|  |
| --- |
| Damion Joyner  4-6-2022  Supervisor: Sarina Till |

|  |
| --- |
| RPDA Assignment |
| Hydroponics Monitoring and Dosing System using IoT and Artificial Intelligence to Grow Swiss Chard Spinach |
| A South African Approach |

Contents

[1. INTRODUCTION 3](#_Toc108015652)

[2. LITERATURE REVIEW 6](#_Toc108015653)

[2.1 Theoretical approach 6](#_Toc108015654)

[2.2 Conceptualisation 6](#_Toc108015655)

[2.3 Review of previous research 8](#_Toc108015656)

[2.3.1 Predictive and Prescriptive Data Analytics in Smart Farming 8](#_Toc108015657)

[2.3.2 Monitoring and Adjusting Hydroponics Systems using Smart Farming and Data Analytics 10](#_Toc108015658)

[3. METHODOLOGY 12](#_Toc108015659)

[3.1 Experiment 12](#_Toc108015660)

[3.2 Data Collection 14](#_Toc108015661)

[3.3 Data Analysis 14](#_Toc108015662)

[3.4 Model Evaluation and Success Metrics 16](#_Toc108015663)

[4. Proposed Contributions 17](#_Toc108015664)

[5. References 18](#_Toc108015665)

[6. Budget 20](#_Toc108015666)

[7. Timeline 20](#_Toc108015667)

# Abstract

This paper analyses the capabilities of an AI trained machine learning model to assist an IoT based smart hydroponics system in producing higher yields of Swiss Chard Spinach. IoT devices are used to measure levels of the hydroponics system (pH, EC, temperature, humidity) which are given as inputs to the model which then provides actions to be taken by other IoT devices to adjust these levels to provide optimal growing conditions for the crop.

The purpose of the research is to provide insights towards sustainable food growth and provide methods of increasing food growth to mitigate global hunger, both UN Sustainable Development Goals.   
To answer these questions a fuzzy logic model was developed to accurately determine adjustment outputs, based on inputs received from IoT sensors concerning the nutrient solution and environment conditions that a crop was exposed to. Further research was also proposed, to provide a direction as to the next steps needed to develop a completely autonomous intelligent IoT based hydroponics system.

We further propose the utilization of technology futures to mitigate the high initial costs of such a system.  
Our results showed that a fuzzy logic model has a precise ability to make adjustments to a hydroponics system, given the environmental factors and nutrient solution levels that the system is exposed to.

# INTRODUCTION

This research aims to determine whether, with the use of IoT, a fuzzy logic AI model will yield any significant cost improvements to a hydroponic based smart farming system when used to automate the required processes.

Soil degradation, land appropriation for farming and continuous urbanisation are a few key factors contributing to the modern environmental breakdown of our planet. Hydroponics is a method of farming where nutrient rich water is continuously pumped over the roots of a plant, thus removing the need for soil and soil regeneration processes and opening up the possibilities of where the farming must be conducted, such as rooftops, indoors or terraces. By leveraging the benefits of IoT and AI with the use of pre-existing hydroponic techniques in a multitude of situations the amount of food grown globally could increase drastically. Not only this, but the techniques mentioned are increasingly more sustainable than traditional methods of farming.

Crop farming in South Africa is typically conducted in the traditional method of growing the selected plant in the soil of a field of a varying size dependant on the type and scale of the farm. Irrigation of these crops in areas that do not have a reliable rainfall can be expensive and tedious. South Africa is also mirroring the global trend of rising food costs mostly due to the rising costs of energy and water prices which are both input resources that are administered by government policy [1]. Although South Africa is nearly food self-sufficient in almost all major food products, the level of food securities is declining rapidly because of year-on-year rise in inflation rates for food items, specifically, from September 2010 to January 2012 this rate rose from 1.2% to 10.3%. According to the Development Bank of Southern Africa in 2011, 60% of local South African households are food insecure [1]. For this reason, South Africa is currently evaluating energy and water options with a focus on the encouragement of energy sustainable solutions.

One such sustainable solution being commonly adopted is that of hydroponics. “Hydroponics is put forward as a solution to combat climate change, to reduce the environmental damage and species extinction caused by overexploitation and intensive farming” [2]. Hydroponics also allows for a more rotational use of water, a less labour-intensive farming process and an easier to control environment, all factors which contribute greatly to food securities, especially in underdeveloped countries such as South Africa. Nutrient Film Technique (NFT) is a hydroponics method of growing crops without the use of soil. It allows farmers to grow vertically and in confined spaces, reuse water and exert more control over the conditions the crop experiences over its growth cycle.

Smart farming is the use of Internet of Things (IoT) devices to monitor and automate one or more farming processes. “It is an emerging concept that refers to managing farms using technologies like IoT, robotics, drones and AI to increase the quantity and quality of products while optimizing the human labour required by production” [3]. Modern technologies are also applied to sense and adjust elements such as temperature, humidity, water flow and nutrient levels to provide a crop with optimum conditions to ensure maximum growth.

