhu, H. 2020. Hyper-parameter optimization: a review of algorithms and applications. *arXiv preprint arXiv:2003.05689*: 1–56.

Zaini, A., Kurniawan, A. & Herdhiyanto, A.D. 2018. Internet of Things for monitoring and controlling nutrient film technique (NFT) aquaponic. In the *2018 International Conference on Computer Engineering, Network and Intelligent Multimedia (CENIM)*. Surabaya, Indonesia: IEEE: 167–171.

Zaras, A., Passalis, N. & Tefas, A. 2022. Neural networks and backpropagation. In Iosifidis, A., & Tefas, A.(eds.). *Deep Learning for Robot Perception and Cognition*. Greece: Academic Press, 17–34.

Zeebaree, M. & Aqel, M. 2019. A Comparison Study between Intelligent Decision Support Systems and Decision Support Systems. *The ISC Int'l Journal of Information Security*, 11(3): 187–194. http://www.isecure-journal.org [29 October 2021]

Zhang, C. 2021. Intelligent Internet of Things service based on artificial intelligence technology. In the *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering, ICBAIE 2021*. Nanchang, China: IEEE: 731–734.

Zhang, C. & Lu, Y. 2021. Study on artificial intelligence: the state of the art and future prospects. *Journal of Industrial Information Integration*, 23: 1–9. https://doi.org/10.1016/j.jii.2021.100224.

Zhang, S., Li, J. & Li, Y. 2023. Reachable distance function for KNN classification. *IEEE Transactions on Knowledge and Data Engineering*, 35(7): 7382–7396.

Zhang, Y. 2023. Research of boosting algorithm machine learning in logistics enterprise financial risk prediction. In the *2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*. IEEE: 94–98.

Zhang, Y., Liu, J. & Shen, W. 2022. A review of ensemble learning algorithms used in remote sensing applications. *Applied Sciences*, 12(17): 1–20.

Zhao, R., Hong, L., Ji, H., Zhang, Q., Zhang, S., Li, Q. & Gong, H. 2025. Decision tree-based parameter identification and state estimation: application to reactor operation digital twin. *Nuclear Engineering and Technology*: 1–9.

Zhou, Z. 2012. *Ensemble methods: foundations and algorithms*. New York: CRC Press.

APPENDICES

Appendix A: Ethical clearance

Office of the Research Ethics Committee
Faculty of Informatics and Design
Room 2.09
80 Roeland Street
Cape Town
Tel: 021-469 1012
Email: ndedem@cput.ac.za
Secretary: Mziyanda Ndede

02 June 2022

Mrs Anila Mundackal
c/o Department of Information Technology
CPUT

Reference no: 220635323/2022/12

Project title: A Decision Support System for yield prediction and monitoring of aquaponics based on the Intelligent Internet of Things

Approval period: 02 June 2022 – 31 December 2023

This is to certify that the Faculty of Informatics and Design Research Ethics Committee of the Cape Peninsula University of Technology conditionally approves the methodology and ethics of Mrs Anila Mundackal (220635323) for Doctor of Technology: Information and Communications Technology.

Any amendments, extension or other modifications to the protocol must be submitted to the Research Ethics Committee for approval.

The Committee must be informed of any serious adverse event and/or termination of the study.

Dr Blessing Makwambeni
Acting Chair: Research Ethics Committee
Faculty of Informatics and Design
Cape Peninsula University of Technology

Appendix B: Individual consent for research participation

FID/REC/ICv0.1

FACULTY OF INFORMATICS AND DESIGN

Individual Consent for Research Participation

Title of the study: **A Decision Support System for yield prediction and monitoring of aquaponics based on the Intelligent Internet of Things**

Name of researcher: Anila Mundackal
Contact details: 0713794662 email: 220635323@mycput.ac.za phone:

Name of supervisor: Prof. Justine O Daramola
Contact details: email: DARAMOLAJ@cput.ac.za phone: 021 460 3184

Purpose of the Study: The research aims to develop a decision support system for aquaponics yield prediction and monitoring to aid farmers in making decisions to achieve maximum productivity.

Participation: My participation will consist essentially of feedback from aquaponics farmers.

Confidentiality: I have received assurance from the researcher that the information I will share will remain strictly confidential unless noted below. I understand that the contents will be used only for thesis, journal articles and that my confidentiality will be protected by the participant's anonymity will be always maintained throughout. All data received from participants will be treated as confidential.

Anonymity will be protected in the following manner avoid disclosing information of personal/identity details.

