AI-Driven Optimization of Hydroponic Farming through IoT-Based Water Quality Monitoring and Predictive Analytics

Lindsay Cooper
Fontys University of Applied Sciences
AI Systems Minor

January 16, 2026

Abstract

Hydroponic systems are an increasingly vital part of sustainable agriculture but require continuous monitoring of water parameters to ensure optimal plant growth. Artificial Intelligence (AI) combined with Internet of Things (IoT) sensors can transform these systems into intelligent, self-regulating environments that minimize resource waste and maximize yield. This paper investigates how AI can be applied to optimize hydroponic farming by integrating IoT-based monitoring of critical water parameters with predictive analytics for sustainable resource management. The study explores key water quality factors, suitable IoT sensor technologies, data pre-processing methods, AI modeling approaches, and system integration strategies. The outcome aims to design a scalable, data-driven architecture that supports real-time decision-making and contributes to resource-efficient, sustainable food production.

1 Introduction

Agriculture accounts for a significant share of global water and energy use. Hydroponics, a soil-free cultivation technique, has emerged as a sustainable alternative that enables high-density crop production with efficient water use. However, maintaining the delicate balance of nutrient solutions requires continuous monitoring of factors such as pH, electrical conductivity (EC), temperature, and dissolved oxygen.

Traditional manual monitoring is labor-intensive and prone to error. Recent advances in IoT and AI provide new opportunities to automate and optimize these processes. IoT sensors can collect high-resolution data in real-time, while AI models can analyze this data to predict fluctuations, detect anomalies, and make recommendations.

This research investigates how AI can enhance hydroponic farming efficiency by integrating IoT-based water monitoring and predictive analytics.

1.1 Research Question

Main Question:
How can AI be applied to optimize hydroponic farming by combining IoT-based monitoring of water parameters with predictive analytics for sustainable resource management?

1.2 Sub-Questions

1. Which water parameters (e.g., EC, pH, temperature, dissolved oxygen) are most critical in hydroponic farming, and how do fluctuations affect crop growth and resource efficiency?
2. What IoT sensors and data pipelines can be used to reliably measure and transmit these parameters in real-time within a hydroponic setup?
3. How should sensor data be preprocessed (e.g., cleaning, normalization, anomaly handling) to make it suitable for predictive modeling?
4. Which predictive analytics and AI techniques (e.g., regression, time series forecasting, anomaly detection) are most effective for forecasting water quality and detecting irregularities?
5. How can IoT monitoring and AI models be integrated into a system architecture that supports real-time decision-making and control in hydroponics?
6. What forms of visualization or feedback (dashboards, alerts, recommendations) can make the system usable and valuable for farmers?
7. To what extent can AI-driven optimization reduce water and nutrient waste, and what is the ecological and financial impact compared to traditional/manual approaches?
8. What risks, limitations, and future opportunities arise when applying AI and IoT to hydroponic farming for sustainable resource management?
9. What anomalies might occur?

2 Literature Review

This section synthesizes research on AI, IoT, and hydroponic management.

2.1 Hydroponics and Water Quality Parameters

In hydroponic farming, water quality is a critical determinant of plant growth and yield. Because plants are grown without soil, the nutrient solution becomes the sole medium through which water, minerals, and dissolved gases are supplied. As a result, any imbalance in water composition can directly affect plant health, nutrient uptake efficiency, and overall system performance. Understanding the chemical and physical properties of water, as well as its dissolved components, is therefore essential for effective hydroponic management.

2.1.1 Water as a Compound

Water is a chemical compound composed of two hydrogen atoms and one oxygen atom (H_2O). These atoms are bonded through covalent bonds formed by shared electrons. Due to the difference in electro-negativity between hydrogen and oxygen, the water molecule is polar, with a partial negative charge on the oxygen atom and partial positive charges on the hydrogen atoms. This polarity gives water several unique physical and chemical properties, including high solvent capacity, high specific heat capacity, surface tension, and capillary action.

While pure water consists only of H_2O molecules, water used in agricultural systems is never chemically pure. It typically contains dissolved minerals, salts, gases, and organic compounds. The type and concentration of these dissolved components can significantly influence plant growth, either positively or negatively, depending on their balance.

2.1.2 Dissolved Minerals

Water used in hydroponic systems contains dissolved minerals that serve as the primary source of nutrients for plants. These minerals are absorbed by plant roots in ionic form and are essential for physiological processes such as photosynthesis, enzyme activation, and cell development. The composition and concentration of dissolved minerals directly influence plant growth, yield, and overall health.

Dissolved minerals are generally classified into macro-nutrients and micro-nutrients. Macro-nutrients are required in relatively large quantities, while micro-nutrients are needed in trace amounts. Table 1 summarizes common dissolved minerals and their roles in plant development.

In hydroponic systems, maintaining the correct concentration and ratio of dissolved minerals is essential. Nutrient imbalances can quickly lead to conditions such as nutrient lockout or toxicity. Excessive sodium concentrations or deficiencies in calcium are common examples of imbalances that negatively affect plant health. Continuous monitoring of water quality parameters such as electrical conductivity (EC) and pH helps ensure that nutrient levels remain within optimal ranges.

Table 1: Dissolved Minerals in Water

Component	**Example**	**Role/Effect**
Macro-nutrients	Nitrates, Phosphates, Potassium, Calcium, Magnesium, Sulfates	Essential for plant growth—photosynthesis, enzyme function, cell development
Micro-nutrients	Iron, Manganese, Zinc, Copper, Boron, Molybdenum	Required in small amounts for metabolic and enzymatic activities

2.1.3 Dissolved Gases

Water in hydroponic systems contains dissolved gases, primarily oxygen and carbon dioxide. Dissolved oxygen is critical for root respiration, enabling plants to generate energy and effectively absorb nutrients. Insufficient oxygen levels in the root zone can result in reduced growth, increased susceptibility to root diseases, and anaerobic conditions.

Carbon dioxide plays a more indirect role in hydroponic water chemistry. It contributes to pH buffering and supports microbial activity within the nutrient solution. Maintaining appropriate levels of dissolved gases is therefore important for sustaining a healthy root environment.