Frankenfield (2021) defines Data analytics as the science of analysing raw data to make conclusions about that information. It is becoming increasingly more commonplace for organisations to include these techniques along with traditional business analytics techniques to better understand their business, make better decisions, and improve performance [4]. This includes the farming sector, from a business analytic point of view to the use of machine learning models to improve a specific farming process.

“Fuzzy logic is an approach to computing based on "degrees of truth" rather than the usual "true or false" (1 or 0) Boolean logic on which the modern computer is based” [5]. In machine learning (ML) and artificial intelligence (AI), fuzzy logic (FL) is used to imitate human understanding of variable degrees rather than strictly binary cases of truth. FL uses the 1 and 0 of a binary situation as the extremes but allows for intermediate degrees of truth between these boundaries. FL is commonly used in engineering to regulate outputs from systems which require multiple input/parameters that may exist in a range without clear certainties or uncertainties such as temperature control systems. For this reason this paper intends to determine whether a FL AI model would be suitable to control a hydroponics system with a specific crop in mind.

Unfortunately, both hydroponics and smart farming currently have high initial costs generally attributed to the amount of hardware required in the form of tents, holding bins for the plants, pipes, pumps, IoT devices etc. However, this can be addressed in the study of Technology Futures, which is the identification, evaluation and growth of potentially high impact emerging technologies [5]. Porter et al, (2003) would argue that an emerging technology cannot be disregarded on its current feasibility alone, and a Technology Futures Analysis may conclude that continued research and adoption of the technology would lead to a decrease in the cost benefit ratio of the technology itself.

Hydroponics also requires a significant human contribution for monitoring and maintaining required pH, humidity, temperature and light levels.

The purpose of this research is to reduce the level of human effort required by automating the monitoring and adjustment of the aforementioned levels by implementing a smart/IoT system which after recording inputs from sensors, will conduct adjustment actions based on feedback from a machine learning model.

Depending on the success of the model the model could be adopted in a plethora of situations. Many South African businesses already make use of hydroponic systems, adapting these systems to use a proven IoT and machine learning model would result in significant yield and monetary improvements. South Africa has a substantial number of subsistence farmers living in rural areas, a successful smart hydroponics system, which would likely need to be initially subsidised, could produce more food, require less water and less effort. Many people living in urban areas do not have the time or space to maintain a personal vegetable garden, a near self-sustaining hydroponic option would likely be adopted by many people wanting to grow their own plants. These implementations would not only add to national food production, but all ultimately directly contribute to a decrease in fertile land requirements around the country and therefore indirectly contribute to environmental sustainability.

In 2015 all members of the United Nations, including South Africa adopted an Agenda for Sustainable Development. At the core of this agenda are 17 Sustainable Development Goals (SDGs) [6]. With the above-mentioned benefits of a successful intelligent IoT based hydroponics system in mind, we can see that the implementation of such directly aligns with multiple UN SDGs, namely Goal 2: Zero Hunger, to help achieve food security and promote sustainable agriculture, Goal 12: Responsible Consumption and Production, to ensure sustainable production patterns, and Goal 15: Life on Land, which aims to protect terrestrial ecosystems and halt or reverse land degradation [6].

# LITERATURE REVIEW

## 2.1 Theoretical approach

The theoretical approach of this paper is that of Technology Futures. This is the process of “identifying, evaluating, and growing technologies that have the potential for significant impact” [5]. It is also the understanding that while a specific technology may not prove affordable or viable at the time a study is conducted, this does not dictate that further research into the technology will not be beneficial in the future. An example to put technology futures into perspective is the commonly cited statement from IBM chairman and CEO of the time Thomas J. Watson Jr. who in the 1940s predicted “a global market for about five computers” [7]. Of course, today we know how incorrect this statement was, however it took time and investment from pioneers such as John Mauchly or manufacturers like Hewlett-Packard who believed that this new technology had a significant role in the future of mankind, to springboard computers into the commonplace item they are now.

## 2.2 Conceptualisation

**Hydroponics** - A technique for farming certain plants in culture solutions rather than in soil [8].

**Nutrient Film Technique (NFT)** – A hydroponic technique where nutrient rich water is recirculated past the plants bare roots [8].

**Internet of Things (IoT)** – The network of physical objects with sensors and software that communicate to exchange data [9].

**Arduino** – Open-source hardware and software, generally in the form of single-board microcontrollers, easy to connect input sensors and simple scripts for logic [10].

**Electrical Conductivity (EC)** – The ability of a solution to pass an electrical current due to dissolved salts and inorganic chemicals, giving an indication of salinity [11].

**pH** – The measure of how acidic or basic water is ranging from 0 to 14 [12].

**Artificial Neural Network (ANN)** – A network of nodes, or artificial neurons, which can communicate with all other nodes in the network. The signal sent to each node has a specific weighting which adjusts as the model learns [13].

**Recurrent neural network (RNN)** – A class of ANN where the connections between nodes form a graph and inputs are stored in memory to allow for sequential data to be processed over time [14].