Conservation of data: The data collected will be kept in a secure manner the data from the survey will strictly be kept on the Belgium Campus server securely. Any other survey data will be kept on the domain laptop device that is password protected. This data will be kept for 2-3 years.

Voluntary Participation: I am under no obligation to participate and if I choose to participate, I can withdraw from the study at any time and/or refuse to answer any questions, without suffering any negative consequences. If I choose to withdraw, all data gathered until the time of withdrawal will be destroyed.

Additional consent: I make the following stipulations (please tick as appropriate):

1

	In thesis	In research publications	Both	Neither
My image may be used:	√	√	√	
My name may be used:	√	√	√	
My exact words may be used:	√	√	√	
Any other (stipulate):				

Acceptance:
I, *Prof Michael Rudolph,*

agree to participate in the above research study conducted by Anila Mundackal of the Faculty of Informatics and Design Information Technology at the Cape Peninsula University of Technology, which research is under the supervision of Prof. Justine O Daramola .

If I have any questions about the study, I may contact the researcher or the supervisor. If I have any questions regarding the ethical conduct of this study, I may contact the secretary of the Faculty Research Ethics Committee at 021 469 1012, or email naidoove@cput.ac.za.

Participant's signature: __ __ Date: *03 05 2022*

Researcher's signature: ________________ Date: 2022-05-04

____

2

OFFICE OF THE DIRECTOR
CENTRE FOR ECOLOGICAL INTELLIGENCE
FACULTY OF ENGINEERING AND THE BUILT ENVIRONMENT

Professor Michael Rudolph
Telephone: +27 82 492 4768
E-mail: michaelr@uj.ac.za

3 May 2022

I, Michael Rudolph, in my capacity as Director of the Centre for Ecological Intelligence, Faculty of Engineering and Built Environment at the University of Johannesburg give consent in principle to allow Anila Mundackal, a student at the Cape Peninsula University of Technology, to collect data in this company/from me as part of her doctoral research. The student has explained to me the nature of their research and the nature of the data to be collected.

This consent in no way commits any individual person to participate in the research, and it is expected that the student will get individual consent from any participants.

I reserve the right to withdraw this permission at any time.

In addition, the company's/my name may or may not be used as indicated below (tick as appropriate):

	Thesis	Conference paper	Journal article	Research poster
Yes	√	√	√	√
No				

Michael Rudolph

3 May 2022

BELGIUM CAMPUS iTversity
It's the way we're wired

I Dr JC Mentz, in my capacity as Academic Dean at Belgium Campus ITversity give consent in principle to allow Anila Mundackal, a student at the Cape Peninsula University of Technology, to collect data in this company as part of her Doctoral research. The student has explained to me the nature of their research and the nature of the data to be collected.

This consent in no way commits any individual person to participate in the research, and it is expected that the student will get individual consent from any participants. I reserve the right to withdraw this permission at any time.

In addition, Belgium Campus ITversity's name may be used as indicated below (tick as appropriate):

	Thesis	**Conference paper**	**Journal article**	**Research poster**
Yes	X	X	X	X
No				

Dr JC Mentz

26 April 2022

Belgium Campus ITversity
1 NPC 2001/017971/08

138 Berg Avenue
Heatherdale AH
Pretoria
South Africa

PO Box 60327
Karenpark 0118
South Africa

Tel: +27 (0) 10 593 5368
Fax: +27 (0)12 542-1617

info@belgiumcampus.ac.za
www.belgiumcampus.ac.za

Appendix C: Experimental data

Table C.1: Plant diameter prediction using *Gridsearch* with 5-fold CV

ML models	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R-squared	Adjusted- R-squared
Linear Regression	0.02	0.15	0.11	0.54	0.53
Random Forest	0.00	0.05	0.03	0.93	0.93
KNN	0.01	0.09	0.06	0.84	0.83
XGBoost	0.00	0.05	0.03	0.93	0.93
MLP	0.01	0.08	0.06	0.82	0.81

Table C.2: Plant diameter prediction using *Randomsearch* with 5-fold CV

ML models	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R-squared	Adjusted- R-squared
Linear Regression	0.02	0.15	0.11	0.54	0.53
Random Forest	0.00	0.05	0.03	0.93	0.93
KNN	0.01	0.09	0.06	0.82	0.82
XGBoost	0.00	0.05	0.03	0.94	0.94
MLP	0.01	0.07	0.05	0.85	0.86

