POLITECNICO DI TORINO
Repository ISTITUZIONALE

# Hybrid machine learning and physics-based model for estimating lettuce (Lactuca sativa) growth and resource consumption in aeroponic systems

*Original*
Hybrid machine learning and physics-based model for estimating lettuce (Lactuca sativa) growth and resource consumption in aeroponic systems / Fasciolo, B., Grasso, N., Bruno, G., Chiabert, P.. - In: SCIENTIFIC REPORTS. - ISSN 2045-2322. - ELETTRONICO. - 15:1(2025). [10.1038/s41598-025-02763-9]

*Availability:*
This version is available at: 11583/3001523 since: 2025-07-03T13:33:41Z

*Publisher:*
springer nature

*Published*
DOI:10.1038/s41598-025-02763-9

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

19 November 2026

scientific reports

Check for updates

OPEN

# Hybrid machine learning and physics-based model for estimating lettuce (Lactuca sativa) growth and resource consumption in aeroponic systems

Benedetta Fasciolo✉, Nicolò Grasso, Giulia Bruno & Paolo Chiabert

**As the global population is expected to reach 10.3 billion by the mid-2080s, optimizing agricultural production and resource management is crucial. Climate change and environmental degradation further complicate these challenges, impacting crop productivity and food security. Traditional farming methods struggle with efficiently managing nutrients and water while ensuring high-quality products, leading to resource wastage and food safety concerns. This study aims to develop a hybrid model combining machine learning and physics-based techniques to predict fresh weight, leaf area, nitrate levels, and water consumption in lettuce grown in aeroponic systems, thereby enhancing resource management and product quality. We integrated a physics-based model with machine learning algorithms to create a dynamic hybrid framework. The model was validated with real-time data from aeroponic systems, showing good predictive performance, particularly for fresh weight and total leaf area. In contrast, predictions of nitrate content and water consumption were less accurate, due in part to smaller training datasets and limitations of the physics-based component under soilless conditions. Despite these challenges, the hybrid model offers a promising solution for optimizing controlled environment agriculture, addressing critical challenges in modern agriculture by improving efficiency and sustainability.**

**Keywords** Hybrid model, Machine learning, Physics-based model, Aeroponic, IoT, Predictive model

Optimizing resources and agricultural production is crucial to addressing the growing global population and urbanization challenges. The world population will grow over the next sixty years from 8.2 billion in 2024 to around 10.3 billion people in the mid-2080s[1]. A significant increase in agricultural production will be required in the coming decades to ensure food security. However, growing urbanization and the environmental crisis pose significant threats. Climate change strongly impacts plant abiotic stress, such as high temperatures[2], salinity[3], and heavy metals[4], compromising agricultural productivity. Moreover, the limited availability of essential resources like macro and micronutrients[5] and soil pH variations further exacerbate the situation[4]. On the other hand, agricultural activities like deforestation[6] and intensive use of fertilizers and pesticides[7,8] contribute to the environmental crisis by reducing the soil organic carbon pool and causing water and air pollution. Food waste is another critical issue with significant environmental and food security implications. Approximately 14% of all food produced is estimated to be lost between the post-harvest stage and just before the retail stage[9]. Developing more efficient agricultural practices and adopting innovative technologies to reduce losses during distribution can significantly decrease environmental impact and ensure greater food availability for the growing global population.

A potential method that can partially face these challenges is Controlled Environment Agriculture (CEA). It refers to the practice of producing crops by adjusting environmental conditions such as lighting, temperature, $CO_2$ levels, humidity, irrigation, fertigation, and other factors that affect a plant's physiological responses[10]. CEA systems involve diverse production methods, all characterized by their ability to regulate the growing environment. Crops can grow either in single layers exposed to natural sunlight or in vertically stacked layers under artificial lighting[11]. As a result, these systems are less impacted by external factors like changes in precipitation and temperature, increasing their resilience to environmental stresses. Integrating IoT technologies

Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy. ✉email: benedetta.fasciolo@polito.it

allows CEA systems to precisely control and improve water and nutrient delivery, as well as other environmental parameters such as light and temperature, ensuring optimal growing conditions[12]. The controlled environment also minimizes the spread of pests and diseases, significantly reducing the need for pesticides or chemicals[13]. Finally, CEA systems provide improved nutrient management and recirculation, especially in hydroponic and aeroponic configurations[14]. CEA uses mainly hydroponics, aquaponics, aeroponics and bioponics as irrigation techniques. Among these techniques, aeroponics stands out for its ability to improve nutrient uptake and oxygenation. Plant roots are suspended in the air and periodically misted with a fine spray of nutrient solution, ensuring precise control over moisture levels and reducing water consumption[13].

Predictive models can be integrated into CEA, which optimizes inputs (environmental resources) to improve outputs (crop growth quality and resource consumption). According to[15,16], two types of predictive models can be used. Physics-based models are computational or mathematical frameworks that use the fundamental principles of physics to simulate and predict the behavior of complex systems. In the case of agricultural modeling, a physics-based model can simulate physiological processes such as light interception, photosynthesis, and respiration. However, these models are inherently constrained by the validity of their underlying assumptions and parameter ranges, meaning they can only provide accurate predictions within specific environmental conditions. When applied beyond these calibrated ranges, their reliability decreases, leading to potential inaccuracies. On the other hand, data-driven models, particularly those based on machine learning, offer greater flexibility as they can be trained on diverse datasets encompassing a wide range of growing conditions. This adaptability allows for more robust predictions across different environments. However, machine learning approaches have their own limitations: they require large amounts of high-quality data for effective training, and the resulting models often function as "black boxes," making it difficult to interpret the underlying biological mechanisms driving the predictions[17]. To address these challenges, a hybrid approach that integrates physics-based and machine-learning models can be a promising solution. By combining the interpretability and domain knowledge embedded in physics-based models with the adaptability and predictive power of machine learning, this synergy can help overcome limitations encountered in experimental settings.

In this study, we developed a hybrid model in which machine learning algorithms estimate several crop growth parameters. One of those serves as input to the physics-based model, which then predicts another key agronomic factor related to resource consumption. Specifically, our model provides dynamic estimations of fresh weight, total leaf area, nitrate content of lettuce, and total water consumption within the cultivation system. Compared to traditional modeling approaches, this hybrid method offers a flexible and scalable solution for Controlled Environment Agriculture (CEA). While physics-based models may struggle with incomplete data or parameter calibration, our approach enables a more adaptive framework that integrates data-driven insights where sufficient information is available, while relying on established physical principles where data is scarce. Previous studies have typically relied on either physics-based or machine learning models in isolation, whereas our method demonstrates how a strategic combination of both can improve predictive accuracy and decision-making in resource-limited scenarios.

