DEVELOPMENT OF A HYDROPONICS SIMULATOR TO GENERATE GUIDELINES FOR DATA COLLECTION IN HYDROPONICS FOR MACHINE LEARNING APPLICATIONS

Jonas W. Swiatek, Felix Kuthe, Loui Al-Shrouf, Mohieddine Jelali

Cologne Lab for Artificial Intelligence and Smart Automation (CAISA), TH Köln - University of Applied Sciences, Betzdorfer Straße 2, 50679 Köln, Germany

Abstract

Hydroponics, a soilless growing method using nutrient-rich water solutions, offers distinct advantages in plant growth and resource efficiency. Data-driven optimization in hydroponics relies on precise control of environmental factors. To effectively apply data-driven optimization, comprehensive and accurate data describing hydroponic plant growth is essential. In the first part of the project and in this paper, we seek to identify critical parameters that influence plant growth, recommend optimal resolutions and measurement frequencies, and address challenges in data collection. Standardization of data collection procedures and establishment of data collection guidelines are fundamental to obtaining high quality, comparable and reproducible data. Establishing a standard for data collection will allow data from different experiments to be compared and used in future research. In the initial phase of the project and throughout this paper, our primary objective is to identify the relevant parameters affecting plant growth in a hydroponic system. These critical parameters encompass aspects such as light intensity, spectral composition, nutrient solution composition, water quality, pH levels, and more. By conducting a thorough investigation, we aim to recommend optimal resolutions for these parameters. Additionally, we will address the challenges associated with collecting data related to these key variables, ensuring that the data obtained is reliable and useful to further research and scientific exploration in the field of hydroponics. Accurate data on these characteristics enables data-driven decisions and methods, such as machine learning and machine vision, which in turn allow optimization of hydroponic productivity. The study conducted a comprehensive analysis and identified 30 essential parameters crucial for characterizing a hydroponic system. The findings are summarized as a list of relevant parameters that need to be measured or recorded to enable data-driven optimization of hydroponic systems.

Keywords: *hydroponics, plant growth parameters, data driven optimization, machine learning, data acquisition, condition monitoring*

1. INTRODUCTION

Hydroponics is a cultivation method that involves growing plants without soil, using nutrient-rich water solutions. This technique offers significant advantages, such as enhanced plant growth and resource efficiency. Hydroponics is exceptionally well-suited for data-driven optimization due to the intricate relationship between various parameters and their impact on plant growth. In traditional soil-based agriculture, it can be challenging to isolate and control specific variables, as soil conditions vary significantly. However, hydroponics offers precise control over environmental factors such as nutrient concentrations, pH levels, temperature, humidity, and light. By accurately measuring and monitoring these parameters, data-driven optimization enables growers to identify and optimize the ideal conditions for plant growth. However, to effectively apply data-driven optimization, it is crucial to have accurate and comprehensive data that accurately describes hydroponic plant growth.

The initial step involves identifying key parameters that directly influence hydroponic plant growth. Accurate measurement of these parameters is essential to obtain reliable data. For this purpose, reliable and calibrated sensors are crucial. These sensors must accurately measure selected variables with high precision and resolution. It is important to ensure that the chosen sensors are suitable for the unique demands of the hydroponic environment and can withstand prevailing environmental conditions. However, the selection of sensors is not extensively covered in this study, as it depends on experiment design, duration, and budget constraints. Instead, our focus in this paper is to explore the key parameters affecting plant growth, while building upon existing knowledge regarding sensor suitability and performance in hydroponic systems.

Standardizing data collection procedures is another important aspect of acquiring good data in hydroponics. Developing standardized protocols ensures consistency across multiple measurements and different operators. Clear and well-defined data collection guidelines should be established. This helps minimize errors and variations in data collection, leading to more reliable and consistent results. Another benefit is that standardizing data collection produces large quantities of usable information that can be harnessed for Machine Learning (ML) purposes. By standardizing the data collection procedures and establishing consistent guidelines, it can be ensured that the data is reliable, comparable, and reproducible. This, in turn, enables ML researchers and practitioners to build robust models and implement ML-based applications with higher accuracy and efficiency. When data is collected in a standardized format, it becomes easier to process, analyze, and feed into ML algorithms, unlocking the potential for advancements in various fields. Whether it's for predictive modeling, computer vision, or any other ML-driven application, standardized data serves as the foundation for realizing the true potential of Machine Learning technologies. Good data describes the complete growth process of hydroponic plants and enables reproductability of the results. Accurate and comprehensive data provides a foundation for analysis and modeling.

This paper aims to discuss the essential parameters in hydroponic plant growth, their recommended resolution and the challenges associated with data acquisition.

2. MATERIALS AND METHODS

2.1 Literature review

In this paper, the identification of relevant process parameters influencing plant growth was conducted through a comprehensive analysis of existing literature and scientific papers. By systematically reviewing the available research, we aimed to identify the key parameters and variables that significantly impact plant growth in hydroponic systems. In order to find a suitable resolution for quantifying these parameters, a comprehensive analysis of the relevant scientific literature was conducted. This investigation aimed to discern the specific magnitude and nature of alterations in each parameter that exert a noteworthy impact on plant growth.. By examining the magnitude of variations and their effects on plant performance, we aimed to establish recommended resolutions for accurate data acquisition.

2.2 Hydropnic simulator

All hydroponic systems possesses distinct characteristics and operate in unique manners. Variations in behavior and substantial differences in inertia are commonly observed among these systems. Understanding the inertia of environmental changes can assist in finding suitable measurement frequencies. To gain insights into this inertia, we utilized a hydroponic simulator, allowing us to approximate and assess how rapidly environmental changes may propagate within the system.