2.1.4 Salts and Other Ions

Water may also contain various salts and ions, such as sodium and chloride. While small concentrations of these ions may not be harmful, excessive salt accumulation can lead to osmotic stress. Osmotic stress disrupts normal cellular processes by reducing the plant's ability to absorb water, which can ultimately result in stunted growth or plant death. Effective nutrient management and regular monitoring help prevent salt buildup in recirculating hydroponic systems.

2.1.5 pH and Buffering Ions

The pH level of the nutrient solution strongly influences nutrient solubility and availability. Most hydroponically grown crops perform optimally within a pH range of approximately 5.5 to 6.5. Values outside this range can impair the uptake of essential nutrients such as iron, phosphorus, and calcium.

Buffering ions, particularly bicarbonates, can increase solution alkalinity and cause gradual pH drift. In practice, pH adjustments are commonly made using acid or base solutions to maintain stability within the optimal range.

2.1.6 Electrical Conductivity

Electrical conductivity (EC) represents the total concentration of dissolved ions in a nutrient solution and is typically expressed in units of mS/cm or dS/m. Only substances that dissociate into charged ions contribute to EC measurements; neutral molecules do not affect conductivity.

In hydroponic systems, most nutrient additives are salts that dissociate into positively and negatively charged ions when dissolved in water. The greater the ionic concentration, the higher the EC value. EC is therefore widely used as an indirect measure of nutrient strength. Maintaining EC within crop-specific target ranges helps prevent under-fertilization, nutrient toxicity, and osmotic stress.

2.2 Sensors, Micro-Controllers, and Data Pipeline

A reliable and accurate IoT system consists of sensors, the correct micro-controller hardware, and a robust data pipeline. The data pipeline should ensure a continuous but low latency transmission. Each water quality parameter(pH, EC, DO, Temp, etc.) requires a specialized sensor with the appropriate signal conditioning and communication interfaces.

Table 2: IoT Sensors for Water Quality

Sensor	**Output Type**	**Importance**
pH probe	Analog, I^2C, UART	pH determines nutrient solubility; deviations cause nutrient lockout.
Electrical Conductivity (EC) Sensor	Analog or UART/I^2C	EC indicates nutrient strength and allows automated dosing.
Temperature Sensor	1-Wire, analog, digital	Temperature affects nutrient uptake and oxygen solubility.
Dissolved Oxygen (DO) Sensor	I^2C, UART	Root oxygen availability determines growth rate and prevents anaerobic conditions.
Oxidation-Reduction Potential (ORP) Sensor	Analog, UART/I^2C	Water cleanliness, microbial activity, root-zone biological stability.

2.3 Communication from Sensor to Micro-controller

Sensors need reliable communication with micro-controllers. The protocol you choose can effect or determine accuracy, stability, wiring complexity and even noise immunity. It's important to understand the communication protocols in IoT systems. In a hydroponic environment digital protocols(I^2C, UART, 1-Wire) are preferred due to high possibility of noise. Digital protocols encode data into packets with error-checking which makes them far more stable than raw analog voltage signals.

Table 3: Communication Protocols for IoT Hydroponic Sensors

Protocol	Best For	Key Characteristics and Benefits
I^2C	EC, DO, temperature sensors with DSPs	Two-wire bus (SDA, SCL) Multiple sensors share one bus (addressed) More stable than analog on long cables Highly noise-resistant Ideal for digital water-quality sensors ESP32/Arduino compatible
UART	Smart probes (EZO), ORP, industrial sensors	TX/RX two-wire serial Point-to-point communication Very stable over long distances Digital encoding reduces noise Used by lab-grade pH/EC/DO/ORP sensors Supports sensors with onboard calibration
Analog (ADC)	Low-cost pH/EC sensors	Outputs voltage (0–3.3V / 0–5V) Prone to pump/doser/cable electrical noise Drift increases over time Requires frequent calibration Works with filtering or isolation boards Best for low-budget builds
1-Wire	DS18B20 temperature sensing	Single wire for power + data Many sensors on one bus Unique 64-bit ID per sensor Very stable for temperature Supports long waterproof cable probes Works for water, air, reservoir monitoring

2.3.1 Micro-controllers for Real-Time Data Collection

Micro-controllers are essentially miniature computing components providing processing power and control to modern electrical projects. This component is critical making projects responsive and smart. There are many factors to think over when choosing the right micro-controller like cost, power efficiency, performance, scalability, and reliability.

Table 4: Micro-controllers

Type	**Benefits**
ESP32	Best option for hydroponics; Built-in bluetooth + wifi; Supports I^2C, UART, ADC, Low power consumption
ESP8266	Supports UART, I2C, SPI, Wi-Fi; Low Memory and low power consumption; good for simple projects, affordable
Andruino	Easy for beginners, Less efficient than ESP32 + native wifi
Raspberry Pi	For complex processing, Useful for running local dashboards or predictive models

2.3.2 Data Pipeline

An effective hydroponics system requires several pipeline stages:

- **Data Acquisition Layer:** The ESP32 or similar microcontrollers read sensor outputs, apply initial filtering, apply timestamps, and format data into structured packets (e.g., JSON, CSV, or binary).

- **Local Processing:** Before transmission, the microcontroller may apply calibration equations, outlier detection, short-term buffering, and local decision-making.

- **Data Transmission Layer:** Wireless communication typically uses Wi-Fi via the ESP32 and the MQTT protocol, a lightweight publish–subscribe system suitable for limited-bandwidth or low-power environments. HTTPS/REST APIs can also be used for cloud logging and data synchronization.

- **Cloud Integration Layer:** Data is pushed to cloud services such as InfluxDB, Firebase, or Google Cloud IoT Core, where it is stored for analysis and visualization.

- **Processing & Analytics Layer:** Cloud or local servers perform long-term storage, anomaly detection, trend analysis, and machine learning inference for predictive monitoring.

- **User Interface Layer:** Displaying information on a dashboard via web-based UI(Grafana, custom React/Vue), mobile apps, real-time alert systems(email or sms). The way of providing the grower information and continuous insight.

A reliable system can be carefully crafted by selecting appropriate IoT sensors and pipeline considerations. Coupling these sensors with an ESP32 micro-controller, MQTT-based communication, and cloud-integrated analytics can create intelligent automated system.