The paper is organized as follows. Section "Related work" provides a detailed description of the Related work and the novelty of the presented model. Section "Material and methods" provides a detailed description of the aeroponic chamber setup, the experimental activity carried out, and the structure of the hybrid model employed in the study. Section "Result" focuses on the results and performance of the model, with a detailed examination of each sub-model. Section "Discussion" offers a discussion on the advantages and limitations of the hybrid model. Section "Conclusion" concludes the paper.

## Related work

Recent advances in crop modeling within CEA have demonstrated the growing relevance of machine learning and deep learning techniques. A variety of models have been proposed, differing in their methodologies (e.g., regression, neural networks, time-series forecasting), input types (e.g., environmental data, spectral data, historical crop data), and outputs (e.g., yield, physiological traits, resource usage).

Regression-based models have primarily focused on yield prediction and basic phenotypic traits. For instance[18], used multiple linear regression to estimate fresh weight, nitrate content, and leaf area. Similarly[19], explored yield prediction across different hydroponic setups using classic Machien Learning models, relying mainly on environmental inputs.

Sensor-integrated deep learning models push further into non-destructive phenotyping[20]. used RGB-D cameras with deep regression networks to estimate lettuce morphological features[21]. blended neural networks with fuzzy logic to enhance yield prediction in IoT-enabled hydroponic systems.

Time-series approaches have added temporal depth to prediction[22]. used historical data and statistical models to forecast plant size. A more advanced application is seen in[23], which applied attention-based deep learning for multivariate, weekly forecasting using environmental time-series inputs.

Multimodal and meta-learning methods have been introduced to handle complex, heterogeneous data. In[24], meta-learning was used to integrate spectral, thermal, and environmental data for real-time physiological state recognition.

Deep learning for resource monitoring is also emerging. The use of LSTM networks and autoencoders in[25] allowed modeling of nutrient availability in aquaponic systems based on latent time-dependent patterns.

Environmental-feature-driven models, such as in[26] highlight the strong predictive power of models that use light intensity, EC, and temperature as primary inputs, particularly in yield estimation tasks.

Most models remain either strictly data-driven or solely physics-based. There is limited exploration of hybrid models that systematically combine the interpretability of physics with the adaptiveness of machine learning. Given the limitations inherent in both physics-based and machine learning approaches, hybrid modeling offers a promising pathway by integrating their respective strengths. Physics-based models, while interpretable and

grounded in biological principles, often struggle with parameter calibration and environmental variability. Conversely, data-driven models excel at capturing complex, non-linear relationships but are highly data-dependent and lack transparency. The novelty of this work lies in its integrated hybrid architecture, where machine learning is not merely used for direct output prediction but strategically supports and augments a physics-based model. Specifically, ML algorithms are trained to estimate intermediate crop growth parameters such as fresh weight, total leaf area, and nitrate content, which are then used as inputs to a physics-based module that simulates total water consumption. By embedding Machine Learning derived growth estimates into a mechanistic water model, we retain the interpretability and domain knowledge of physical modeling while dynamically adapting to data availability. Moreover, while previous studies focused predominantly on yield prediction or single-parameter forecasting, our model provides a multi-output dynamic framework that integrates growth and resource-related predictions.

Building on this modeling approach, the model is designed to accept environmental parameters as inputs over the course of the crop cycle. Based on these, it estimates key agronomic outputs such as fresh weight, leaf area, nitrate content, and water consumption. These estimates can support growers in predicting crop performance under given environmental scenarios, enabling them to make informed decisions. For example, if the environmental inputs are fixed and not controllable, the model can help anticipate outcomes. Alternatively, if environmental conditions can be adjusted, the model can guide growers in optimizing these inputs to achieve desired outcomes.

## Materials and methods

The first part of the Materials and methods section describes the structure of the controlled environment growth chamber (Section "System construction and cyber-physical system development"), the experimental activities, and the data collected (Section "Experimental activity") used to train the model. Lastly, there is a section detailing the hybrid model, its characteristics, and its components (Section "Modelling tools").

### System construction and cyber-physical system development

The cultivation systems for the experimental setup consisted of three fully automated and sensorized aeroponic modules, each designed to monitor and regulate key environmental parameters. Every module was housed within a grow tent measuring $150 \times 150 \times 200$ cm, which allowed for a semi-open environment, with the interior lining featuring a reflective mylar to enhance light efficiency. Inside each tent, plants were cultivated on a plastic tray ($60 \times 40$ cm) with a structured bed made of rock wool dowels, allowing for the simultaneous growth of up to 228 plants. LED support system enabled precise adjustment of the lighting distance to optimize plant exposure. The illumination setup comprised three LED lamps, each 70 cm long and powered by a 24 V supply, ensuring uniform light distribution across the cultivation area.

Each module was equipped with a set of sensors and actuators to monitor and regulate the growing environment. Sensors were placed in two areas: the nutrient solution reservoir and the growth chamber. In the reservoir, pH and electrical conductivity (EC) sensors assessed nutrient concentration, while environmental sensors measured temperature, humidity, and $CO_2$ levels within the chamber. Communication between sensors and the control unit was managed via the $I^2C$ protocol, a widely adopted serial communication standard[27].

The actuators were divided into environmental and climate control and irrigation. Regarding irrigation, the nutrient delivery system utilized a primary pump to spray the solution through atomization nozzles. Additional peristaltic pumps controlled pH regulation by dosing acids or bases as needed and managed the distribution of macro and micronutrients. Regarding environmental and climate control actuators, the lighting system, consisting of three LED lamps, provided a light spectrum with 17% blue light (400–500 nm), 45% green light (500–600 nm), and 38% red light (600–800 nm), supporting optimal plant growth. Internal climate conditions were regulated by a heating panel and a cooler to adjust the temperature. Finally, an exhaust fan extracted air when the humidity was too high, and a ventilation fan circulated the air.

For system control and data acquisition, each module incorporated a Raspberry Pi single-board computer (SBC). The Raspberry Pi managed actuators via General-Purpose Input/Output (GPIO) pins for on/off operations, while the four dosing pumps, two dedicated to pH adjustments and two for EC regulation, communicated using the $I^2C$ protocol. The open-source software MyCodo was installed on the Raspberry Pi to facilitate system automation, utilizing both time-based and event-driven commands. All collected data was stored locally for further analysis. Figure 1 illustrates one of the modules used in the experimental activity and the aeroponic cyber-physical system.