An indoor vertical Deep Flow Technique (DFT) hydroponic system is utilized for growth assessment. Please refer to **Error! Reference source not found.** for a schematic of the hydroponic simulator components. The system is comprised of two grow layers and a water reservoir. Each grow layer possesses dimensions of 60 by 80 cm, with a water depth of 16 cm, and a total water capacity of 120 liters. Each Layer can accommodate up to 20 plants. The measurement of essential parameters, such as electrical conductivity (EC), pH, and water temperature, was conducted using an Atlas Scientific hydroponics kit. Additionally, air temperature and humidity were monitored using a DHT22 Sensor from Joy-It. For air flow two 120mm fan were used. Data collection and recording were automated through the implementation of a Raspberry Pi 3, enabling continuous logging of the acquired data. Throughout the experiment, the system was operated without regulation to study the inherent variability of the environmental parameters during plant growth assessment.

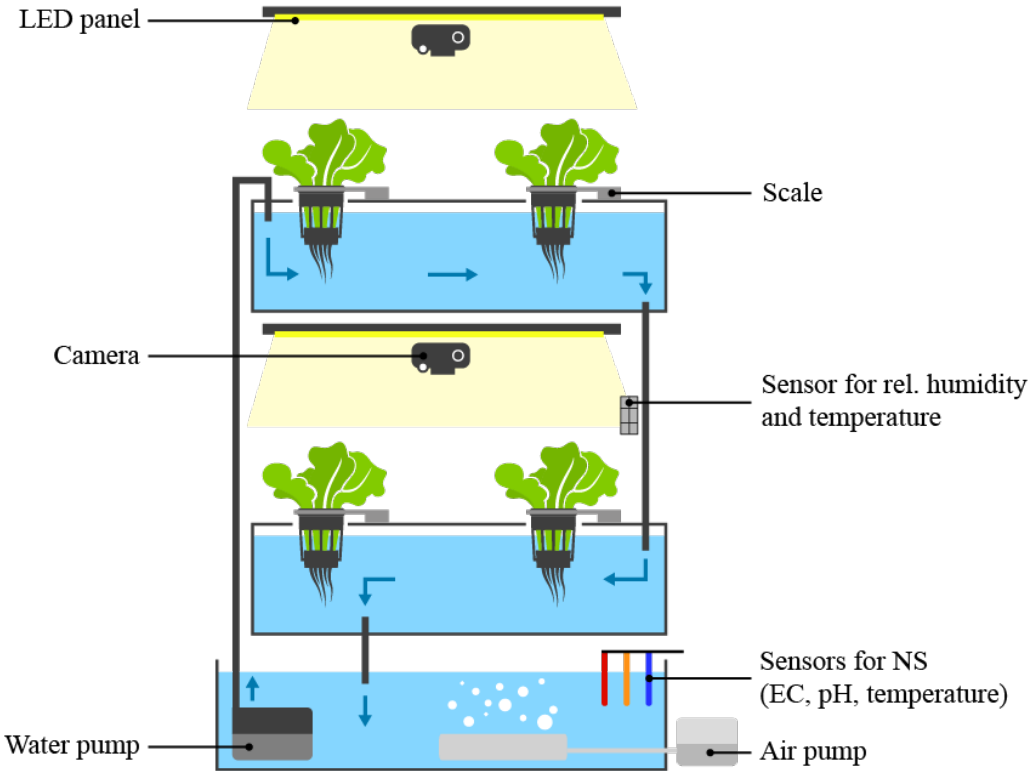

Fig. 1. Schematic components of the hydroponic simulator

2.3 Growth assessment

The assessment of success in hydroponic plant growth varies depending on plant species and project objectives. Relevant parameters used to measure success can include plant size, plant mass, number of fruits, and fruit size. Size serves as an indicator of overall plant growth and is measured by parameters, such as height, color, width, or canopy size. Plant mass reflects biomass accumulation, indicating nutrient uptake and utilization efficiency. The number of fruits harvested is crucial for yield-focused projects, reflecting successful pollination, flowering, and fruit development. Fruit size is significant for projects emphasizing quality and marketability. The selection of success indicators should align with specific plant species and project goals. Thus, choice of measurement methods depends on the specific project requirements and goals. To continuously collect data throughout the entire growth cycle of a plant, it is necessary to ensure that the measurements taken are non-invasive. Otherwise, they may introduce biases and skew the results. For this reason, only noninvasive methods of measurement are explored further. Based on the research of Chen et al. (2016), scales are used for continuous measurement of the plant weight. Visual data is collected with two Panda Latte 5MP USB cameras located above the plants.

3. RESULTS

The findings of this study are grouped into light, nutrient solution, greenhouse environment, and growth assessment. At the end of the chapter, a summary of the results is presented.

3.1 Light

Light influences plant growth in three important ways: photosynthesis, photomorphogenesis and photoperiodism. The first mechanism, photosynthesis, is a fundamental process, by which plants use light energy to convert carbon dioxide and water into glucose and oxygen. Light serves as the primary energy source, driving the biochemical reactions essential for the synthesis of carbohydrates that sustain plant growth and development (Corré, 1983). The first crucial measurement here is the Photosynthetic

Photon Flux Density (PPFD). PPFD represents the quantity of photosynthetically active radiation (PAR) reaching the plant canopy per unit area per second (in μmol· m^{-2} s^{-1}) and is a metric to evaluate the light intensity. Research conducted by Zheng et al. (2019) on strawberries demonstrated, that excessive or insufficient light intensity beyond a range of 90 μmol·m^{-2} s^{-1} can hinder photosynthetic activity and result in decreased photosynthetic activity. Similarly, Nguyen et al. (2019) investigated the effect of light intensity on hydroponically cultivated spinach, using light intensities of 90, 140, 190, and 240 μmol·m^{-2} s^{-1}. Other studies have also utilized similar resolution ranges for their research. Thus, a resolution of more than ±10 μmol·m^{-2} s^{-1} does not appear necessary for accurately evaluating the influence of light intensity on plant growth (Kang et al., 2013).