2.4 Sensor Data Preprocessing

Preprocessing sensor data is a very important step into creating accurate predictive models. Raw measurements that come from the sensors usually contain noise, drift, missing values, and irregular sampling levels. In order to create clean, consistent, and statistically reliable models used for forecasting and anomaly detection proper preprocessing should be completed. Proper preprocessing includes the following steps: data cleaning, synchronization, normalization, feature engineering, and handling anomalies.

2.4.1 Data Cleaning

2.4.2 Removing Physically Invalid Values

Sensors occasionally generate readings outside realistic physical ranges. Examples include:

- pH values below 0 or above 14
- EC values jumping from 1.2 mS/cm to above 20 mS/cm within one sample
- Temperature values below 0°C or above 50°C

These readings should be removed or replaced with interpolated values.

2.4.3 Noise Reduction

Hydroponic environments introduce electrical and mechanical noise. Useful smoothing methods include:

- Rolling or moving average filters
- Median filtering to remove sharp spikes
- Savitzky–Golay filtering to preserve overall trends

2.4.4 Handling Missing Data

Missing or dropped packets can be resolved by:

- Linear interpolation for short gaps
- Forward fill for slowly varying sensors (e.g., temperature)
- Leaving large gaps unfilled and flagged for model awareness

2.4.5 Timestamp Alignment and Resampling

Sensors often log values at irregular or asynchronous intervals. Time-series models require consistent spacing. Standard steps include:

1. Convert all timestamps to a unified time standard (UTC or local with offset).

2. Resample signals to a fixed interval (e.g., 10 s or 1 min).

3. Interpolate missing values created by resampling.

This produces a synchronized feature matrix suitable for modeling.

2.4.6 Normalization and Scaling

Machine learning models benefit from normalized input values. Common approaches include:

- **Min–Max Normalization:**

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

- **Z-score Standardization:**

$$x' = \frac{x - \mu}{\sigma}$$

- **Log Transformation** for skewed variables such as EC or ORP.

Normalization should be fitted on the training set only to avoid data leakage.

2.4.7 Drift Correction and Calibration

Hydroponic sensors, particularly pH, EC, and ORP probes, exhibit gradual drift. Preprocessing should include:

- Applying manufacturer calibration curves

- Using rolling windows to detect slow drift trends

- Excluding startup values before sensors reach equilibrium

2.5 Anomaly Detection and Outlier Handling

Anomalies can severely distort model performance. Detection methods include:

- Z-score thresholds (e.g., values beyond 3σ)

- Interquartile range (IQR) rules

- Time-series anomaly detection methods such as rolling-median deviation or Isolation Forest

Outliers may be removed, replaced, or labeled depending on the modeling objective.

2.5.1 Feature Engineering

To improve predictive accuracy, additional features should be derived from raw measurements:

- Rolling means (e.g., 5-minute pH average)
- Rate of change features, such as $\Delta EC/\Delta t$
- Time-of-day encodings using sine/cosine functions
- Interaction features (e.g., EC adjusted by temperature)

These engineered variables enable models to capture system dynamics more effectively.

2.5.2 Train/Test Splitting for Time-Series

Time-dependent data must be split chronologically rather than randomly. Recommended approaches include:

- Train–validation–test split in time order
- Walk-forward validation
- Sliding windows for forecasting models

2.6 Most effective AI techniques for forecasting and irregularity detection

For forecasting water-quality trajectories (e.g., predicting EC/temperature trends), time-series approaches are typically effective. Deep sequence models such as LSTM have been used successfully in controlled-environment agriculture and soilless/hydroponic contexts, including forecasting root-zone EC and other variables (Eraliev et al., 2023; Moon et al., 2018, 2019). For smaller datasets or when interpretability is important, classical regression and tree-based models can provide competitive baselines and clearer explanations of feature influence (Sharmin et al., 2025). For detecting irregularities, a layered strategy is practical: rule-based limits (hard bounds for pH/EC/temperature), statistical tests, and ML-based anomaly detection. Surveys of IoT anomaly detection highlight unsupervised methods as common choices when anomaly labels are scarce (Chatterjee et al., 2022). In practice, combining thresholds (fast and explainable) with an unsupervised detector (robust to complex patterns) often improves reliability.

2.7 Integrating IoT monitoring and AI into a real-time system architecture

A workable architecture is: sensors → microcontroller/edge gateway → message broker → time-series storage → analytics (forecasting + anomaly detection) → user interface + alerting. MQTT is a standard fit for the broker layer in IoT telemetry ("MQTT — The Standard for IoT Messaging", n.d.). Time-series storage supports rapid retrieval for dashboards and model features ("Time series database explained", n.d.). AI inference can run in the cloud for flexibility or on-device for low latency; TensorFlow Lite supports on-device inference using a lightweight interpreter, enabling edge deployment where connectivity or latency is a concern ("TensorFlow Lite inference guide", n.d.). For decision support, the system should produce: predicted parameter values over a short horizon, anomaly flags with confidence, and recommended corrective actions (e.g., adjust dosing, increase aeration, investigate pump issues).

2.8 Visualization and feedback for farmer usability

Usable systems typically combine: real-time dashboards, threshold-based alerts, and action-oriented recommendations. Dashboards should show time-series plots for pH/EC/temperature/D recent anomalies, and short-term forecasts. Grafana is a widely used open-source dashboard platform that supports time-series visualization and alerting, making it suitable for IoT telemetry ("Grafana OSS: Visualization and dashboarding technology", n.d.; "Time series visualization — Grafana documentation", n.d.). Alerts (SMS/email/app) should trigger on sustained threshold violations (e.g., pH outside range for >N minutes) and on high-confidence anomalies. Recommendations should be simple and contextual (e.g., "EC rising faster than normal; check dosing pump and reservoir mixing").

2.9 Impact on water/nutrient waste and ecological/financial outcomes

Hydroponics can be more water-efficient than soil cultivation because water is recirculated and delivered directly to roots; controlled comparisons have reported lower product water use and higher water-use efficiency in hydroponic systems relative to soil (Verdoliva et al., 2021). More recent analyses also report substantial reductions in water use (and sometimes fertilizer) depending on crop and system design (Regmi et al., 2024). AI-driven optimization can further reduce waste by preventing over-dosing (via tighter EC control), reducing emergency drain/refill events (via early anomaly detection), and maintaining conditions closer to optimal (reducing yield loss and input waste). Financial impacts follow from reduced consumables (nutrients, water), avoided crop loss, and lower labor time spent on manual testing; these should be quantified in your evaluation using measurable baselines (manual monitoring vs AI-supported monitoring) and cost models of inputs and labor.