**Long Short-Term Memory (LSTM)** – An extension of RNN which extends memory use by deciding which data is pertinent and which can be forgotten [14].

**K-Nearest Neighbours (KNN)** – Simple supervised machine learning algorithm which can be used for regression or classification and identifies an observation based on the identity of its nearest neighbours [15].

**Fuzzy Logic** – The concept in computing with a logical approach of degrees of truth as opposed to Boolean logic of true or false [16].

**Deep Neural Network** – An ANN with multiple hidden layers between input and output and can model complex non-linear relationships [13].

**Regression** – A technique of modelling which estimates the relationship between a dependant variable and one or multiple independent variables [15].

**Classification** – The process of classifying a given set of data into classes [15].

## 2.3 Literature Review

Prior research with similar intentions conducted in various locations around the world indicate three common trends when implementing machine learning models for use in IoT controlled hydroponics systems, predictive analytic models used to forecast crop yield, prescriptive analytic models used to assist the farmer in decision making and a combination of models used to monitor and adjust hydroponics systems automatically.

### 2.3.1 Predictive and Prescriptive Data Analytics in Smart Farming

S. Bhatt, D. Mody and R. Rao conducted research [8] that was broken up into three parts, the first was on gathering sensor data for air and water temperature, pH, and light intensity sensors. Secondly was a control system which monitors the data from the sensors and passes it to the third part, a machine learning algorithm in the form of an Artificial Neural Network [8]. Like the other studies, sensors gather data, the microcontroller passes the data through a machine learning model and adjusts levels accordingly. The reason Neural Networks were chosen as the as the Machine Learning model is because of the non-linearity of the biological processes involved in hydroponic systems. Other estimation models were not deemed appropriate.

The NN model consisted of 3 layers, input, hidden and output layers. 2 architectures, a 1 hidden layer and a 2 hidden layer topology, were trained and tested with a multitude of training algorithms and compared. A 1-HL NN with 9 hidden nodes trained with the quasi-Newton backpropagation algorithm was selected.

In another paper [9] the authors simply used pH, humidity, temperature, water flow and light intensity sensors to communicate with a Recurrent Neural Network which provided predictions and suggestions of how to adjust the hydroponics system. Humidity and temperature could then be adjusted manually by altering the water flow and using a sprinkler [9].

RNN was used as it retains a history of past mappings of inputs to outputs and therefore suggestions become more robust over time. It is the authors intention to use the error values from the RNN model to automatically adjust the hydroponics system in the future.

In a similar fashion, a study [10] done in India proposed a prototype where tomatoes grown in a hydroponics system would be monitored with various input sensors. By passing the input parameters through a deep neural network an action could be classified and either the farmer would be alerted to take action or automated aspects of the prototype could be performed [10].

F. Balducci, D. Impedovo and G. Pirlo took a different approach for their research [11], rather than using the live data from sensors to train a machine learning model, used data from pre-existing datasets to conduct and prove certain proposed implementations with the intention of giving direction to agricultural business of where to invest efforts [11]. One dataset was used to develop a forecasting model of crop yield, another was used to detect faulty hardware sensors and forecast and reconstruct erroneous data, while the last was used with prediction in mind, taking specific metrics of different culture species into account. Various supervised machine learning models were developed, including KNN, Decision Trees, neural networks and polynomial predictive models.

A study [12] was done in Thailand to simply measure the accuracy of machine learning models in predicting the crop yield of lettuce given certain environmental input parameters [12].

The models were trained on data from an earlier study gathered over a year. Input variables considered were: “light intensity, humidity, air temperature, water temperature, EC, and pH value of nutrient solution” [12].

Multiple different linear models were developed to have a comparison of accuracy including SGD Regression, Bayesian Ridge, ARD Regression, SVR, MLR and ANN, with SVR generally proving the most accurate.

Experiments were then conducted over a 6-week growth cycle, with plants being measured weekly and actual yield being compared to predicted yield.

Similarly, another study [13] performed no automation, but rather used sensor data to the predict crop yield of hydroponically grown lettuce [13]. Four machine learning models were trained namely, support vector regressor (SVR), extreme gradient boosting (XGB), random forest (RF) and deep neural network (DNN). Data was cultivated from three different hydroponic techniques: nutrient film technique, pyramidal aeroponic system, and tower aeroponic system. The models were also evaluated using three scenarios with varying combinations of input parameters. DNN using leaf number, dry weight and water consumption was the preferred method and proved excellent at predicting fresh lettuce yield.

### 2.3.2 Monitoring and Adjusting Hydroponics Systems using Smart Farming and Data Analytics

T. Kaewwiset and T. Yooyativong conducted a study [14] in Thailand with the purpose of automatically adjusting the electrical conductivity (EC) and pH levels of a hydroponics system using a micro-controller implementing an equation derived from linear regression.

“EC of the nutrient solution is a good indicator of the amount of available ions of salt concentration in the root zone” [14]. Essentially EC is a measurement of the value of essential nutrients in the solution feeding the hydroponics system and pH is the measurement of acidity or alkalinity of the same solution.