Photomorphogenesis is the process through which light influences plant morphology and development. Different wavelengths and intensities of light play crucial roles in regulating specific aspects of plant growth, such as stem elongation, leaf expansion, and branching architecture (Corré, 1983). In the study by Brown et al. (1995), 'Hungarian Wax' pepper plants were grown under different light sources. When exposed solely to red LEDs without supplemental blue wavelengths, the plants had reduced biomass compared to those under blue fluorescent lamps or Metal-halide lamp. However, the addition of far-red radiation to the red LEDs resulted in taller plants with greater stem mass. Interestingly, plants exposed to red or red and additional far-red radiation had fewer leaves than those under blue wavelengths. These findings emphasize the importance of considering the spectral composition of light in influencing plant growth. Accurate measurements of the spectral composition are essential to fully comprehend and optimize the effects of light on plant development. Goins et al. (1997) also showed that light composition does have an influence on plant growth. Understanding these interactions can lead to targeted light manipulation strategies for enhanced plant growth and productivity. The light composition of modern grow LEDs are usually provided by the manufacturer. If the information is not provided, a spectrometer can be used to determine the light spectrum. The spectrum does not need to be measured continuously, one measurement per dataset (DS) is sufficient as the spectrum does not change unless there are defects in the lighting.

The third mechanism is photoperiodism, which regulates flowering and reproductive stages in various plant species (Song et al., 2015). In the study conducted by Avgoustaki et al. (2020), the growth rate of Genovese basil plants was analyzed under two different lighting schemes. The aim was to investigate the possibility of reducing energy demand without compromising plant growth and quality. The first lighting system involved 16 hours of continuous light, while the second system had a 14-hour photoperiod with intermittent light. Both treatments maintained consistent light intensity. The results revealed that intermittent light exposure, with short ten-minute cycles during the night and four-hour periods of daylight, did not have a negative impact on the quality and quantity of basil plants. This indicates that energy consumption can be reduced through intermittent lighting without compromising plant growth and development (Avgoustaki et al., 2020). A study by Kang et al. (2013) evaluated the effects of different light intensities and photoperiods on lettuce growth. Four light intensity treatments ranging from 200 to 290 μmol· m^{-2} s^{-1} PPFD were provided by LEDs, combined with three photoperiods: 18/6 (1 cycle), 9/3 (2 cycles), or 6/2 (3 cycles) of light and dark periods. Results showed that a high light intensity of 290 μmol·m^{-2} s^{-1} PPFD combined with a 6/2 (light/dark) photoperiod resulted in optimal plant growth and development. Light intensities of 230 or 260 μmol m^{-2} s^{-1} PPFD with longer photoperiods of 18/6 and 9/3 (light/dark) also led to good growth and enhanced photosynthetic capacity. To summon this metric into one number, PPFD can be converted into the Daily Light Integral (DLI). It quantifies the total amount of PAR received by the plants over a 24-hour period and is given in mol m^{-2} d^{-1}.

Hence, it is crucial to not only record the duration of day and night but also document the specific patterns and characteristics of the light and dark periods.

3.2 Nutrient solution

The Nutrient solution (NS) that supplies the plants with all essential elements is made by dissolving fertilizer salts in water (Resh, 2022, p. 35). These elements can be provided in different chemical forms. It is important to note that not all of them have the same solubility, so it is not only the total amount of each element that is important, but also the chemical state it is in. The chemical composition of fertilizers is provided by the manufacturer. The amount of macro- and microelements needs to be documented. Additionally, the amount of nutrient solution used is relevant, as the amount of nutrient solution per plant influences the inertia of the system. If the system is replenished with additional nutrient solution during the growth cycle, this also needs to be documented.

3.2.1 Water composition

The initial dissolved minerals in the water used to prepare the nutrient solution is of importance. Minerals that are already dissolved can influence the final composition of the nutrient solution (Resh, 2022, pp. 90–91). In addition, the replenishment of lost volume of nutrient solution with water due to consumption and evaporation leads to an accumulation of salts such as sodium or chloride (Resh, 2022, p. 90; Savvas et al., 2007). It is therefore necessary to analyze the water used in the creation of the nutrient solution and test it for dissolved salts. This data needs to be documented. Furthermore, the amount of water used to replenish the system since this also introduces more salts to the system.

3.2.2 EC-value

During plant growth, it is not possible to directly monitor the exact amount of dissolved nutrients in the nutrient solution. Alternatively, the electronic conductivity (EC) gives a measure of the solution's ability to conduct an electric current. This EC value reflects the total concentration of dissolved salts in the nutrient solution (Resh, 2022, p. 83). It serves as an indicator of the solution's strength and nutrient content, where a higher EC value indicates a higher concentration of nutrients. This can be beneficial for plants during specific growth stages when they require increased nutrient availability. However, excessively high EC levels can lead to nutrient imbalances, resulting in nutrient toxicity and detrimental effects on plant health. The EC is measured in millisiemens (mS) per centimeter and ranges in practical applications for hydroponics systems between 0.2 to 4 mS cm^{-1}. In studies higher values have been used for research purposes. The resolution of EC values in past studies was not higher than 0.1 mS cm^{-1} (Kappel et al., 2021; Resh, 2022, pp. 83–84; Sublett et al., 2018).