### Experimental activity

Seeds of butterhead lettuce (A3036, Gautier semences, FR) (Lactuca sativa var. capitata) were planted in 1 $cm^2$ rock wool cubes. Each cube, that host one seed, was placed in a tray with 228 slots and watered with tap water, then kept in the dark at 22 °C ± 2°C for 72 h to germinate. After three days, once the seeds were germinated (two cotyledons per sprout), the entire tray with the young sprouts was moved to the aeroponic system described in Section "System construction and cyber-physical system development". Each growth cycle lasted three weeks (3 days of germination plus 21 days of growing cycle). During the first week, around 228 plants were cultivated, each with approximately 10.5 $cm^2$ of growth space. Then, due to disruptive analysis, in the second week, the 114 plants remaining had 21 $cm^2$ of growth space, and in the third week, the 25 plants remaining had 96 $cm^2$ of growth space.

Throughout the cultivation period, the photoperiod was maintained at 16 h of light and 8 h of darkness[28]. The relative humidity was kept at 68% ± 12%[29], and the $CO_2$ level was not controlled and thus kept at atmospheric levels (420–450 ppm)[30]. The plant roots were sprayed with water and a nutrient solution for 2 min every 15 min[31,32]. The pump sprayed 6 L per minute through a total of 7 nozzles using a commercially available Mills

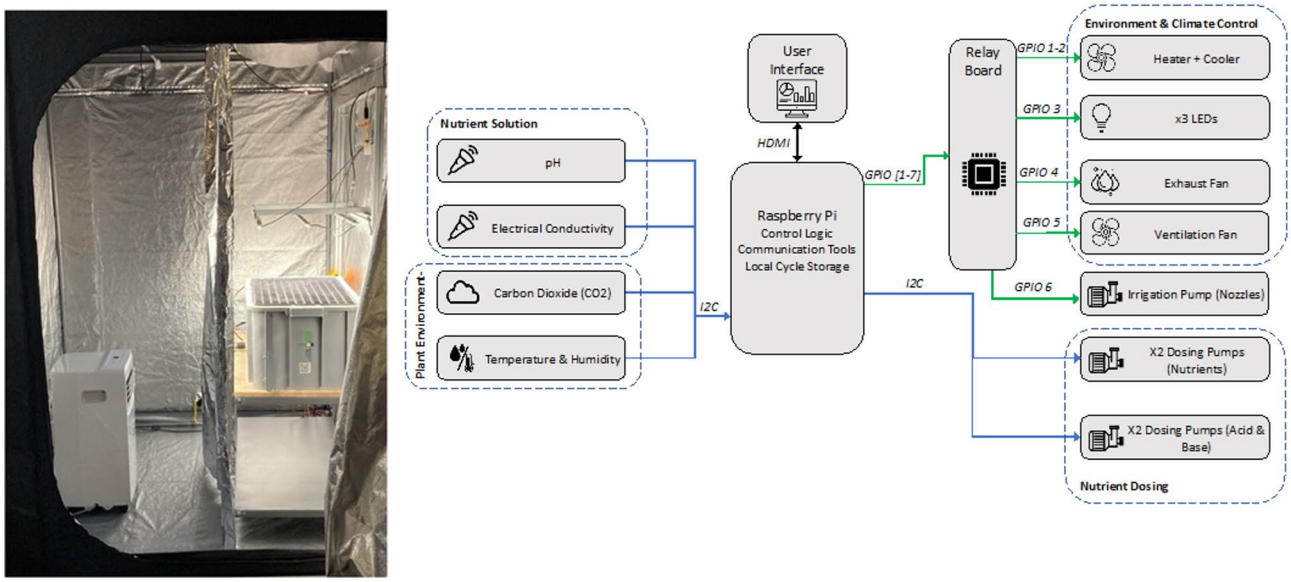



**Fig. 1**. Aeroponic Module (left) with cyber- physical system (right).

| Levels | Temperature | Light | Electrical Conductivity |
|---|---|---|---|
| High | 24.7 °C ± 1.2 °C | 253 µmol/m$^2$s ± 19 µmol/m2s | 1600 µS/cm ± 100 µS/cm |
| Low | 20.3 °C ± 1 °C | 143 µmol/m$^2$s ± 14 µmol/m$^2$s | 1000 µS/cm ± 100 µS/cm |

**Table 1**. Levels of the parameters tested during the experimental activity.

nutrient solution containing NH4 0.3%, NO3 3.2%, K2O 4.3%, CaO 5.1%, Fe 0.042%, P2O5 4.3%, MgO 1.1%, SO3 2.2%, B 0.0045%, Cu 0.0025%, Mn 0.02%, Mo 0.001%, and Zn 0.01%. The pH of the nutrient solution was adjusted with acetic acid and kept between 5.5 and 6.5[33].

During the experimental activity, we followed the Design of Experiment (DoE) approach investigating three parameters across two levels each: EC of the nutrient solution[34,35], light intensity[36,37], and temperature of the cultivation chamber[29,38]. This involved conducting a total of 8 growth cycles with one replica for a total of 16 growing cycles to test all possible combinations of these parameters. Each growing cycle lasted 21 days after the 3 days of germination. Table 1 summarizes the values of the tested parameters for each level.

The LEDs of the cultivation modules were adjusted to two different heights to provide the light intensity shown in Table 1. Respectively, at 42 cm and 14 cm from the tray, the light intensities of 253 µmol/m$^2$s and 143 µmol/m$^2$s were measured by a full-spectrum quantum sensor.

During each growing cycle, we conducted weekly measurements of fresh weight (FW), total leaf area (T-LA), nitrate content (N), and water consumption ($H_2O$). Regarding the FW and T-LA measurements, a total of 6 plants were collected and analyzed weekly. Thus, the tray was divided into three sectors, and for each sector, two plants were randomly selected for a total of six plants. First, each plant was cut at the root-crown level for immediate weighing to collect fresh weight data. Subsequently, every leaf of each plant was scanned and measured using the Petiole Pro APP. Then, the sum of all the plant leaves was done to acquire the total leaf area measurement of each crop. Figure 2 shows some samples collected after 7, 14, and 21 days in the growing chamber.