The method was to measure the EC in the solution, use linear regression to determine the amount of fertiliser to add for adjustment and slowly add it using a microcontroller and solenoid valve. The same was done for pH, by adding nitric acid. The solution was measured every 10 seconds and pH and fertiliser was released in 1 second intervals.

The experiment was to determine whether the model could keep the solution within the required range of EC and pH for green and red oak lettuce. The model managed an 80.8% accuracy for EC, and 95% for pH.

Again, in a study [15] conducted in Indonesia, we see again the intention of automatically adjusting the EC and pH levels of the solution with the addition of the flow of the nutrient solution through the hydroponics system. A 12-rule fuzzy logic model was also implemented as opposed to a neural network [15].

The paper addressed the use of an NFT system and used 2 pumps to adjust the pH up and down and 2 pumps to adjust the EC up and down.

The fuzzy logic model considered 2 input variables (pH and EC) and 4 output variables (4 pumps). The variables were fuzzified to determine the degree of membership. The Tsukamoto model was used to set the logic rules, which evaluates when pumps should be off or on based on EC and pH levels.

The model was able to adjust the EC and pH levels of the nutrient solution successfully, however the study results focused on response times for correcting the solution when levels were outside of the required range which was around 15 minutes.

Another study [16] in India proposed a network of soil moisture and soil temperature sensors and irrigation mechanisms in a field crop, controlled by a central microprocessor using a KNN model [16].

The inputs were used to classify the soil in a specific area into varying degrees of wet or dry and adjust accordingly. Future intentions are to include spraying the crops with appropriate chemicals.

A paper [17] done in the Philippines discussed a 4-step process to using IoT and machine learning in a hydroponics system [17]. The first step was the development of a network of sensors which monitored and gathered information of the hydroponics system. Step two was to “generate a predictive analysis model using Bayesian Network that automates the hydroponics farm system” [17]. The third was developing a user interface that the farmer could access for monitoring and controlling outputs and the last step was to test and analyse the model performance and improvement gain versus a manual farming process. The automated system outperformed the manual process in every aspect.

The final piece of literature [18], also conducted in Indonesia, proposed a similar system as previously discussed papers, where pH and EC of a nutrient solution in a hydroponics system were monitored and adjusted using suggestions from a forward chaining decision tree model controlled by IoT devices [18]. The study also included the control of a cooling fan to adjust temperature levels inside the hydroponics system. The authors performed 10 experiments to compare the adjustments made by the decision tree to those recommended by an expert and achieved a 100% match.

# 3. METHODOLOGY

## 3.1 Experiment

As stated previously the intention of the experiment is to develop a fuzzy logic AI model which will read inputs from a hydroponics system with the use of IoT sensors, and provide suitable output back to a microcontroller, which will then be able to make adjustments to the system to provide an optimal growing environment for Swiss Chard Spinach. In order to better understand how the model will operate a description will first be given of the proposed hydroponics system and IoT network that the model is intended to operate. An explanation of the methods used to develop and evaluate the model will then be given.

### Hydroponics System

The proposed system to grow Swiss Chard spinach will consist of a 72-hole NFT Hydroponics grow stand, made of food grade PVC with 4 x 115cm tiers. A large water sump will hold the solution for the system and a submersible pump will feed water through the system (Fig 1.).

Figure : Hydroponics Grow Stand

As the study is to determine how beneficial a machine learning model will be in assisting the hydroponics system, the environment in which the proposed system is housed should be as controlled as possible. For this reason, the system will be enclosed in a 1.2m x 2.4m x 2m nylon grow tent with 2 x extraction fans and an array of adjustable LED light strips. This will allow for the control of light levels, temperature, and humidity of the environment as well as limit the affect pests have on the yield.

### IoT Network

To monitor the system, we propose a network of input sensors: humidity and temperature sensor (DHT22), water pH monitor (Analog pH Sensor), water electric conductivity meter (E-201), light sensor (Analog Ambient Light Sensor), and a water flow sensor (YF-B6 Sensor), which read data from the environment and pass it to a microcontroller (ESP8266 WeMos D1 R2). The microcontroller will in turn pass the data to a cloud-based data processing platform ([[1]](#footnote-1) Thingspeak), which both makes the sensor data available to the user on a web-based dashboard and passes the data to the Fuzzy Logic AI model which will determine what actions need to be performed to remedy anomalies in the environment, see figure 2.

The last step in the process is the microcontroller receiving the output from the Neural Network and feeding instructions to the relevant environmental control outputs: humidity (Extractor Fans), temperature (LED lights), solution pH (pH Control pump), and the solution dissolved solids (Solution Control Pump). Sensor reading will occur every minute, and adjustments will be made in intervals of 5 sensor readings to give the solutions EC and pH time to stabilise after an adjustment.



Figure : IoT Diagram

### Fuzzy Logic AI Model

As stated earlier fuzzy logic is used to assist systems in providing an output based input parameters, so the first step in developing the FL AI model was to determine what inputs are available and what outputs they should affect. As shown in figure 3, the pH, Electronic Conductivity, Air Temperature, and Humidity percentage are inputs that are available from the hydroponics system. The outputs required are pH Control (whether to add acid or alkali), EC Control (whether to add salts or water), Lights and Fans, both of which can be adjusted to affect both humidity and temperature.