3.2.3 pH

The pH value is another crucial parameter to monitor the nutrient solution. The pH value of the nutrient solution has a significant effect on plant growth in hydroponic system, as it directly affects the availability of nutrients and their uptake by the plant roots (Resh, 2022, p. 16). Different plant species have specific pH requirements for optimal nutrient absorption. If the pH of the nutrient solution deviates from the appropriate range, nutrient deficiencies or toxicities can occur, hindering plant growth and development. A pH that is too high or too low can limit the availability of essential nutrients, leading to nutrient imbalances and impaired physiological processes (Resh, 2022, p. 16). Wortman (2015) showed that high pH and low EC values can reduce the marketable yield of marketable Basil by up to 76%. He looked at pH levels of 5.5, 6 and 7.2. Finer Resolution are not common in research of the influence of the pH level. Other studies also did not use a higher resolution than 0.1 (Hopkinson and Harris, 2019; Kane et al., 2006; Resh, 2022).

3.2.4 Dissolved oxygen & aeration

A well-oxygenated root zone environment is vital for maintaining a healthy root system, facilitating nutrient uptake, supporting root growth, and preventing root-borne diseases. Insufficient oxygen levels in the root zone can adversely affect root and plant performance, leading to increased disease susceptibility (Chérif et al., 1997). The solubility of oxygen depends on water temperature and the partial pressure of oxygen in the air, lower temperature means lower amounts of oxygen. Suyantohadi et al. (2010) have shown that in these lower temperature environments (12°C) oxygen enrichment of the nutrient solution can lead up to 2,1 fold increase in plant growth. In addition to the amount of dissolved oxygen, the method of oxygen enrichment is also important. Here, Ikeura et al. (2018) have shown that there can be a difference in plant growth due to the type of aeration method used. Park and Kurata (2009) have shown that the application of microbubbles to hydroponics solution promotes lettuce growth. The lettuce samples treated with microbubbles exhibited significantly higher fresh and dry weights, ranging from 2.1 to 1.7 times greater than those untreated with microbubbles. In addition to that, Ikeura et al. (2018) did show that the way microbubbles are produced also effects plant growth. Abu-Shahba et al.

(2021) also found that the morphological characteristics of lettuce plants grown in hydroponics with microbubbles were significantly better than those of lettuce in the other treatments (no bubbles, macro bubbles, soil).

Based on this, it is necessary to document the position of the aeration, the method used to aerate the nutrient solution, the amount of air flow (bubbles) and the dissolved oxygen. Additional data specific to the aeration method such as used air stones should also be provided.

3.3 Greenhouse environment

3.3.1 Airflow

Airflow in hydroponics serves multiple vital functions, including uniform temperature regulation, transpiration facilitation, humidity control, and prevention of stagnant air (Resh, 2022, p. 311). Proper air circulation helps to regulate the temperature within the hydroponic system, dissipating excess heat and maintaining an optimal thermal environment for plant growth. Lee et al. (2013) showed that increased horizontal airflow over hydroponic lettuce can reduce the occurrence of tipburn due to improved transpiration and calcium transport within the plants. In their study horizontal airflow rates of 0.28 $m \cdot s^{-1}$ (Low), 0.55 $m \cdot s^{-1}$ (Medium), and 1.04 $m \cdot s^{-1}$ (high) were used. Peiro et al. (2020) conducted a recent study that demonstrated varying airflow rates in a closed hydroponic environment resulted in significant differences in the dry weight of the lettuce. Air speeds above 0.6 $m \cdot s^{-1}$ were associated with a decrease in lettuce plant growth. Since airflow can vary within a greenhouse or indoor farm, it should be measured at the canopy or as close as possible. A resolution of more than 0.1 $m \cdot s^{-1}$ is not feasible or useful in most cases due to turbulence and other factors affecting the effective airflow. As this value is usually dictated by the operator, constant measurement is not necessary. Changes in plant size and geometry may require additional measurements at canopy level.

3.3.2 CO2-level

CO2 is essential for photosynthesis, the process by which plants convert light energy into chemical energy. Increased CO2 levels stimulate photosynthesis, resulting in increased plant growth, biomass production, and yields. Adequate CO2 concentrations improve nutrient uptake as the energy produced from photosynthesis is utilized for efficient nutrient absorption. Higher CO2 levels also reduce photorespiration, a process that competes with photosynthesis and hampers plant growth. By minimizing photorespiration, plants can allocate more energy to growth and development. With access to sufficient CO2, plants allocate more resources to growth, resulting in faster growth rates, earlier flowering, and shortened crop cycles in hydroponics. While CO2 enrichment initially enhances growth in some plants, long-term exposure can lead to varied results (Makino and Mae, 1999). Certain plant species may exhibit reduced photosynthesis under prolonged CO2 enrichment (Makino and Mae, 1999). CO2 Concentration is measured in ppm. In the atmosphere, the usual concentration is around 385 ppm. Resolutions higher than 5 ppm are not commonly used.

3.3.3 Relative humidity

Humidity refers to the amount of moisture present in the air. High humidity means there is a higher concentration of water vapor in the air, while low humidity indicates a lower water vapor concentration. The main way on influencing the growth is by governing the transpiration of plant leaves. Transpiration is the process by which plants lose water vapor through their leaves. High humidity reduces the rate of transpiration, as the surrounding air is already saturated with moisture. This can lead to reduced water uptake by plants and potentially cause nutrient imbalances. On the other hand, low humidity can increase transpiration, potentially leading to excessive water loss and stressing the plants. Humidity is usually measured in relative humidity (RH), the value is given in percent. A Study on the relative influence of humidity on Lactuca Sativa "suggests that the optimal relative humidity that could improve plant growth is 85±2%. As low humidity below optimum level will increase stomatal resistance" (Chia and Lim, 2022). The resolution used here is not higher that 1%, which is consistent with other studies (Shibata et al., 1995; Vanhassel et al., 2015).

3.4 Growth Assessment

Measuring the success of a process is essential and is required at various stages. For the plant growing process, we will focus here on plant weight and image analysis.