Regarding the N measurements, for each harvest week, a minimum of 3 g of lettuce per sector was collected and promptly frozen for 2–3 weeks for future nitrate analysis, conducted in accordance with BS EN 12014-2:2017[39]. The collected samples were homogenized using a mixer, and approximately 1 gram was analyzed. This sample was transferred to a 100 ml beaker containing 40 ml of ultrapure (MilliQ) water at 80 °C, brought to boiling point, and boiled for 15 min. After cooling to room temperature, the solution was filtered through a 25 µm filter (Whatman filter grade 4) and collected in a 50 ml volumetric flask (class A, Vitlab GmbH, DE). To purify the solution, 200 µl of the filtered solution was processed using a nitrate kit (LCK340, Hach, DE), then analyzed at 220 nm using a spectrophotometer (Hach, DE).

Table 2 details the structure of the dataset used to train the hybrid model.

In the dataset, all outputs (FW, T-LA, N, $H_2O$) have a mean sector value that changes weekly, except for water consumption, which has a mean weekly value. The temperature and humidity input values are reported as mean weekly values, while pH and EC are reported as a single mean value for the entire growth cycle. Light intensity has a mean sector value that remains constant throughout the cycle. The mean of the three sector values represents the high or low level of the light intensity values tested in the experimental activity reported in Table 1.

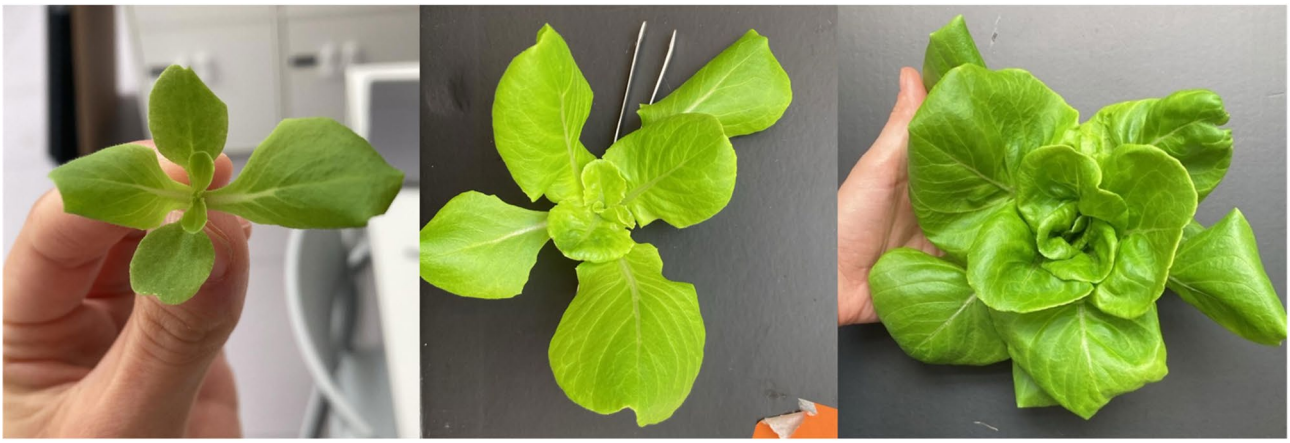


**Fig. 2**. Example of samples collected respectively at the first, second and third week of growth for the FW and T-LA analysis.

| Temp [°C] | Hum [%] | EC [µS/cm] | pH | L. int. [µmol/m2s] | Week | Sector | FW [mg] | T-LA [cm2] | *N* [mg/grFW] | $H_2O$ [L] |
|---|---|---|---|---|---|---|---|---|---|---|
| 25.03 | 75.7 | 1100 | 6 | 137 | 1 | A | 123 | 7.9 | 0.57 | 0.02 |
| 25.03 | 75.7 | 1100 | 6 | 156 | 1 | B | 132 | 8.7 | 0.72 | 0.02 |
| 25.03 | 75.7 | 1100 | 6 | 135 | 1 | C | 134.5 | 7.6 | 0.76 | 0.02 |
| 25.20 | 76.3 | 1100 | 6 | 137 | 2 | A | 1893.5 | 107.95 | 0.70 | 0.03 |
| 25.20 | 76.3 | 1100 | 6 | 156 | 2 | B | 2004.5 | 107.4 | 0.90 | 0.03 |
| 25.20 | 76.3 | 1100 | 6 | 135 | 2 | C | 1729.5 | 102.6 | 0.42 | 0.03 |
| 25.30 | 73.1 | 1100 | 6 | 137 | 3 | A | 9965 | 451.9 | 0.61 | 0.09 |
| 25.30 | 73.1 | 1100 | 6 | 156 | 3 | B | 11,886 | 508.4 | 0.64 | 0.09 |
| 25.30 | 73.1 | 1100 | 6 | 135 | 3 | C | 8886.5 | 405.3 | 0.745 | 0.09 |

**Table 2**. Dataset structure that reports the data of the first growing cycle.

## Modelling tools

The hybrid model developed in this study evaluates four output parameters: fresh weight (FW), total leaf area (T-LA), nitrate content (N), and water consumption ($H_2O$). These parameters are estimated on a weekly basis. They were selected because they represent the three fundamental dimensions for assessing the effectiveness of a cultivation system: productivity (FW and T-LA), quality (N), and sustainability ($H_2O$)[40]. Fresh weight is an indicator of the plant's biomass[41], while total leaf area reflects the crop's photosynthetic capacity. Additionally, nitrate levels that exceed certain thresholds can be harmful to human health[42], and water consumption is one of the primary resources used within controlled environment systems. A representation of the complete hybrid model with the four elements is illustrated in Fig. 3

Each sub-model receives specific inputs, derived both from data measured in the cultivation chamber and from the outputs of the other sub-models. This approach aligns with what is commonly referred as "chained regression" in the literature[43].

The outputs of the hybrid model have a weekly cadence as do the inputs of all sub-models. Nitrate content, fresh weight, and total leaf area are estimated on a weekly basis using three distinct sub-models that employ machine learning algorithms. In contrast, water consumption is estimated weekly using a physics-based model. As illustrated in Fig. 3, the model's structure varies between the first week and the subsequent weeks. During the first week, the sub-models for fresh weight and total leaf area do not incorporate inputs from the previous week, as the growing cycle has just begun. Similarly, for the water consumption sub-model, the total leaf area input is omitted during the first week because, as discussed in Section "Physics-based sub-model", the leaf area is smaller than the available growing space. However, in the following weeks, the sub-models for fresh weight and total leaf area include the previous week's inputs, $FWw_{i-1}$, and $T\text{-}Law_{i-1}$, respectively, while the water consumption sub-model incorporates the $T\text{-}LAw_i$ input.