2.10 Risks, limitations, and future opportunities

Key risks include sensor drift and calibration burden (especially pH/EC probes), electrical noise in wet environments, network outages, and data quality issues (missingness, timestamp errors). From the AI perspective, limitations include distribution shift across seasons/crops, insufficient labeled anomalies, and overfitting on short datasets. Security and reliability are also relevant in connected systems, as IoT deployments may be exposed to attacks or misconfiguration (Zahoor et al., 2025). Future opportunities include: edge inference for rapid local decisions (e.g., running lightweight models with TensorFlow Lite), multimodal sensing (adding flow, level, light intensity), and closed-loop control with safe-guards (human-in-the-loop and constraints) to move from advisory outputs to semi-automated control.

2.11 Anomalies in Hydroponic Water Quality and Sensor Data

Typical anomalies fall into three categories: (i) sensor/measurement anomalies (probe drift, calibration loss, stuck-at values, noise spikes, disconnected probes); (ii) process anomalies (rapid EC jumps from dosing errors, pH runaway from buffer depletion, temperature spikes from heater failures, DO drops from pump/aerator failure); and (iii) communication/data anomalies (dropped packets, duplicated timestamps, out-of-order data). Because labeled anomaly data is often scarce, IoT anomaly detection is commonly approached with unsupervised or semi-supervised techniques, supplemented by domain thresholds for hard safety limits (Chatterjee et al., 2022). Several anomalies may occur:

- **Sensor Drift:** Gradual deviation in pH or EC sensor readings due to electrode wear ((Brown & Chen, 2019)).
- **Sudden Parameter Spikes:** Rapid increase in EC caused by nutrient overdosing or evaporation.
- **pH Instability:** Upward drift caused by high alkalinity or microbial activity.
- **Dissolved Oxygen Drops:** Caused by pump failure or high root biomass.
- **Temperature Anomalies:** Heater malfunction or sunlight exposure causing overheating.
- **Communication Failures:** Missing sensor packets leading to incomplete data streams.

- **Electrical Noise:** Power fluctuations introducing false spikes in sensor values.

Critical parameters include pH, EC, temperature, and dissolved oxygen. Studies indicate that fluctuations outside optimal thresholds can lead to nutrient imbalance and reduced growth efficiency ((Hernandez & Cruz, 2023)).

2.12 Literature Review Synthesis and Knowledge Gap

The reviewed literature validates that hydroponic water quality is strongly influenced by a small set of critical parameters, primarily pH, electrical conductivity, temperature, and dissolved oxygen. Previous studies consistently show that deviations in these parameters can negatively impact nutrient uptake efficiency, plant health, and overall yield stability.

IoT-based monitoring systems supply an effective means of monitoring these parameters in real time. Prior research highlights the use of low-cost micro controllers, digital water-quality sensors, and lightweight communication protocols such as MQTT to achieve reliable data transmission in resource-constrained agricultural environments. Time-series databases and cloud-based platforms further enable scalable storage and visualization of high-frequency sensor data.

Recent advancements in artificial intelligence and predictive analytics offer additional benefits by enabling early detection of abnormal system behavior and short-term forecasting of water quality trends. Time-series models, including recurrent neural networks and classical regression-based approaches, have been successfully applied in controlled-environment agriculture. However, many existing solutions rely on centralized cloud processing or focus on large-scale commercial systems.

Despite these advances, a gap remains in the availability of affordable, integrated solutions that combine IoT-based water quality monitoring with predictive analytics for small- to medium-scale hydroponic systems. Existing studies often address sensing, data analytics, or control in isolation, rather than as a unified system. This research aims to address this gap by designing and evaluating a scalable architecture that integrates real-time IoT sensing, data preprocessing, predictive modeling, and user-centered visualization to support sustainable resource management in hydroponic farming.

2.13 State of the Art in AI- and IoT-Based Hydroponic Systems

Recent advances in controlled-environment agriculture have led to significant progress in the integration of Internet of Things (IoT) technologies and artificial intelligence (AI) for hydroponic farming. Existing research demonstrates that continuous monitoring of water quality parameters such as pH, electrical conductivity (EC), temperature, and dissolved oxygen can substantially improve crop stability, nutrient efficiency, and yield consistency.

2.13.1 Current Achievements

State-of-the-art hydroponic systems commonly employ IoT sensor networks to enable real-time monitoring and remote supervision of nutrient solutions. Several studies report successful automation of pH and EC regulation using sensor feedback loops, reducing manual intervention and improving system stability. In parallel, machine learning techniques have been applied to predict nutrient consumption, detect abnormal system behavior, and forecast short-term water quality trends. Time-series models, including regression-based approaches and recurrent neural networks, have shown promising results in forecasting EC and temperature dynamics in controlled environments.

More recent systems increasingly incorporate edge computing concepts, where preliminary data processing and decision-making occur close to the sensors. This reduces latency and improves responsiveness for time-critical events such as rapid pH drift or oxygen depletion. Open-source hardware platforms and cloud-based dashboards further contribute to system accessibility and scalability.

2.13.2 Limitations of Existing Approaches

Despite these advancements, several limitations remain. Sensor drift, particularly in pH and EC probes, continues to affect long-term reliability and necessitates frequent calibration. Dissolved oxygen and ORP sensors remain relatively expensive, limiting their adoption in small-scale or hobbyist hydroponic systems. Additionally, many AI-driven solutions rely heavily on centralized cloud processing, which can introduce latency, increase dependency on network connectivity, and raise concerns regarding robustness in offline scenarios.

From an analytical perspective, existing studies often focus on isolated components of the system, such as sensing accuracy or model performance, without evaluating the end-to-end integration of data acquisition, preprocessing, predictive modeling, and user interaction. Furthermore, many proposed solutions are designed for industrial-scale operations and are not easily transferable to low-cost or small-scale deployments.