3.4.1 Plant weight

In the past, it was a common practice to evaluate the success of a growth cycle by measuring the weight of the plant. This could only be done, usually at the final stage, since it requires the plant to be harvested and weighted (Andriolo et al., 2005; Fu et al., 2012). The process can be performed in a fresh or dry state to measure the weight of the plant with or without water content. However, this approach is inadequate for comprehensive monitoring and data analysis since a continuous measurement is required.

In this study, it was tested to plant lettuce plants in small baskets on scales that continuously measure the weight of each plant. In contrast to the study done by (Chen et al., 2016) the plants were planted in an commercially typical way so that a closed canopy is crated. This is done to maximize the yield form lighting and space. While this worked well in early stages of growth in later stages the overlapping of leaves leads to significant faults and fluctuation in the measurements. It is therefore important to ensure that the distance between plants is sufficient for the final size and shape of the plants or choose a different kind of measurement.

3.4.2 Images

Advancements in digital camera technology have introduced a new tool for measuring plant growth, which is enabled by high-resolution imaging capabilities that capture intricate plant structures. This facilitates detailed analyses of growth patterns and developmental processes. Moreover, the integration of multispectral and hyperspectral imaging techniques enables comprehensive assessments of plant health, stress responses, and biochemical composition. This allows for valuable insights into plant physiology. A noteworthy advantage is the non-invasive and continuous nature of these measurements, which enables the monitoring of plant growth throughout the entire growth cycle without disruptions.

Despite the high availability and low cost of high-resolution cameras, machine vision applications face a significant challenge due to the large amount of data generated by image capture compared to traditional numeric sensor data. In machine vision applications, the primary challenge is balancing image resolution and data storage. Higher resolution images provide finer details and more precise visual data for machine vision algorithms. However, they also require greater storage capacity and may cause resource limitations.

Given the substantial variations in ML applications and data requirements based on the specific type of ML application utilized, relying solely on past studies to establish visual data collection guidelines for hydroponics is impractical. Consequently, a theoretical approach was adopted for the analysis to address this challenge effectively.

The primary objective is to determine the best image size that meets the requirements of machine vision tasks while minimizing data storage. This can be achieved by identifying the critical resolution level necessary for accurate and reliable analysis without needlessly overloading the storage capacity. By striking this balance, researchers can harness the full potential of high-resolution imaging for measuring plant growth as well as optimizing resource allocation and cost-effectiveness.

A successful machine vision system design is based on a detailed specification and consists of several steps. The general design process we follow is given in Hornberg (2017), of which we will address and adapt here the relevant parts for hydroponic machine vision systems. The first part in this guideline is choosing the camera scan type, as this influences the following choice of hardware and image acquisition. For hydroponic systems, area scan cameras are sufficient, as they are easier to set up, control and trigger and don't require movement of the part or camera.

In order to be able to determine the resolution of the sensor, it is first necessary to define the smallest feature that needs to be identified. By understanding the smallest feature's size, we can establish a baseline spatial resolution required by the camera sensor to accurately capture and avoid losing information due to insufficient resolution. In Hornberg (2017), the spatial resolution is calculated with the size of the smallest feature S_{f} and the number of pixels to map to the smallest feature N_{f} as

$$R_{\mathrm{s}} = \frac{S_{\mathrm{f}}}{N_{\mathrm{f}}}.$$

It is important to select a suitable N_{f} for the application to ensure that the data is a faithful representation of the actual scene. This requires complying with the Shannon criterion. This principle stipulates that to accurately reconstruct a signal, the sampling frequency must be at least twice the signal's highest frequency (Shannon, 1948). For machine vision, this criterion translates to capturing the hydroponic scene with a frequency at least twice that of the smallest feature size, so $N_{\mathrm{f}} \geq 2$. The range of features of hydroponically grown plants varies widely. In addition, the importance of the features for monitoring plant growth has not yet been determined and differs between plants. Some of the smallest common features in hydroponics are new flowers, diseases and pest infestations. For example, the smallest feature that could be of interest is a broad mite infestation, as they are less than 0.2mm in length (Resh, 2022, p. 564). This would theoretically require a minimal spatial resolution of $R_{\mathrm{s}} = \frac{S_{\mathrm{f}}}{N_{\mathrm{f}}} = \frac{0.2\ \mathrm{mm}}{2\ \mathrm{pixel}} = 0.1 \frac{\mathrm{mm}}{\mathrm{pixel}}$. This can be used to calculate the necessary minimal resolution of the camera. In Hornberg (2017) the camera resolution R_{c} is calculated with the field of view (FOV) and the spatial resolution R_{s} with

$$R_{\mathrm{c}} = \frac{\mathrm{FOV}}{R_{\mathrm{s}}}.$$

The FOV is the area visible to the camera at the working distance. It dictates the extent of the hydroponic scene that the camera can capture within a single image. Achieving an appropriate FOV is essential to encompass the regions of interest while avoiding unnecessary data acquisition. This can vary between a single fruit (M. S. Tasrif Anubhove et al., 2020), to multiple plants (Lee et al., 2018). Based on the chosen working distance and the necessary resolution R_{c}, a suitable lens and image sensor can be selected, which affect crucial factors such as depth of field and image distortion.

The next step is to decide whether to use a color or monochrome sensor type. To successfully monitor a hydroponic system, it is necessary to capture the full visible light spectrum as this helps to distinguish between different plant pigments, detect disease and monitor nutrient levels. Monochrome sensors, on the other hand, tend to be more sensitive to light, making them more suitable for monitoring plant growth in low-light conditions.