The model begins with the fresh weight sub-model, which predicts the FW of the lettuce on a weekly basis. The inputs used for this sub-model are the environmental parameters measured during the experimental activity and the fresh weight from the previous week ($FWw_{i-1}$). The FW sub-model was tested in two versions. The first version uses the predicted FW of the previous week from the FW sub-model as input $FWw_{i-1}$. The second version uses the actual FW value measured during the experimental activity as input $FWw_{i-1}$. Since the performance of the two versions were similar, it was decided to use the predicted FW for training the model. Further details on the performance of the predicted and actual FW are reported in Section "Fresh weight sub-model performances".

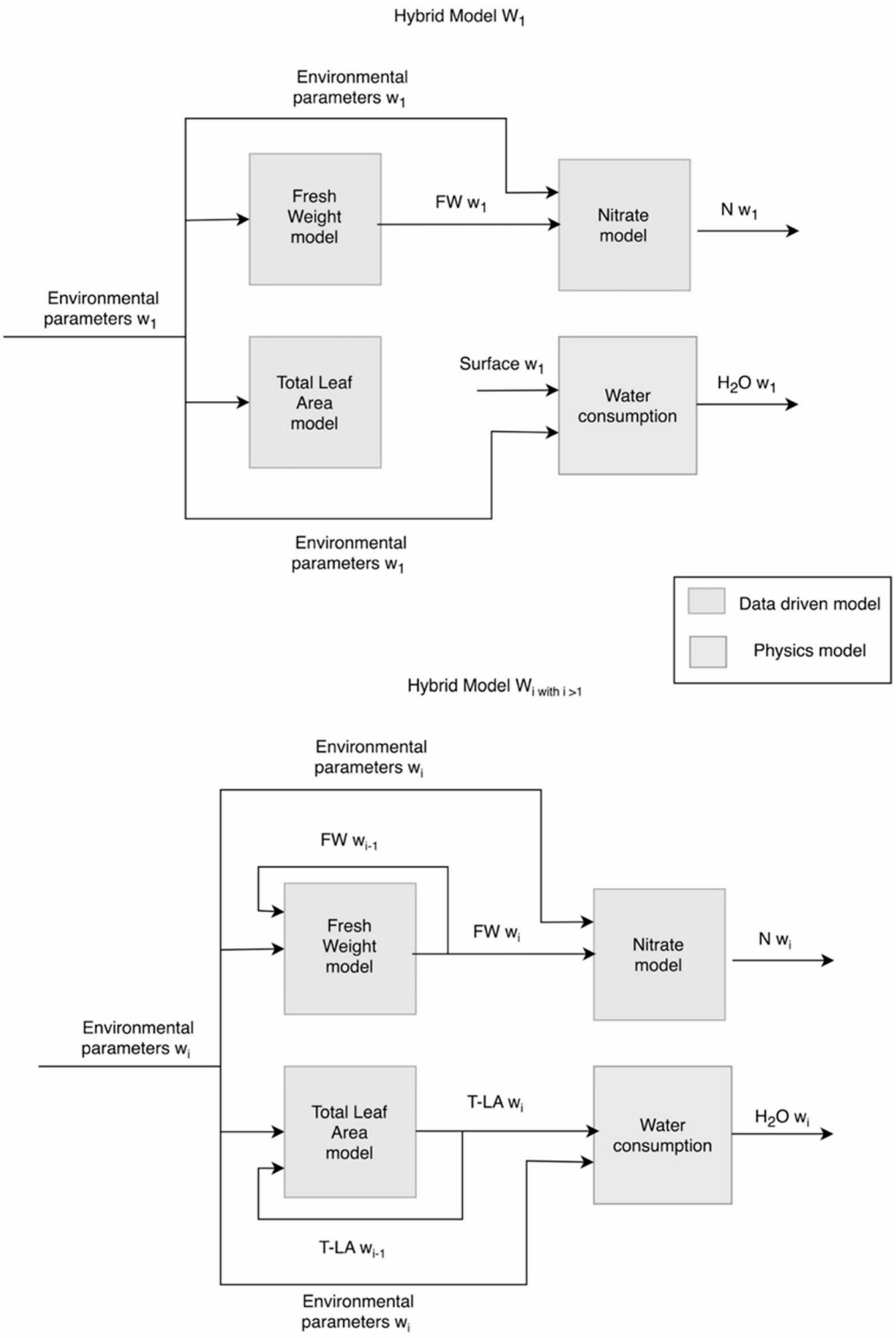



**Fig. 3**. Hybrid model.

The output of the FW sub-model, the predicted FW, is then used as input for the N sub-model along with the environmental parameters measured during the experimental activity. Finally, the total leaf area of the lettuce is also estimated weekly through the T-LA sub-model. To estimate the T-LA, the environmental variables and the leaf area from the previous week (T-LA $w_{i-1}$) are used as inputs. Similar to the FW model, the performance of the T-LA model was tested and compared using both the predicted T-LA $w_{i-1}$ and the actual T-LA $w_{i-1}$ derived from the experimental activity. The output of this sub-model was then used as input for the physics-based model to estimate the area occupied by the lettuce.

The choice of inputs for the machine learning models was made taking into consideration biological knowledge. For the FW and T-LA sub-models, environmental variables were utilized as inputs because the growth of crops in terms of weight and size is influenced by the environmental conditions to which they are subjected[44]. Moreover, fresh weight and leaf area from the previous week were included in the FW sub-model and the T-LA sub-model, respectively, as they represent the starting point for the weight and leaf area from which the plant will develop in the following week. In the N sub-model the environmental conditions were used as an input. The accumulation of nitrates in crops is closely related to the physiological state, which in turn is influenced by conditions in which it grows[44]. Additionally, the predicted weekly fresh weight was used as an input for the N sub-model. Fresh weight reflects the amount of biomass produced by the plant and can be correlated with nitrate accumulation, as a rapidly growing plant may have a higher nitrate demand[45]. The nitrate content from the previous week was not used as an input in the nitrate sub-model. Nitrate content can vary significantly from week to week due to changes in environmental conditions. The nitrate concentration in lettuce is much more dependent on environmental conditions at the current time than on the nitrate concentration in the previous week.