2.13.3 Opportunities and Research Directions

Current research trends highlight several opportunities for further development. Combining multi-sensor data fusion with predictive analytics can improve robustness against sensor noise and drift. The use of lightweight machine learning models deployed at the edge offers potential for faster anomaly detection and local decision support without continuous cloud dependency. Additionally, open-source frameworks and low-cost micro-controllers create opportunities to design affordable, modular hydroponic monitoring systems accessible to small-scale growers.

2.13.4 Research Gap

Although prior work demonstrates the feasibility of IoT-based monitoring and AI-assisted analysis in hydroponics, there remains a lack of integrated, low-cost systems that combine real-time sensing, data preprocessing, predictive analytics, and user-centered visualization within a unified architecture. This research addresses this gap by designing and evaluating an AI-driven hydroponic monitoring system that emphasizes affordability, scalability, and sustainability for small- to medium-scale hydroponic applications.

3 Project Definition and Research Scope

This section defines the project objectives, scope, deliverables, assumptions, and risks. It serves as a project charter that formally structures the research and guides the system design and evaluation phases.

3.1 Project Objectives

The primary objective of this research is to design and evaluate an AI-driven hydroponic monitoring system that integrates IoT-based water quality sensing with predictive analytics. The system aims to support sustainable resource management by improving monitoring accuracy, early anomaly detection, and decision support.

The specific objectives are:

- To monitor critical hydroponic water parameters using IoT sensors
- To design a reliable data acquisition and transmission pipeline
- To apply predictive analytics for forecasting water quality trends
- To detect anomalies that may indicate system failure or plant stress
- To evaluate the potential for reducing water and nutrient waste

3.2 Project Scope

3.2.1 In-Scope

The scope of this project includes:

- Analysis of hydroponic water quality parameters
- Selection and evaluation of IoT sensors and micro-controllers
- Design of a data pipeline for real-time monitoring

- Data preprocessing and feature engineering
- Application of predictive analytics and anomaly detection models
- Visualization and feedback mechanisms for system users

3.2.2 Out-of-Scope

The following aspects are explicitly excluded:

- Fully autonomous control of nutrient dosing or actuators
- Commercial-scale deployment
- Robotic harvesting or computer vision-based plant inspection

3.3 Key Deliverables

The expected deliverables of this research are:

- A reviewed body of literature on AI, IoT, and hydroponics
- A system architecture for IoT-based hydroponic monitoring
- Preprocessed datasets suitable for predictive modeling
- Predictive and anomaly detection models
- Evaluation of system performance and sustainability impact

3.4 Assumptions and Constraints

3.4.1 Assumptions

- Sensor data is available at sufficient quality and frequency
- Open-source tools provide adequate performance
- The experimental setup reflects typical hydroponic conditions

3.4.2 Constraints

- Limited project timeline
- Budget constraints on sensor hardware
- Limited computational resources for edge devices

3.5 Risk Analysis

Potential risks and mitigation strategies include:

Table 5: Project Risks and Mitigation Strategies

Risk	**Impact**	**Mitigation**
Sensor drift or failure	High	Regular calibration and validation
Noisy or missing data	Medium	Filtering and interpolation techniques
Model overfitting	Medium	Cross-validation and regularization
Hardware limitations	Low	Lightweight models and optimization

4 Methodology

4.1 Research Approach

This research follows a design science research methodology (DSRM), which is suitable for developing and evaluating practical solutions to real-world problems. The study combines literature research, system design, prototyping, and experimental evaluation to investigate how AI-driven predictive analytics can enhance IoT-based hydroponic water quality monitoring.

The research is exploratory and applied in nature. Rather than aiming to develop novel machine learning algorithms, the focus is on selecting, integrating, and evaluating existing AI and IoT technologies within a coherent system architecture for sustainable resource management.

4.2 System Architecture Overview

The proposed system consists of five main components: sensing, edge processing, data transmission, analytics, and user interaction. Water quality parameters (pH, EC, temperature, and dissolved oxygen) are measured using IoT sensors connected to a microcontroller (ESP32). Sensor readings are timestamped, preprocessed, and transmitted via MQTT to a cloud-based data storage service.

A time-series database stores the incoming data and provides input to predictive analytics models. These models perform short-term forecasting and anomaly detection on water quality parameters. The results are presented to the user through a dashboard interface, along with alerts and actionable recommendations.

4.3 Functional and Non-Functional Requirements

The system requirements were derived directly from the research objectives and sustainability goals of hydroponic farming. These requirements guided architectural decisions, data handling strategies, and the selection of analytics techniques.

4.3.1 Functional Requirements

The system is required to:

- Monitor key hydroponic water quality parameters, including pH, electrical conductivity (EC), temperature, and dissolved oxygen.
- Timestamp and store sensor measurements for time-series analysis.
- Provide real-time visualization of current and historical sensor data.
- Generate alerts when parameters deviate from predefined safe ranges.
- Apply predictive analytics for short-term forecasting and anomaly detection.

4.3.2 Non-Functional Requirements

The system must additionally satisfy the following constraints:

- **Reliability:** Continuous operation with minimal data loss.
- **Scalability:** Support for additional sensors and extended logging.
- **Maintainability:** Modular and extensible software architecture.
- **Security:** Encrypted communication and authenticated dashboard access.
- **Energy Efficiency:** Compatibility with ESP32 low-power operation modes.

A complete functional requirements specification is provided in Appendix A.

4.4 Data Collection and Data Sources

Two complementary data sources are used:

4.5 Data Collection

Water quality data will be collected from a hydroponic setup using IoT sensors measuring pH, electrical conductivity, temperature, and dissolved oxygen. Sensors will be connected to an ESP32 micro-controller, which samples each parameter at fixed time intervals (e.g., every 5 minute).

If access to a physical hydroponic system is limited, additional datasets may be generated through controlled simulations or obtained from open-source hydroponic or controlled-environment agriculture datasets. All collected data will include timestamps and sensor identifiers to support time-series analysis.

4.5.1 IoT Sensor Data from a Prototype Setup

Water-quality data is collected from a hydroponic reservoir via sensors attached to an ESP32. Measurements include pH, EC, temperature, and DO (optional ORP). Each sample is assigned a timestamp at the edge device to preserve temporal ordering. The sampling interval is configurable; in the prototype it is selected to balance responsiveness (detecting rapid drift) and power/network constraints. Sensor calibration metadata (e.g., calibration date, reference solutions, slope/offset parameters) is logged to support drift management.