Because plant growth monitoring is characterized as a relatively non-dynamic process, where changes occur gradually over time, the requirements for frame rate and shutter mechanisms are not as high as in fast-moving scenarios, such as industrial robots or production lines. A lower frame rate, such as less than 5 frames per second, is generally sufficient for most monitoring and analysis tasks in this context. A lower frame rate not only conserves storage and processing resources, but also contributes to lower power consumption, an important consideration in long-term monitoring applications.

By adapting the vision system design to the unique nature of hydroponic plant growth, researchers can unlock the full potential of this innovative technology and achieve accurate and reliable imaging for plant growth analysis. A practical example of the described procedure has been done in a study by M. S. Tasrif Anubhove et al. (2020), where they successfully use high resolution cameras to monitor the plant growth of tomato plants.

In order to facilitate a comprehensive understanding of the hydroponic environment and its dynamic changes, a summary table has been created that lists the various characteristics collected, their respective datatype, the required measurement resolution, the recording type, and the unit in which the characteristic is expressed (see **Table 1**). This table serves as a valuable reference for researchers and growers to effectively monitor and optimize the hydroponic system.

The collected features can be broadly categorized into two main types of data: numerical sensor data and metadata. Numerical sensor data represents the majority of the characteristics collected in the hydroponic system. This data is obtained from the camera and other sensors. To ensure accurate and meaningful data analysis, a recommended resolution has been established for each numerical sensor data type. Metadata, on the other hand, includes information about the physical characteristics and specifications of the hydroponic system. Unlike numerical sensor data, metadata is determined only once per dataset and remains constant throughout the duration of the experiment. It serves as a reference for researchers and growers to correlate specific environmental conditions with observed plant responses.

The data acquisition rate can be divided into two primary types: once per dataset or continuous recording. Certain features only need to be determined once and remain constant throughout the dataset. For continuous recording, it is important to select a measurement frequency that is appropriate for the experiment. In most cases, for a simple day/night cycle with one day and one night phase, it is recommended to take at least one image at the beginning and one at the end of the day phase. It is important that the different measurements are carried out at the same time and at the same intervals so that they can be bundled and correlated with each other. While this represents the minimum frequency for monitoring hydroponic systems, the amount of data produced might not be sufficient for certain data-driven optimizations. For example, a study in Korea acquired 364 images per day between 9:00 a.m. and 9:00 p.m. to train an algorithm for a plant phenotyping system with multiple machine-learning algorithms (Lee et al., 2018).

Table 1. Static parameters for data acquisition in hydroponic systems

Name	Datatype	Resolution	Recording type	Unit
		Light		
PPFD	numerical	±5	once per DS	µmol m^{-2} s^{-1}
DLI	numerical	±0,1	once per DS	mol m^{-2} d^{-1}
Day/Night composition	numerical	±1	once per DS	h
Day hours	numerical	-	once per DS	h
Night hours	numerical	-	once per DS	h
Light spectrum	metadata	-	-	-
Natural/Artificial light	metadata	-	-	-
Lamps used	metadata	-	-	-
		Nutrient solution		
EC-value	numerical	±0,1	continuous	mS cm^{-1}
pH-value	numerical	±0,1	continuous	-
Temperature of NS	numerical	±1	continuous	°C
Dissolved oxygen	numerical	±1	continuous	ppm
Replenished NS	numerical	-	-	l
Replenished water	numerical	-	-	l
Time of replenishment	numerical	-	-	-
Amount of NS	metadata	-	-	-
Mineral composition of used water	metadata	-	-	-
Nutrient composition of fertilizers	metadata	-	-	-
Means of oxygen enrichment	metadata	-	-	-

		Greenhouse environment		
Airflow	numerical	±0,1	continuous	$m\ s^{-1}$
CO2	numerical	±1	continuous	ppm
Relative humidity	numerical	±1%	continuous	-
Temperature	numerical	±1	continuous	°C
Aeration device used	metadata	-	-	-
Amount of airflow	metadata	-	-	l/h
Position of aeration	metadata	-	-	-
		Plant growth		
Weight	numerical	±1	continuous	g
Visual documentation	image	-	continuous	-

4. DISCUSSION

Among the identified 28 parameters, only 10 necessitate continuous recording throughout the growth cycle to obtain comprehensive data. Furthermore, essential parameters such as EC, pH, water and air temperature, humidity, and CO2 are typically already measured and controlled to maintain proper environmental conditions hydroponic system.

Recent research findings, particularly those regarding the impact of microbubbles or intermittent lighting on plant growth, remain relatively unexplored, and their necessity for data collection cannot be conclusively established at this stage. Nonetheless, in the pursuit of a comprehensive dataset, it is prudent to gather this data regardless. Failing to collect this data in the context of data-driven research renders the dataset incomplete, resulting in an inadequate depiction of the hydroponic system. By adopting a proactive approach to data collection, we ensure that potentially relevant factors, such as the influence of microbubbles or intermittent lighting, are thoroughly examined and considered. Though the immediate significance of certain data elements might not be apparent, their inclusion can enhance the dataset's value for subsequent analyses and research endeavors. As a result, our understanding of hydroponic practices is enriched, and any potential correlations or effects can be better identified and understood.

Measuring plant weight during the growth process is relevant when the weight serves as a crucial indicator of the success of the growth cycle. This is particularly applicable to plants like salad, basil, or others that are entirely harvested and sold. However, to derive accurate and reliable insights from such measurements, it is essential to effectively address external influences that may affect the measurement process. Mitigating external factors such as air flow against the leaves, variations in root buoyancy due to fluctuating water levels, or entangled plant leaves, is crucial to maintain measurement integrity.

Consequently, achieving complete datasets requires relatively minimal additional investment. The primary focus lies in implementing a robust data infrastructure to facilitate data collection and archival.