*Machine learning sub-models*

In this study, the prediction of fresh weight, nitrate content, and total leaf area parameters was performed using machine learning algorithms. The following algorithms were tested: Random Forest (RF), Deep Neural Network (DNN), Support Vector Machine (SVM), k-Nearest Neighbor (k-NN), Generalized Linear Model (GLM), and Gradient Boosted Trees (GBT). The performance of each algorithm was compared, and the best-performing algorithm was selected for each sub-model.

For the implementation of the algorithms, the dataset was divided into test and training sets using the leave-one-out cross-validation technique, being a robust evaluation strategy, particularly effective for small datasets, as it maximizes the use of available data[46]. The performance indicators considered to evaluate the accuracy of each model were the relative error (MAPE)(1) and squared correlation (R2)(2).

$$MAPE = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} \frac{\left| y_{obs,i} - y_{pred,i} \right|}{\left| y_{obs,i} \right|} \tag{1}$$

Where $y_{obs,\,i}$ is the observed value and $y_{pred,\,i}$ is the predicted value for the i-th observation. This indicator measures the average relative difference between the observed and predicted values, expressing the error as a fraction of the observed value.

$$R^2 = 1 - \frac{\sum \left( y_{obs} - y_{pred} \right)^2}{\sum \left( y_{obs} - \bar{y}_{obs} \right)^2} \tag{2}$$

Where $y_{obs}$ is the mean of the observed values. This indicator measures the proportion of the total variance in the data that is explained by the model. An $R^2$ value close to 1 indicates a high ability of the model to explain the variability of the observed data.

*Physics-based sub-model*

To estimate the weekly water consumption per plant, the Penman-Monteith evapotranspiration model was utilized[47]. This model is based on the crop evapotranspiration index ($ET_c$), which represents the total amount of water lost due to evapotranspiration from a crop in a specific area [mm/day]:

$$ET_c = ET_0 \times K_c \tag{3}$$

As seen from the formula (3), $ET_c$ is determined by the multiplication of $ET_0$, which represents the reference evapotranspiration of a well-irrigated grass surface, and $K_c$ which is the crop coefficient that primarily depends on the type of crop and its growth stage. Thus, the effect of both crop transpiration and soil evaporation are integrated into a single crop coefficient.

$ET_c$ is a unit-based index. To find the evapotranspiration per plant, $ET_c$ can be multiplied by the area of interest (the canopy cover estimated by the T-LA sub-model):

$$ET_{cp} = ET_c \times CC \tag{4}$$

where $ET_{cp.}$ is the evapotranspiration per plant and CC is the canopy cover. Canopy cover refers to the proportion of the ground area that is shaded by the foliage of crops when viewed from above[47].

$\underline{Et_0 \text{ calculation}}$ The calculation of $ET_0$ was performed using the Penman-Monteith equation:

$$ET_0 = \frac{0.408\Delta \ (R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta \ + \gamma \ (1 + 0.34 u_2)} \tag{5}$$

where:

- $ET_0$ is the reference evapotranspiration [mm $day^{-1}$],

- $R_n$ is the net radiation at the crop surface [MJ $m^{-2}$ $day^{-1}$] calculated through the Eq. (6),
- G is the soil heat flux density [MJ $m^{-2}$ $day^{-1}$] (assumed to be 0),
- T is the mean daily air temperature at 2 m height [°C] obtained by the temperature sensor readings,
- $U_2$ is the wind speed at 2 m height [m $s^{-1}$] obtained by the datasheet of the working ventilation fan,
- $e_s$ is the saturation vapor pressure [kPa] calculated through the specific formula,
- $e_a$ is the actual vapor pressure [kPa] calculated through the specific formula,
- $e_s - e_a$ is the saturation vapor pressure deficit [kPa],
- $\Delta$ is the slope vapor pressure curve [kPa $°C^{-1}$] calculated through the specific formula,
- $\gamma$ is the psychrometric constant [kPa $°C^{-1}$] calculated through the specific formula.

All parameters of the Eq. (5) were calculated as indicated by the physics-based model[47] based on the environmental input data measured through sensors during the experimental activity. Light radiation was converted to solar radiation following the Eq. (6).

$$R_n = \sum mol * w_i * 6{,}02214076x10^{23} * E_i * h * A \tag{6}$$

where:

- i represent a specific wavelength,
- w is the percentage of the specific wavelength.
- mol is the total number of moles of photons,
- E is the energy of a single photon at a specific wavelength,
- h is the photoperiod,
- A is the cultivation area.

$K_c$ calculation To calculate the crop coefficient ($K_c$), it is essential to first identify the growth stages of the crop. The crop is typically divided into four growth stages: initial (the crop covers 10% of the ground), development (it ends at approximately 70–80% ground cover), mid-season (from full cover to the start of maturity), and late season[47]. Given the same type of crop, the development rate of vegetation cover and the time required to achieve effective full cover are both influenced by weather conditions. Generally, once the effective full cover for a plant canopy has been reached, the rate of further phenological development (flowering, seed development, ripening, and senescence) depends more on plant genotype and less on weather conditions.

From our experimental activity, we hypothesized that for cycles where the temperature and light intensity had high levels, the initial stage ended around day 7, while the development stage ended around day 20. In contrast, when the temperature and light intensity had low levels, the initial stage ended on day 7, and the development stage ended around day 30.

Once the growth stages are identified for the first, second, and third weeks, the selected $K_c$ coefficients need to be adjusted for the frequency of wetting or climatic conditions during each growth stage following the procedure detailed by (Allen & Food and Agriculture Organization of the United Nations., 1998). Subsequently, a crop coefficient curve is constructed, allowing for the determination of $K_c$ values for any period during the growing period.

Water consumption calculation Once $ET_c$ is calculated using the formula (3), the water consumption per plant is estimated by multiplying $ET_c$ by the canopy cover. During the experimental activity, CC was estimated to be 42% in the second week of growth and 32% in the third week relative to the total leaf area. For the first week, since the cultivation grid area dedicated to each lettuce plant is larger than the area covered by the lettuce, $ET_c$ is multiplied by the total area allocated to each lettuce plant.

To evaluate the model's accuracy, observed values from the experimental activity were compared with model estimates using the MAPE and $R^2$ indices, described by their respective formulas (1) and (2).

## Result

Various machine learning algorithms were compared for each sub-model, selecting the one with the best performance. The RapidMiner tool[48] was exploited to train and test the models on our dataset. Table 3 reports the machine learning algorithms and the specific hyperparameters used in the whole hybrid model. To select the best combination of hyperparameters for each model, the grid search tuning strategy has been used. This procedure systematically tested multiple combinations of hyperparameters for each algorithm, such as the number of trees and maximum depth for Random Forest, epochs and learning rate for DNN, and gamma, C, and kernel type for SVM. The process allowed us to identify the best-performing configurations, which led to improvements in the whole model's performance.