Table C.3: Plant diameter prediction using *Randomsearch* with 10-fold CV

ML models	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R-squared	Adjusted- R-squared
Linear Regression	0.02	0.15	0.11	0.54	0.53
Random Forest	0.00	0.05	0.03	0.94	0.93
KNN	0.01	0.09	0.06	0.82	0.82
XGBoost	0.00	0.05	0.03	0.94	0.94
MLP	0.01	0.07	0.05	0.85	0.86

Table C.4: Plant height prediction using *Gridsearch* with 10-fold CV

ML models	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R-squared	Adjusted-R-squared
Linear Regression	0.01	0.09	0.06	0.80	0.80
Random Forest	0.00	0.06	0.05	0.92	0.92
KNN	0.00	0.07	0.05	0.91	0.91
XGBoost	0.00	0.06	0.05	0.92	0.92
MLP	0.00	0.06	0.05	0.92	0.92

Table C.5: Plant height prediction using *Randomsearch* with 5-fold CV

ML models	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R-squared	Adjusted-R-squared
Linear Regression	0.01	0.09	0.06	0.80	0.80
Random Forest	0.00	0.06	0.05	0.92	0.92
KNN	0.00	0.07	0.05	0.91	0.91
XGBoost	0.00	0.06	0.05	0.92	0.92
MLP	0.00	0.06	0.05	0.92	0.91

Table C.6: Plant height prediction using *Randomsearch* with 10-fold CV

ML models	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R-squared	Adjusted-R-squared
Linear Regression	0.01	0.09	0.06	0.80	0.80
Random Forest	0.00	0.06	0.05	0.92	0.92
KNN	0.00	0.07	0.05	0.91	0.91
XGBoost	0.00	0.06	0.05	0.92	0.92
MLP	0.00	0.06	0.05	0.92	0.91

Table C.7: Water pH prediction using *Gridsearch* with 5-fold CV

ML models	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R-squared	Adjusted- R-squared
Linear Regression	0.03	0.19	0.15	0.55	0.54
Random Forest	0.02	0.14	0.10	0.77	0.76
KNN	0.02	0.13	0.09	0.79	0.78
XGBoost	0.02	0.13	0.09	0.79	0.79
MLP	0.03	0.17	0.14	0.65	0.64

Table C.8: Water pH prediction using *Randomsearch* with 5-fold CV

ML models	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R-squared	Adjusted- R-squared
Linear Regression	0.03	0.19	0.15	0.55	0.54
Random Forest	0.02	0.14	0.10	0.77	0.76
KNN	0.02	0.13	0.09	0.78	0.77
XGBoost	0.02	0.15	0.12	0.71	0.71
MLP	0.03	0.18	0.14	0.59	0.59

Table C:9 Water pH prediction using *Randomsearch* with 10-fold CV

ML models	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R-squared	Adjusted- R-squared
Linear Regression	0.03	0.19	0.15	0.55	0.54
Random Forest	0.02	0.13	0.09	0.77	0.77
KNN	0.02	0.13	0.09	0.78	0.77
XGBoost	0.02	0.15	0.12	0.71	0.71
MLP	0.03	0.18	0.14	0.59	0.59

Table C.10: Water TDS prediction using *Gridsearch* with 5-fold CV

ML models	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R-squared	Adjusted- R-squared
Linear Regression	0.00	0.01	0.01	0.99	0.99
Random Forest	0.00	0.02	0.01	0.99	0.99
KNN	0.00	0.02	0.01	0.99	0.99
XGBoost	0.00	0.02	0.01	0.98	0.98
MLP	0.00	0.03	0.02	0.98	0.98

Table C.11: Water TDS prediction using *Randomsearch* with 5-fold CV

ML models	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R-squared	Adjusted- R-squared
Linear Regression	0.00	0.01	0.01	0.99	0.99
Random Forest	0.00	0.03	0.01	0.98	0.98
KNN	0.00	0.02	0.01	0.99	0.99
XGBoost	0.00	0.03	0.01	0.98	0.98
MLP	0.00	0.02	0.02	0.99	0.99

Table C.12: Water TDS prediction using *Randomsearch* with 10-fold CV

ML models	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	R-squared	Adjusted- R-squared
Linear Regression	0.00	0.01	0.01	1.00	1.00
Random Forest	0.00	0.03	0.01	0.98	0.97
KNN	0.00	0.02	0.01	0.99	0.99
XGBoost	0.00	0.03	0.01	0.98	0.98
MLP	0.00	0.02	0.02	0.99	0.99