The advantages of gathering complete standardized datasets are manifold. Firstly, it allows the combination of datasets from different sources to construct large datasets for machine learning application. Secondly, researchers can revisit and reevaluate old data using novel research inquiries, providing fresh insights and perspectives. Additionally, complete datasets support crowd-sourcing efforts, fostering collaboration and enhancing the overall value of the collected information. Isolated datasets from individual growth cycles hold limited significance without the ability to be compared and analyzed in a broader context. Finally standardized data collection allows data reproducibility.

5. CONCLUSIONS

The process of establishing guidelines for hydroponic datasets for Machine Learning (ML) is inherently iterative. The guidelines need to be put into practice, and the outcomes should be subject to rigorous examination and critical analysis. By applying the guidelines in real-world scenarios and research, the effectiveness of the guidelines can be evaluated, and any potential shortcomings identified. Based on the literature findings and experimental results, three preliminary guidelines can be formulated:

Growth data should be consistently collected in a comprehensive manner, encompassing all 28 identified parameters. This inclusive approach ensures that a holistic view of the growth process is captured, leaving no critical aspects unaddressed.

Parameters should be assessed to determine their relevance and applicability in a specific use case. If necessary, additional data should be collected to complement the existing parameters and facilitate a more accurate analysis of the growth process.

The principle of "more data is better than less data" should be upheld, with the data collection frequency being tailored based on an evaluation of the expected dataset size and any potential data storage limitations.

The practical implementation of these guidelines in the hydroponics simulator seems relatively straightforward, with minimal challenges foreseen. Most of the relevant process parameters are readily accessible and merely require accurate data documentation. Future data collection in the hydroponic simulator will allow to verify the preliminary guidelines and further improve them. By continuously evaluating and updating the guidelines based on practical implementation and empirical evidence, they can provide a significant improvement in the availability of ML datasets for hydroponic applications. This iterative approach ensures that the guidelines remain relevant and effective.

REFERENCES

Abu-Shahba, M. S., Mansour, M. M., Mohamed, H. I. and Sofy, M. R. (2021) 'Comparative Cultivation and Biochemical Analysis of Iceberg Lettuce Grown in Sand Soil and Hydroponics With or Without Microbubbles and Macrobubbles', *Journal of Soil Science and Plant Nutrition*, vol. 21, no. 1, pp. 389–403.

Andriolo, J. L., Da Luz, G. L., Witter, M. H., Godoi, R. d. S., Barros, G. T. and Bortolotto, O. C. (2005) 'Growth and yield of lettuce plants under salinity', *Horticultura Brasileira*, vol. 23, no. 4, pp. 931–934.

Avgoustaki, D. D., Li, J. and Xydis, G. (2020) 'Basil plants grown under intermittent light stress in a small-scale indoor environment: Introducing energy demand reduction intelligent technologies', *Food Control*, vol. 118, p. 107389 [Online]. DOI: 10.1016/j.foodcont.2020.107389.

Brown, C. S., Schuerger, A. C. and Sager, J. C. (1995) 'Growth and Photomorphogenesis of Pepper Plants under Red Light-emitting Diodes with Supplemental Blue or Far-red Lighting', *Journal of the American Society for Horticultural Science jashs*, vol. 120, no. 5, pp. 808–813 [Online]. DOI: 10.21273/JASHS.120.5.808.

Chen, W.-T., Yeh, Y.-H. F., Liu, T.-Y. and Lin, T.-T. (2016) 'An Automated and Continuous Plant Weight Measurement System for Plant Factory', *Frontiers in plant science*, vol. 7, p. 392.

Chérif, M., Tirilly, Y. and Bélanger, R. R. (1997) 'Effect of oxygen concentration on plant growth, lipidperoxidation, and receptivity of tomato roots to Pythium F under hydroponic conditions', *European Journal of Plant Pathology*, vol. 103, no. 3, pp. 255–264.

Chia, S. Y. and Lim, M. W. (2022) 'A critical review on the influence of humidity for plant growth forecasting', *IOP Conference Series: Materials Science and Engineering*, vol. 1257, no. 1, p. 12001.

Corré, W. J. (1983) 'GROWTH AND MORPHOGENESIS OF SUN AND SHADE PLANTS I. THE INFLUENCE OF LIGHT INTENSITY', *Acta Botanica Neerlandica*, vol. 32, 1-2, pp. 49–62.

Fu, W., Li, P. and Wu, Y. (2012) 'Effects of different light intensities on chlorophyll fluorescence characteristics and yield in lettuce', *Scientia Horticulturae*, vol. 135, pp. 45–51 [Online]. DOI: 10.1016/j.scienta.2011.12.004.

Goins, G. D., Yorio, N. C., Sanwo, M. M. and Brown, C. S. (1997) 'Photomorphogenesis, photosynthesis, and seed yield of wheat plants grown under red light-emitting diodes (LEDs) with and without supplemental blue lighting', *Journal of Experimental Botany*, vol. 48, no. 7, pp. 1407–1413.

Hopkinson, S. and Harris, M. (2019) 'Effect of pH on Hydroponically Grown Bush Beans (phaseolus vulgaris)', *Int. J. Environ. Agric. Biotechnol.*, vol. 4, no. 1, pp. 142–145.

Hornberg, A. (ed) (2017) *Handbook of Machine and Computer Vision: The Guide for Developers and Users* [Online], 2nd edn, Weinheim, Wiley-VCH Verlag GmbH & Co. KGaA. Available at http://onlinelibrary.wiley.com/book/10.1002/9783527413409.

Ikeura, H., Takahashi, H., Kobayashi, F., Sato, M. and Tamaki, M. (2018) 'Effects of microbubble generation methods and dissolved oxygen concentrations on growth of Japanese mustard spinach in hydroponic culture', *The Journal of Horticultural Science and Biotechnology*, vol. 93, no. 5, pp. 483–490.