4.5.2 Synthetic Dataset for Model Development

To support model development and experimentation when physical data is limited, a synthetic dataset is generated using a simulation of hydroponic dynamics and an "expert" dosing policy. The simulator produces realistic trajectories for pH, EC, temperature, DO, and ORP with measurement noise, drift-like behavior, and occasional anomalies (e.g., aeration failure causing DO drops). The labels represent recommended corrective actions (nutrient A/B dosing and pH up/down dosing). Synthetic data enables controlled experimentation, repeatable training, and ablation studies of feature engineering choices.

4.5.3 Database Storage and Access

Sensor and synthetic data are stored in a time-series database. The database schema includes a timestamp column, run identifiers (for independent runs or growth cycles), raw sensor parameters, and any actuator/dosing events. Data is retrieved for training via an API query endpoint or via a PostgreSQL-compatible interface depending on deployment constraints. Time-indexed storage enables efficient range queries for dashboards and model feature extraction.

4.6 Data Preprocessing

Before modeling, raw sensor data will undergo preprocessing steps as described in Section 2. These steps include removal of physically invalid values, noise reduction using rolling or median filters, timestamp alignment, and normalization.

Drift correction will be applied using calibration curves and rolling-window analysis. Missing values caused by transmission issues will be handled using interpolation for short gaps and flagged for larger gaps. Feature engineering techniques such as rolling averages and rate-of-change features will be derived to support predictive modeling.

4.6.1 Data Cleaning and Validity Checks

Physically impossible readings are removed or flagged (e.g., pH < 0 or > 14, temperature outside plausible range, unrealistically large one-step jumps). Sensor-specific bounds are used as hard safety constraints. Duplicate or out-of-order samples are handled by sorting by timestamp and de-duplicating based on sensor ID and timestamp.

4.6.2 Noise Reduction and Smoothing

To mitigate electrical noise and mechanical disturbances, smoothing is applied using rolling means or median filters. Median filters are preferred for suppressing spikes without distorting step changes caused by dosing events. Smoothing parameters (window length) are chosen relative to the sampling interval.

4.6.3 Missingness Handling

Short gaps caused by dropped packets are filled using interpolation (linear for EC/pH, forward-fill for slowly varying variables such as temperature where appropriate). Longer gaps are retained as missing and flagged to avoid introducing misleading trends.

4.6.4 Time Alignment and Resampling

Multi-sensor streams are resampled to a uniform interval to create a synchronized feature matrix. All timestamps are converted to a unified time standard (UTC) and aligned to fixed bins (e.g., 1-minute or 5-minute). Resampling is followed by interpolation where needed.

4.6.5 Normalization and Leakage Prevention

Normalization (min–max scaling or z-score standardization) is fitted on the training split only to avoid leakage. Scaling parameters are stored with the model artifact to ensure consistent preprocessing during inference.

4.7 Feature Engineering

In addition to base sensor measurements, derived features are computed to capture system dynamics:

- **Time features:** hour-of-day and day-of-week encodings for periodic effects (e.g., ambient temperature cycles).
- **First differences:** Δfeatures (e.g., ΔpH, ΔEC) per run to capture rate of change.
- **Exponentially weighted moving averages (EMA):** short-horizon smoothing for each parameter, capturing recent state while retaining responsiveness.
- **Rolling statistics:** rolling mean and rolling standard deviation (optional) to characterize local volatility.

Feature computation is performed per run identifier to avoid cross-run contamination.

4.8 Predictive Modeling

The analytics layer supports two tasks: (i) forecasting and (ii) decision support via recommendation/label prediction.

4.8.1 Forecasting

Short-horizon forecasting predicts near-future values of water parameters (e.g., pH and EC) from historical windows. Candidate models include:

- Linear regression baselines (interpretable, fast),
- Tree-based regressors (e.g., gradient boosting / XGBoost) capturing non-linear interactions,
- Sequence models (e.g., LSTM) for richer temporal dependencies where sufficient data exists.

Model choice is driven by accuracy, interpretability, and deployability (edge vs cloud).

4.8.2 Dose Recommendation as Multi-Output Regression

For recommendation, the target variables are the dosing actions (nutrient A/B and pH up/down). This is modeled as multi-output regression using a strong tabular learner (e.g., XGBoost regressor wrapped in a multi-output strategy). Inputs are the engineered features at time t, and outputs represent recommended doses at time t (or over a short horizon). This provides actionable guidance rather than only parameter predictions.

4.9 Predictive Modeling and Anomaly Detection

Two categories of models will be implemented: forecasting models and anomaly detection models.

For forecasting, time-series models such as linear regression, tree-based regression, and Long Short-Term Memory (LSTM) networks will be evaluated for predicting short-term trends in pH and EC. Model selection will consider prediction accuracy, computational complexity, and suitability for deployment on edge or cloud environments.

For anomaly detection, a layered approach will be used. Rule-based thresholds will detect immediate safety violations, while statistical methods (e.g., z-score, IQR) and unsupervised machine learning models (e.g., Isolation Forest) will identify abnormal patterns in the time-series data.

4.10 Anomaly Detection

Anomalies are detected through a layered approach:

- **Rule-based thresholds:** immediate safety checks for out-of-range values and sustained violations.
- **Statistical detection:** z-score or IQR-based outlier detection on cleaned signals.
- **Unsupervised ML:** detectors such as Isolation Forest on feature vectors to identify multi-parameter abnormal patterns.

Detected anomalies are tagged with type (sensor vs process vs communication) when possible, enabling more actionable alerts.

4.11 Validation Strategy and Evaluation Metrics

4.11.1 Train/Validation/Test Splitting

Because the data is time-dependent, random shuffling is avoided. Splitting is performed chronologically and/or by run identifier:

- **Chronological split:** earlier time segments for training, later segments for validation/testing.
- **Run-based split:** complete growth cycles/runs are held out to prevent leakage across runs and to test generalization.

4.11.2 Metrics

Forecasting accuracy is assessed with:

- Mean Absolute Error (MAE),
- Root Mean Square Error (RMSE),
- R^2 (as a complementary goodness-of-fit indicator).