Two scenarios were developed and compared for both the fresh weight sub-model and the total leaf area sub-model. The first scenario, for the FW and T-LA models, respectively, uses predicted $FWw_{i-1}$ and $T\text{-}LAw_{i-1}$ as input variables. The second scenario uses $FWw_{i-1}$ and $T\text{-}LAw_{i-1}$ measured during the real experimentally collected data as input variables.

In sections "Fresh weight sub-model performances", "Total leaf area sub-model performances", and "Nitrates sub-model performances", the performances of the fresh weight, leaf area, and nitrate sub-models are presented. For the total leaf area and fresh weight sub-models, a comparison between predicted FW and T-LA and measured FW and T-LA is also reported. Section "Water consumption sub-model performances" presents the performance of the physics-based model on water consumption per plant.

| | |
|---|---|
| **FW sub-model $w_1$: Random Forest** | |
| Number of trees | 21 |
| Criterion | MSE |
| Maximal depth | 70 |
| **FW sub-model $w_2$: Random Forest** | |
| Number of trees | 11 |
| Criterion | MSE |
| Maximal depth | 90 |
| **FW sub-model $w_3$: Random Forest** | |
| Number of trees | 41 |
| Criterion | MSE |
| Maximal depth | 29 |
| **T-LA sub-model $w_1$: Random Forest** | |
| number of trees | 90 |
| Criterion | MSE |
| Maximal depth | 60 |
| **T-LA sub-model $w_2$: Random Forest** | |
| number of trees | 60 |
| Criterion | MSE |
| Maximal depth | 21 |
| **T-LA sub-model $w_3$: Random Forest** | |
| number of trees | 70 |
| Criterion | MSE |
| Maximal depth | 70 |
| **N sub-model $w_1$: Deep Neural Network** | |
| Activation function | ReLU |
| Number of layers | 3 |
| Number of neurons | 50 |
| Learning rate | 0.4 |
| Loss function | MSE |
| Optimizer | SGD |
| Epochs | 100 |
| **N sub-model $w_2$: Random Forest** | |
| Number of trees | 21 |
| Criterion | MSE |
| Maximal depth | 29 |
| **N sub-model $w_3$: SVM** | |
| Kernel function | ANOVA |
| C | -1 |
| Gamma | 700 |

**Table 3**. Machine learning algorithms and parameters.

## Fresh weight sub-model performances

For the FW sub-model, the random forest algorithm was used to predict the fresh weight of the lettuce weekly. Table 4 presents the weekly performance of the model using predicted FWwi-1 and measured FWwi-1 as the input variable. The measured FWwi-1 is the fresh weight measured during the experimental activity described in section "Experimental activity".

The comparison of the performance shows that the differences are minimal, thus using the sub-model with predicted $FWw_{i-1}$ is preferable because it doesn't require measuring the actual fresh weight at the end of the first and second week. This approach ensures that if the hybrid model needs to be trained for a new crop or a new range of environmental conditions, it will only be necessary to collect environmental inputs during the experimental activity resulting in faster data collection.

| Week | $MAPE_{pred}$ | $R^2_{pred}$ | $MAPE_{meas}$ | $R^2_{meas}$ |
|---|---|---|---|---|
| Week 1 | 10.97% | 0.93 | 10.97% | 0.93 |
| Week 2 | 12.05% | 0.94 | 11.34% | 0.94 |
| Week 3 | 10.60% | 0.93 | 10.16% | 0.95 |

**Table 4**. Weekly performance of FW sub-model using predicted $FWw_{i-1}$ as input (left) and using measured $FWw_{i-1}$ as input (right).

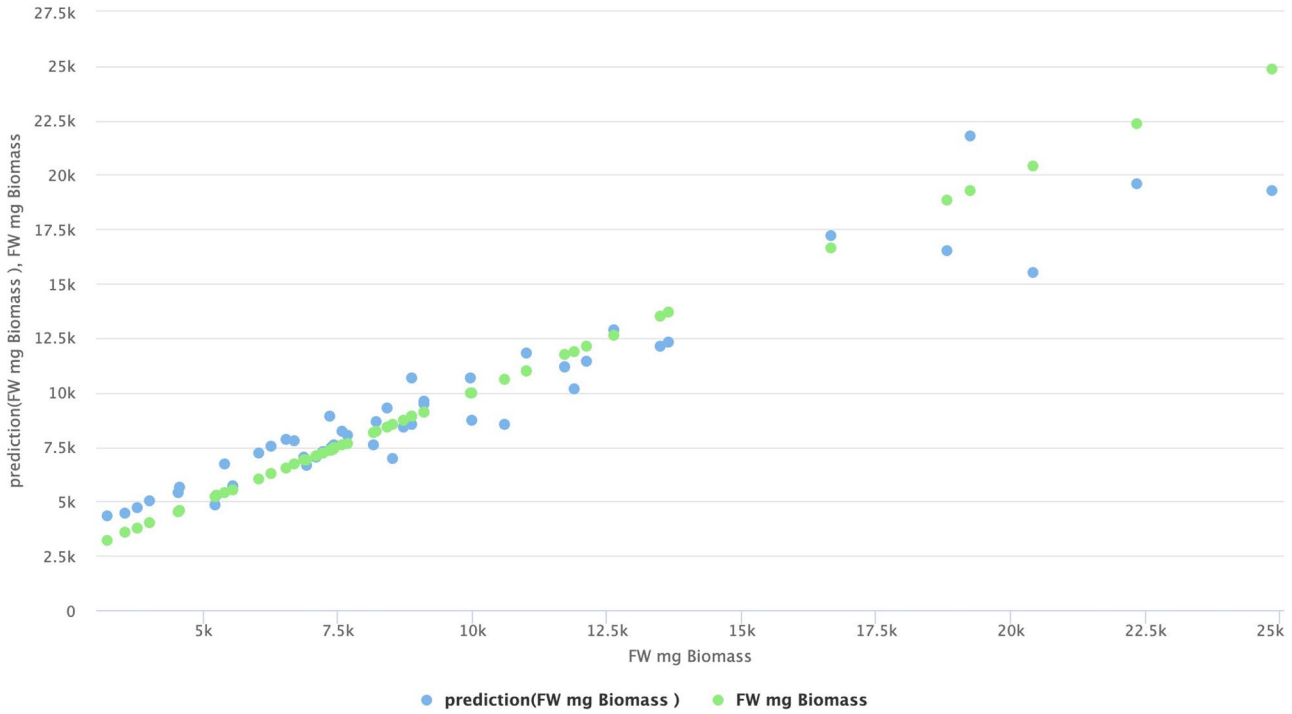



**Fig. 4**. FW predicted vs. FW measured at the third week.

| Week | $MAPE_{pred}$ | $R^2_{pred}$ | $MAPE_{meas}$ | $R^2_{meas}$ |
|---|---|---|---|---|
| Week 1 | 13.80% | 0.81 | 13.80% | 0.81 |
| Week 2 | 10.06% | 0.92 | 11.32% | 0.91 |
| Week 3 | 9.17% | 0.93 | 8.68% | 0.92 |