Kane, C. D., Jasoni, R. L., Peffley, E. P., Thompson, L. D., Green, C. J., Pare, P. and Tissue, D. (2006) 'Nutrient Solution and Solution pH Influences on Onion Growth and Mineral Content', *Journal of Plant Nutrition*, vol. 29, no. 2, pp. 375–390.

Kang, J., Sugumaran, K., Atulba, S. L., Jeong, B. R. and Hwang, S. (2013) 'Light intensity and photoperiod influence the growth and development of hydroponically grown leaf lettuce in a closed-type plant factory system', *Horticulture*, vol. 54.

Kappel, N., Boros, I. F., Ravelombola, F. S. and Sipos, L. (2021) 'EC Sensitivity of Hydroponically-Grown Lettuce (Lactuca sativa L.) Types in Terms of Nitrate Accumulation', *Agriculture*, vol. 11, no. 4, p. 315.

Lee, J. G., Choi, C. S., Jang, Y. A., Jang, S. W., Lee, S. G. and Um, Y. C. (2013) 'Effects of air temperature and air flow rate control on the tipburn occurrence of leaf lettuce in a closed-type plant factory system', *Horticulture, Environment, and Biotechnology*, vol. 54, no. 4, pp. 303–310.

Lee, U., Chang, S., Putra, G. A., Kim, H. and Kim, D. H. (2018) 'An automated, high-throughput plant phenotyping system using machine learning-based plant segmentation and image analysis', *PloS one*, vol. 13, no. 4, e0196615.

M. S. Tasrif Anubhove, N. Ashrafi, A. M. Saleque, M. Akter and S. U. Saif (2020) 'Machine Learning Algorithm based Disease Detection in Tomato with Automated Image Telemetry for Vertical Farming', *2020 International Conference on Computational Performance Evaluation (ComPE)*, pp. 250–254.

Makino, A. and Mae, T. (1999) 'Photosynthesis and Plant Growth at Elevated Levels of CO2', *Plant and Cell Physiology*, vol. 40, no. 10, pp. 999–1006.

Nguyen, T., Tran, T. and Nguyen, Q. T. (2019) 'Effects of light intensity on the growth, photosynthesis and leaf microstructure of hydroponic cultivated spinach (Spinacia oleracea L.) under a combination of red and blue LEDs in house', *International Journal of Agricultural Technology*, vol. 15, pp. 75–90.

Park, J.-S. and Kurata, K. (2009) 'Application of Microbubbles to Hydroponics Solution Promotes Lettuce Growth', *HortTechnology*, vol. 19, no. 1, pp. 212–215.

Peiro, E., Pannico, A., Colleoni, S. G., Bucchieri, L., Rouphael, Y., Pascale, S. de, Paradiso, R. and Gòdia, F. (2020) 'Air Distribution in a Fully-Closed Higher Plant Growth Chamber Impacts Crop Performance of Hydroponically-Grown Lettuce', *Frontiers in plant science*, vol. 11, p. 537.

Resh, H. M. (2022) *Hydroponic food production: A definitive guidebook for the advanced home gardener and the commercial hydroponic grower*, Boca Raton, CRC Press.

Savvas, D., Stamati, E., Tsirogiannis, I. L., Mantzos, N., Barouchas, P. E., Katsoulas, N. and Kittas, C. (2007) 'Interactions between salinity and irrigation frequency in greenhouse pepper grown in closed-cycle hydroponic systems', *Agricultural Water Management*, vol. 91, no. 1, pp. 102–111 [Online]. DOI: 10.1016/j.agwat.2007.05.001.

Shannon, C. E. (1948) 'A Mathematical Theory of Communication', *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423.

Shibata, T., Iwao, K. and Takano, T. (1995) 'EFFECT OF VERTICAL AIR FLOWING ON LETTUCE GROWING IN A PLANT FACTORY', *Acta Horticulturae*, no. 399, pp. 175–182.

Song, Y. H., Shim, J. S., Kinmonth-Schultz, H. A. and Imaizumi, T. (2015) 'Photoperiodic flowering: time measurement mechanisms in leaves', *Annual review of plant biology*, vol. 66, pp. 441–464.

Sublett, W., Barickman, T. and Sams, C. (2018) 'The Effect of Environment and Nutrients on Hydroponic Lettuce Yield, Quality, and Phytonutrients', *Horticulturae*, vol. 4, no. 4, p. 48.

Suyantohadi, A., Kyoren, T., Hariadi, M., Purnomo, M. H. and Morimoto, T. (2010) 'Effect of high consentrated dissolved oxygen on the plant growth in a deep hydroponic culture under a low temperature', *IFAC Proceedings Volumes*, vol. 43, no. 26, pp. 251–255 [Online]. DOI: 10.3182/20101206-3-JP-3009.00044.

Vanhassel, P., Bleyaert, P., van Lommel, J., Vandevelde, I., Crappé, S., van Hese, N., Hanssens, J., Steppe, K. and van Labeke, M.-C. (2015) 'Rise of nightly air humidity as a measure for tipburn prevention in hydroponic cultivation of butterhead lettuce', *Acta Horticulturae*, no. 1107, pp. 195–202.

Wortman, S. E. (2015) 'Crop physiological response to nutrient solution electrical conductivity and pH in an ebb-and-flow hydroponic system', *Scientia Horticulturae*, vol. 194, pp. 34–42 [Online]. DOI: 10.1016/j.scienta.2015.07.045.

Zheng, J., Ji, F., He, D. and Niu, G. (2019) 'Effect of Light Intensity on Rooting and Growth of Hydroponic Strawberry Runner Plants in a LED Plant Factory', *Agronomy*, vol. 9, no. 12.