Figure : Inputs and Outputs for FL AI Model:

Figure 3 also visually indicates which inputs effect which outputs; this is important when establishing rules for the fuzzy logic model at a later stage.

Once the inputs and outputs are established, respectively known as antecedents and consequents in fuzzy logic, it is necessary to determine the membership functions of each. This is the process of fuzzification or establishing at what level an input or output “Crisp” value the model should associate with the corresponding associated human vernacular, known as linguistic variables. In the case of our inputs, the linguistic variables would be ‘High’, ‘Low’ and ‘Optimal’, whereas those for the outputs range depending on the corresponding control mechanism. To get accurate values for what is deemed ‘Optimal’ for the different input variables multiple sources were consulted, including farmers who had experience with growing spinach, research papers and websites which listed optimal conditions for growing various crops.

The following tables show the linguistic variables and their crisp value ranges, known as universes in fuzzy logic, of the antecedents and consequents.

Table :Antecedent Value Ranges

|  |  |  |  |
| --- | --- | --- | --- |
| **Antecedent** | **Low** | **Optimal** | **High** |
| **pH** | 0 - 7 | 6 - 8 | 7 - 14 |
| **EC** | 0 - 2.1 | 1.8 – 2.3 | 2.1 - 5 |
| **Air Temperature** | 0 - 15 | 10 - 20 | 15 - 45 |
| **Humidity** | 0 - 60 | 50 - 70 | 60 - 100 |

Table : Light Value Ranges

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Consequent** | **Off** | **Low** | **Medium** | **High** |
| **Lights** | 0 - 1 | 0 - 2 | 1 - 3 | 2 - 3 |

Table : Fans Value Ranges

|  |  |  |
| --- | --- | --- |
| **Consequent** | **Off** | **On** |
| **Fans** | 0 | 1 |

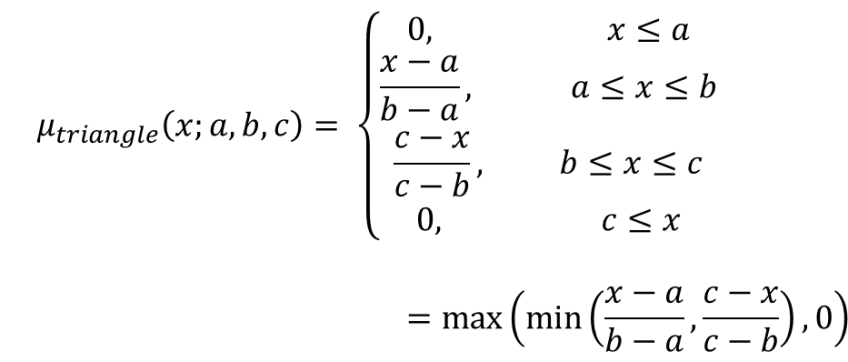
Table : pH Control Value Ranges

|  |  |  |  |
| --- | --- | --- | --- |
| **Consequent** | **Add Acid** | **No Action** | **Add Alkali** |
| **pH Control** | 0 - 1 | 0 - 2 | 1 - 2 |

Table : Solution Control Value Ranges

|  |  |  |  |
| --- | --- | --- | --- |
| **Consequent** | **Add Salts** | **No Action** | **Add Water** |
| **Solution Control** | 0 - 1 | 0 - 2 | 1 - 2 |

With the input and output variables defined and universes allocated to each of them development begins by creating membership functions for each linguistic variable for each antecedent and each consequent. A membership function is used to provide a degree of membership for a crisp input or output value so that a degree of action can be taken. This means that when the model determines that an input is low it will know exactly how low it is, and rather than just making an adjustment on a low reading it can make a precise adjustment based on the degree of membership of said reading. Although multiple fuzzy membership functions exist, all of the inputs and outputs from this system require the Triangular membership function which is defined in equation 1.



Equation : Triangular membership function

The following section shows how the membership functions were defined for the model. For readability the functions will be separated into their relative antecedent or consequent and a figure showing how the functions are mapped will be given.