Anomaly detection is evaluated with:

- Precision and recall (when labeled events exist),
- Event-based qualitative evaluation (expert inspection) for unlabeled scenarios,
- False alert rate (alerts per day) to assess usability.

subsubsectionAcceptance Criteria System performance is evaluated against predefined acceptance criteria (Table **??**), including sensor accuracy, refresh latency, predictive performance, uptime, and dashboard responsiveness.

4.12 System Evaluation and Sustainability Analysis

To evaluate sustainability impact, the study estimates:

- **Nutrient efficiency:** reduction in over-dosing events and stabilization of EC around target ranges.
- **Water efficiency:** reduced need for corrective drain/refill events due to earlier detection of instability.
- **Operational efficiency:** reduced manual measurement frequency and faster response to anomalies.

Where direct measurement is feasible, input consumption (water, nutrients) and time-to-detection for anomalies are recorded. Where not feasible, scenario-based evaluation and cost modeling are applied using documented prices for consumables and estimated labor time.

4.13 Implementation Tools and Reproducibility

The prototype is implemented using ESP32 microcontrollers for sensing, MQTT for telemetry, and a Time-series database QuestDB for storage. Analytics and model development are performed in Python using NumPy, pandas, scikit-learn, and XGBoost/TensorFlow where applicable. Model artifacts (trained model, feature list, preprocessing parameters, and metadata) are versioned and stored to support repeatable inference. Experiments record dataset version, random seeds, train/test splits, and hyperparameters. This supports reproducibility and enables systematic comparison across modeling approaches.

5 System Design Phases

1. **Requirements Analysis:** Identify functional and non-functional system requirements.

2. **Architecture Design:** Define data flow from IoT sensors to AI models and visualization tools.

3. **Prototype Implementation:** Develop a sensor network and data processing pipeline.

4. **Model Development:** Train predictive models for pH and EC forecasting using regression and anomaly detection.

5. **Validation:** Compare results with baseline manual monitoring to evaluate efficiency and sustainability.

5.1 Data Sources

Sensor data (EC, pH, temperature, ORP, dissolved Oxygen) will be collected via IoT nodes or simulated using open datasets if physical setups are unavailable. Data preprocessing will include normalization, outlier removal, and time-series smoothing.

6 Prototype: Purchase Order

Technical Indicators	**Item**	**Price (€)**
Module power supply: 5.00 V; Measurement range: -2000 mV to 2000 mV; Output voltage: 0–4 V; Measuring temperature: 5–70 °C; Accuracy: ±10 mV (25 °C); Response time: ≤20 s / ≤1 min; ORP interface: BNC; Module size: 35 × 26 mm	Oxidation–Reduction Potential Sensor (ORP) 51 / STM32 Water Quality Electrode + Sensor	54.99
Module power: 5.00 V; Measuring range: 0–14 pH; Measuring temperature: 0–80 °C; Accuracy: ±0.1 pH (25 °C); Response time: ≤1 min; BNC connector; Gain adjustment potentiometer; Power indicator LED; Module size: 43 × 32 mm	pH Electrode Sensor Probe 0–14 pH Meter	15.25
Interface: RS-485 with automatic TX/RX control; Baud rate: 110–256000 bps; Power supply: 3.0–30 V; Logic levels: 3.3 V / 5 V; Transmission distance: ≤800 m; Multi-node support: up to 10 devices; Built-in 120 Ω termination; EMI/EMC, surge and lightning protection; Operating temperature: -40 to +85 °C	TTL to RS485 Converter (3.3 V / 5 V) for Arduino	1.34
Supply voltage: 9–24 V DC; Measurement range: EC 0–4400 μS/cm, TDS 0–2000 ppm; Output: 4–20 mA, 0–5 V, 0–10 V, RS485 Modbus RTU; Accuracy: ±2% FS; Automatic temperature compensation	EC Transmitter / TDS Conductivity Sensor Module	49.39
WiFi + Bluetooth dual-core microcontroller; Ultra-low power; 30-pin breakout board	ESP32 Development Board (ESP-WROOM-32)	3.63
Working voltage: 5 V (USB-C); Flash: 32 Mbit; WiFi: 802.11 b/g/n; Bluetooth 4.2 BR/EDR + BLE; Interfaces: UART, SPI, I^2C, PWM, ADC, DAC; Size: 54 × 28 × 10 mm	ESP32 DevKit V1 (CP2102 / CH340)	5.45
12 V, 3 A DC regulated power supply	Power Supply Adapter	11.88
ABS electrical enclosure	ABS Electrical Box	11.24
Input: 4–40 V; Output: 1.25–37 V adjustable; Step-down regulator	LM2596 Buck Converter Module	3.45
80 × 20 mm PCB (1 pc)	PCB Board	0.83
M3 male–female hex standoffs (16 pcs) + M2 standoffs (12 pcs)	Mounting Hardware	1.05
Basic PLA filament, 114.7 g (10 € / kg)	3D Printing Material	1.15
	TOTAL	**158.31**

Excluded for learning basic electrical practices: Basic Starter Kit for ESP32 IoT Development — 9.49 €

6.1 Probe Holder Prototyping

6.1.1 Environmental Constraints

When it came to designing the holder for the probes, it came down to how it could be used. The idea was to create a floatable holder for probe or mountable in a large pipe. The initial though processes was on the materials ability to endure water or chemical exposure for long durations and be food grade safe for bio conscious growers. The material for 3d printing the final product should be PETG. The prototypes were printed in Normal PLA for testing the product as a whole. Whether the probes are easily installable, if the device floats under the weight of the probes, and how it overall functions on water.

6.1.2 3D Design Process

Figure 1: Caption describing the image.

The first prototype was created within FreeCAD, which is not a user-friendly program. The design was a wide two piece cylinder like part that one recesses inside the other. As the smaller recesses inside the larger part a o-ring was to provide sealing. Making the inside an air pocket to compensate probe weight so that the holder would float. The wholes were precisely the size of the wholes for the probes. This lead to it being difficult for inserting. The whole were adjusted through a rat-tail file in order to be able to install all probes. The sealing method of the o-ring, the gap didn't compensate enough for the o-ring. As shown in the picture prototype1, the sleave no longer would recess. The device did not float. A internal silicone membrane was then glued to the inside to provide better sealing of air. The device still did not float. The idea is take it back to the drawing board.