#### pH Membership Functions

Equation : pH Low Membership Function

Equation : pH Optimal Membership Function

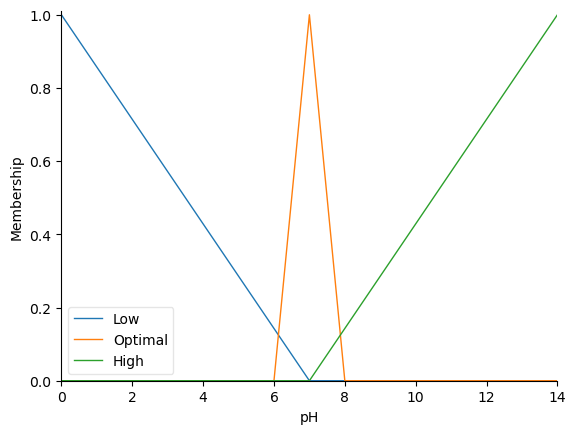
Equation : pH High Membership Function

Figure : Plot of pH Membership Functions

#### EC Membership Functions

Equation : EC Low Membership Function

Equation : EC Optimal Membership Function

Equation : EC High Membership Function

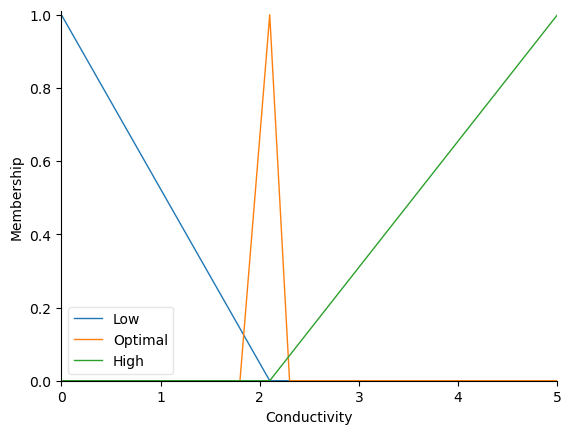


Figure : Plot of EC Membership Functions

#### Air Temperature Membership Functions

Equation 7: Air Temperature Low Membership Function

Equation 8: Air Temperature Optimal Membership Function

Equation 9: Air Temperature High Membership Function

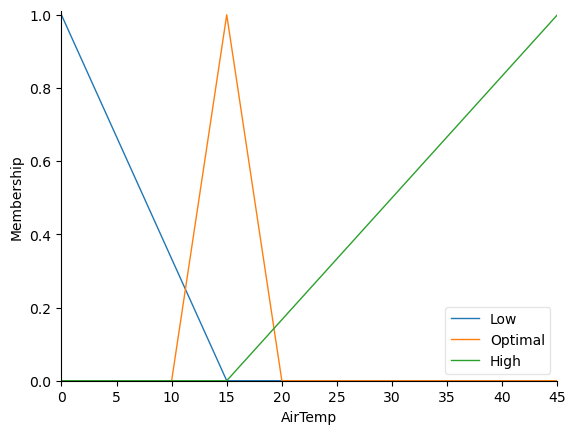


Figure : Plot of Air Temperature Membership Functions

## 3.2 Data Collection

Data will be obtained from two sources. The first being a dataset supplied by the Natural Environment Research Council and made available online [19]. This dataset is comprised of observations of both hydroponic and soil-based plants and contains relevant features such as the pH of the hydroponic solution, the biomass of the plants and different measures of soluble concentrations in the hydroponic solution.   
Solution pH and soluble concentration are necessary measures for plant health, if these levels are incorrect the spinach will not grow at all, and plant biomass is an important feature to gain a deeper understanding of the plants preferred optimal levels of pH and EC in the solution. With correct preparation and analysis of this dataset, along with an understanding of boundaries and ceilings of the solution levels required for optimal growth conditions for Swiss Chard spinach a Classification Decision Tree model will be trained to accurately identify what actions need to be performed in order to remedy the hydroponic environment, based on input from the sensors.

The second source of data is collected from the sensors themselves and the classification variable derived from the Decision Tree model. In the case of the Decision Tree and the ANN model these inputs, namely pH, EC, Water Flow, Temperature and Humidity, are live and shall be received by the models every minute, the RNN however, stores past inputs and decisions in memory as part of the learning process, meaning that a predetermined number of sensor input observations and decisions will be stored as part of a training dataset for the RNN model.

## 3.3 Data Analysis

Before utilising the data from [19] to train a Decision Tree Classification model, the data will need to be cleaned and processed. The data consists of 4 datasets, with related unique identification numbers. As we only require specific features from the datasets we will start by combining the datasets by the identification column (Sample\_name) and then dropping the columns which are not required. Columns kept from table 1 will be Ade, Zeatin and iP, which give an indication of the solutions electrical conductivity. From table 2 we require the Plant\_biomass column to give an indication of what environmental conditions produced the greatest crop yield. No features are needed from table 3 and pH is taken from table 4. Once all relevant features have been extracted, irrelevant observations can be removed. The datasets consist of observations both in hydroponic and soil environments, the soil observations will be removed as they will cause inconsistencies with the model.

Once the data has been cleaned and processed the dataset is split 70/30 and the Decision Tree model can be trained using the larger of the two splits and evaluated using the smaller. This process is repeated, applying varying techniques of analyses until accuracy of the model is optimal.

The proposed ANN and RNN models use Back-Propagation to learn over time. This essentially means that after each output that is reached an algorithm traces back to each neuron in the network that was activated to achieve this output. Each route between two neurons (edge) is then given a weighting which is continuously adjusted to result in the most accurate prediction for following iterations.

The ANN model will be adapted from [8] and consist of 3 layers, an input layer, a hidden layer and an output layer. The inputs being pH, EC, temperature, humidity, ambient light and water flow. Outputs are the adjustment of pH, adjustment of EC, fans on or off, more or less light, and water pump on or off.