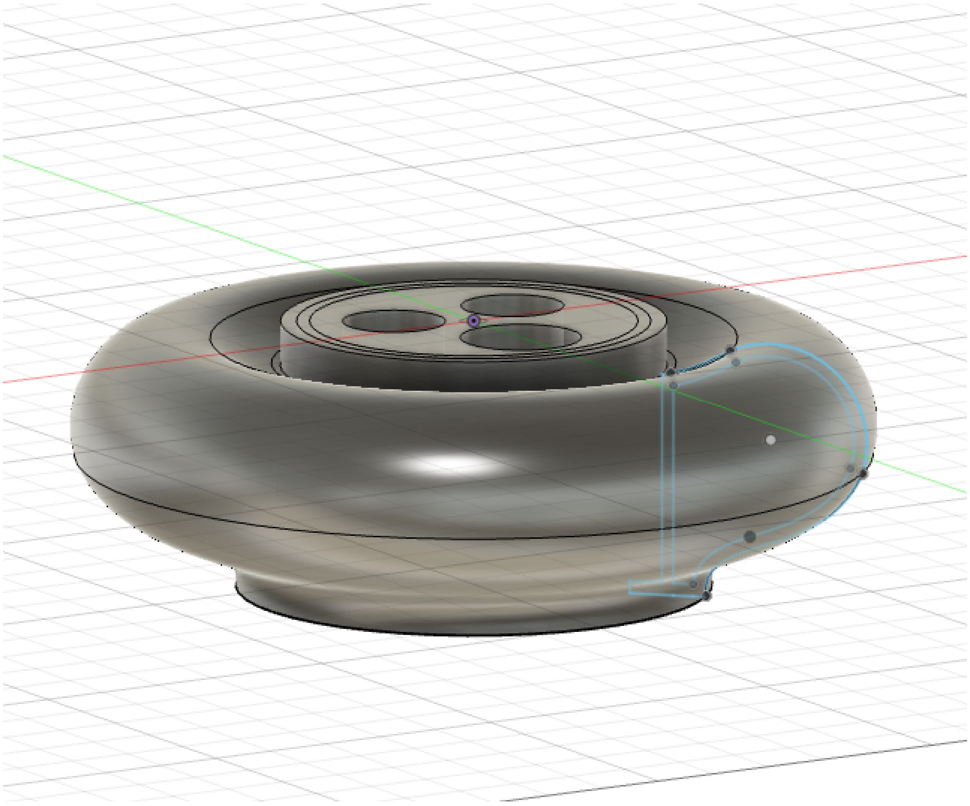

Figure 2: Caption describing the image.

As for the second prototype, it was created within fusion with the first prototypes flaws taken into account. Fusion has the ability to adapt itself to what a user is familiar with if they have solid works experience. This led to a more user-friendly environment. By decreasing the core sleeve, the need for being air tight was removed. Installation became easier. Creating a separate part as the floating method allows for ease of probe removal, and ensured the air tight nature to float. Testing is featured in the two pictures below. The prototype can float its weight of 114,7 grams.

Figure 3: Caption describing the image.

Figure 4: Caption describing the image.

6.1.3 Electrical Principles of IoT

7 Expected Results

The expected outcome is a prototype that:

- Accurately measures and predicts water quality parameters.
- Provides real-time alerts for anomalies.
- Optimizes resource usage, reducing water and nutrient waste.
- Offers a user-friendly dashboard for decision support.

Preliminary literature suggests that AI-driven hydroponic management could achieve up to 15–30% water savings and 10–20% improvement in yield consistency.

8 Discussion

The integration of IoT and AI technologies in hydroponic systems has strong potential for advancing sustainability in agriculture. However, challenges remain regarding sensor calibration, data reliability, and the computational cost of AI inference on low-power devices. Future developments may focus on edge AI and federated learning to make these systems more accessible and privacy-friendly.

9 Conclusion

This study contributes to understanding how AI and IoT can be combined for intelligent water management in hydroponic farming. Through predictive analytics and real-time monitoring, growers can improve decision-making, reduce resource waste, and enhance system resilience. The findings aim to inform both academic research and practical innovation in sustainable agri-tech.

10 Appendix

A Functional Requirements Specification

A.1 Introduction

This document defines the functional and non-functional requirements of a hobby-scale hydroponic monitoring system using ESP32-based sensing and predictive analytics. The system focuses on monitoring water quality parameters and providing decision support through visualization, alerts, and forecasting.

A.2 Functional Requirements

A.2.1 Sensor Measurement

1. The system shall measure pH at configurable intervals.
2. The system shall measure electrical conductivity (EC).
3. The system shall measure water temperature.
4. The system shall measure dissolved oxygen (DO).
5. The system may optionally measure oxidation-reduction potential (ORP).
6. The system shall timestamp each measurement.
7. The system shall store sensor calibration metadata.

A.2.2 Data Storage

1. The system shall store sensor readings locally.
2. The system shall support optional cloud synchronization.
3. The system shall buffer data during network outages.
4. The system shall retain a minimum of 30 days of historical data.

A.2.3 Visualization and Alerts

1. The system shall provide a web-based or mobile dashboard.
2. The dashboard shall display real-time and historical data.
3. Users shall be able to configure alert thresholds.
4. Alerts shall be delivered via email, SMS, or push notifications.

A.2.4 Analytics and Prediction

1. The system shall apply anomaly detection to sensor data.
2. The system shall generate short-term forecasts for pH and EC.

3. The system shall estimate potential nutrient imbalance risks based on trend deviations.
4. The system shall provide daily summary analytics.

A.3 Non-Functional Requirements

A.3.1 Reliability

The system shall operate continuously with at least 95% uptime and less than 5% data loss.

A.3.2 Scalability

The system shall support additional sensors and long-term cloud data storage.

A.3.3 Security

Communication between system components shall use encrypted protocols, and user access shall require authentication.

A.3.4 Energy Efficiency

The system shall support low-power operation and battery backup.

A.3.5 Environmental Resilience

Hardware components shall withstand high humidity, water exposure, and operating temperatures between 5°C and 40°C.