The proposed RNN model is adapted from [9] and will have the same 5 inputs and outputs as the ANN model but have 2 hidden layers, each node of which contains a hidden state (see figure 3), essentially acting as a short term memory function for the models decision making process.

Figure : Architecture of RNN Model

## 3.4 Model Evaluation and Success Metrics

The Classification Decision Tree model will be evaluated in 2 stages. Firstly, once training is complete, the model will be scored and tested against the subset of cleaned data reserved for testing. After which statistical metrics can be viewed, particularly the precision and accuracy of the model in classifying a given observation. The second evaluation will be done against other classification models which serve a similar function using a different algorithm. The models will be evaluated on their f1 score, which is a measure of a model’s performance based on a combination of the model’s precision and recall.

The evaluation of the Neural Networks is also two-fold. Firstly, they are given several random or real inputs, the outputs are then compared against the expectation which is calculated manually. This evaluates the neural network on its ability to make generalised predictions on data that it was not trained with. Secondly, the Artificial Neural Network will be evaluated against the Recurrent Neural Network to ensure that output results are comparable.

Two control groups of equal number of spinach plants will be grown along side those monitored and maintained by the IoT and machine learning models. Control 1, will be exposed to the elements and grown in a regular vegetable garden, while control 2 will be grown inside the hydroponics system in a controlled environment where levels are adjusted by a farmer rather than an artificial intelligence. If successful, the study will provide conclusive results as to the degree of benefits of firstly, a hydroponics system in a controlled environment monitored and adjusted by a farmer with the assistance of IoT devices, and secondly, those of a the above, monitored and adjusted by a combination of a classification machine learning model and a neural network.

# 4. Ethical Considerations and Limitations

As mentioned previously the aim of this research is to provide an understanding of a system which is readily available and easily accessible to assist with sustainable food growth while limiting the human effort required. While ethical considerations for the conduction of the study are non-existent as the data is derived from sensors and plants and no human or personally identifiable data is used, there are limitations of the resulting systems.

While the model derived from the study will be freely available, the initial technology required is costly. This places a limitation on the system, that it will not be readily available to anyone, as a person desiring to use the system would have to acquire the components of the system at a high initial cost. Not only this, but someone wishing to perform further research on this topic based on the findings of this study, would first need to obtain their own hydroponics system and all the sensors required to perform the work.

Another limitation of the research is the technical application of the IoT can be quite complex for anyone trying to adopt the technology that has not worked with Arduinos or IoT sensors before. This has been addressed by providing detailed diagrams and a list of the IoT devices.

# 5. Proposed Contributions

It is intended that this study will contribute to the body of knowledge in a few distinct ways. Firstly, and at the forefront of the study is the proposal that Data Analytics and Machine Learning techniques can help contribute to already existing smart farming methods. The study aims to show that a classification model, trained on pre-existing data, in combination with a neural network can prove a viable option of artificial intelligence in the automation of a hydroponic farming process.

Secondly, an amalgamation of smart farming techniques is intended, based on those of prior research. If successful, this study will contribute refinements and additions to this process.

Lastly, it is the intention to evaluate and refine the process of growing one of South Africa’s most consumed and affordable vegetables, the Swiss Chard spinach. The study will accumulate a significant amount of data and repeat the growth cycle of the spinach numerous times and may produce useful insights when concluded.

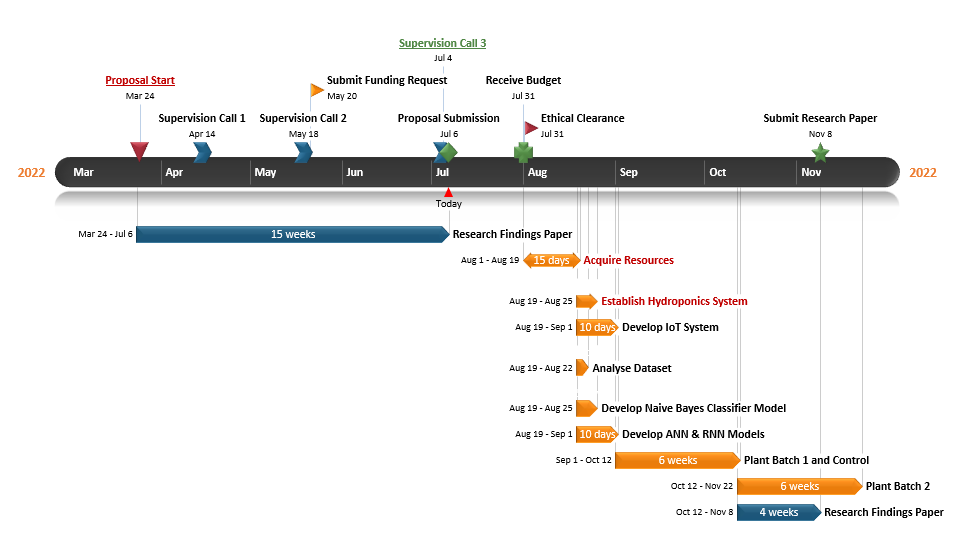
# 6. References

|  |  |
| --- | --- |
| [1] | M. Gulati, I. Jacobs, A. Jooste, D. Naidoo and S. Fakir, “The water-energy-food security nexus: Challenges and opportunities for food security in South Africa,” *Aquatic Procedia 1,* pp. 150-164, 2013. |
| [2] | “What is Hydroponics and Advantages,” 22 04 2021. [Online]. Available: https://www.iberdrola.com/sustainability/what-is-hydroponics-and-advantages#:~:text=Hydroponic%20crops%20are%20based%20on,can%20be%20grown%20at%20home!. |
| [3] | Sciforce, “Smart Farming: The Future of Agriculture,” 22 June 2020. [Online]. Available: https://www.iotforall.com/smart-farming-future-of-agriculture. |
| [4] | S. C. Albright and W. L. Winston, Business Analytics: Data Analysis and Decision Makink, Boston: Cengage Learning, Inc., 2017. |
| [5] | Mission-focused Research, “Technology Futures,” 2021. [Online]. Available: https://www.mitre.org/research/mission-focused-research/technology-futures#:~:text=Technology%20Futures%20focuses%20on%20identifying,impact%20on%20our%20sponsors'%20missions.. |
| [6] | United Nations, “Department of Economic and Social Affairs,” 2021. [Online]. Available: https://sdgs.un.org/goals. |
| [7] | R. Hickin and M. Bechtel, “Technology Futures: Projecting the Possible, Navigating What’s Next,” World Economic Forum, Cologny, 2021. |
| [8] | S. Bhatt, D. Mody and R. Rao, “IMPLEMENTATION OF IoT AND MACHINE LEARNING IN HYDROPONICS: A REVIEW,” *Journal of Emerging Technologies and Innovative Research,* pp. 387-391, 2018. |
| [9] | D. V. Saraswathy, C. Nithiesh and S. Palani Kumaravel, “INTEGRATING INTELLIGENCE IN HYDROPONIC FARMS,” *International Journal of Electrical Engineering and Technology,* pp. 150-158, 2020. |
| [10] | M. Mehra, S. Saxena, S. Sanakaranarayanan, R. Jackson Tom and M. Veeramanikandan, “IoT based hydroponics system using Deep Neural Networks,” *Computers and Electronics in Agriculture,* pp. 473-486, 2018. |
| [11] | F. Balducci, D. Impedovo and G. Pirlo, “Machine Learning Applications on Agricultural Datasets for Smart Farm Enhancement,” *Machines,* pp. 6-38, 2018. |
| [12] | S. Gertphol, P. Chulaka and T. Changmai, “Predictive models for Lettuce quality from Internet of Things-based hydroponic farm,” Kasetsart University, Bangkok, 2018. |
| [13] | A. Mokhtar, W. El-Ssawy, H. He, N. Al-Anasari, S. Sammen, Y. Gyasi-Agyei and M. Abuarab, “Using Machine Learning Models to Predict Hydroponically Grown Lettuce Yield,” Frontiers in Plant Science, Cairo, 2022. |
| [14] | T. Kaewwiset and T. Yooyativong, “Electrical Conductivity and pH Adjusting System for,” *International Conference on Electrical Engineering/Electronics, Computer, Telecommunications,* pp. 761-764, 2017. |
| [15] | D. Yolanda, H. Hindersah, F. Hadiatna and M. Triawan, “Implementation of Real-Time Fuzzy Logic Control for NFT-Based Hydroponic System on Internet of Things Environment,” *International Conference on System Engineering and Technology,* pp. 153-159, 2016. |
| [16] | Y. Shekhar, E. Dagur, S. Mishra, T. Rijo Jackson, M. Veeramanikandan and S. Sankaranarayanan, “Intelligent IoT Based Automated Irrigation System,” *International Journal of Applied Engineering Research,* pp. 7306-7320, 2017. |
| [17] | M. Alipio, A. Dela Cruz, J. Doria and R. Fruto, “A Smart Hydroponics Farming System Using Exact Inference in Bayesian Network,” Global Conference on Consumer Electronics, Manila, 2017. |
| [18] | Y. Nugraha, B. Irawan and R. Saputra, “System Design And Implementation Automation System Of Expert System On Hydroponics Nutrients Control Using Forward Chaining Method,” Asia Pacific Conference on Wireless and Mobile, Bandung, 2017. |
| [19] | A. Rampurawala, “Plant growth: Hydroponic and Soil Compound Dataset,” Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/abtabm/plant-growthhydroponics-and-soil-compound-dataset?select=Table\_1.csv. [Accessed 05 July 2022]. |
| [20] | Oracle, “What is Data Analytics?,” 2022. [Online]. Available: https://www.oracle.com/business-analytics/data-analytics/. |

# 6. Budget



# 7. Timeline



1. https://thingspeak.com/ [↑](#footnote-ref-1)