**Table 5**. Weekly performance of the T-LA sub-model using predicted $T\text{-}LAw_{i-1}$ as input (left) and using measured $T\text{-}LAw_{i-1}$ as input (right).

Figure 4 compares the predicted fresh weight from the model with the actual fresh weight measured during the third week of the experiment.

Both data sets show a similar trend, suggesting that the model predictions generally align with the observed biomass data. However, there are instances where the predicted values either underestimate or overestimate the actual biomass. For fresh weight values below approximately 5000 mg and above 20,000 mg, the discrepancies between predicted and observed values become more pronounced, indicating a reduction in model accuracy for biomass levels. This decline in accuracy could be attributed to the limited availability of training data for higher and lower fresh weight values. Moreover, the model's inputs are restricted to environmental parameters, potentially overlooking other factors that might influence plant growth in experimental cycles with higher or lower fresh weight. For instance, proximity to LED lights or localized temperature could play a significant role but are not explicitly accounted for in the current model.

## Total leaf area sub-model performances

For the T-LA sub-model, the random forest algorithm was used to predict the weekly total leaf area of the lettuce. Table 5 presents the weekly performance of the model using predicted $T\text{-}LAw_{i-1}$ and measured $T\text{-}LAw_{i-1}$ as the

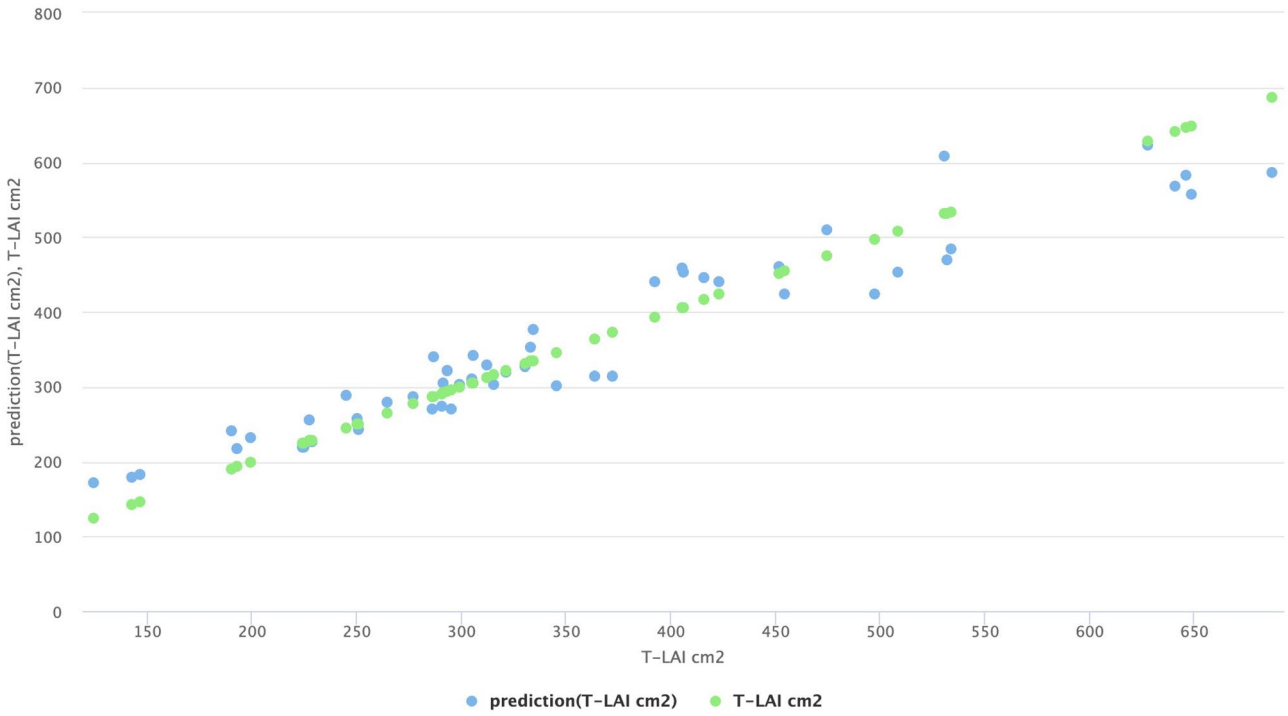



**Fig. 5**. T-LA predicted vs. T-LA measured at the third week.

| Week | MAPE | $R^2$ |
|---|---|---|
| Week 1 | 9.45% | 0.72 |
| Week 2 | 14.87% | 0.51 |
| Week 3 | 20.80% | 0.73 |

**Table 6**. Weekly perfomance of the N sub-model.

input variable. The measured T-LAw$_{i-1}$ is the total leaf area measured during the experimental activity described in section "Experimental activity".

Also in this case, the comparison of performance shows that the differences are minimal, thus the sub-model with predicted T-LAw$_{i-1}$ is used in the hybrid model.

Figure 5 compares the predicted total leaf area from the model with the total leaf area measured during the third week of the experiment.

Similar to the FW sub-model, for total leaf area values below approximately 200 cm² and above 500 cm², there are greater discrepancies between the predicted and observed values. A similar hypothesis as the one proposed for the FW sub-model can be considered.

To enhance the interpretability of the proposed hybrid model, feature importance was examined based on the most interpretable algorithm used, Random Forest. This analysis identified the environmental parameters that most strongly influence crop growth. For both fresh weight and leaf area estimation, the most influential variables were temperature, light intensity, and relative humidity. These findings are consistent with established plant physiological processes: light intensity promotes photosynthesis and biomass accumulation; temperature regulates enzymatic activity and metabolic rates; and relative humidity influences gas exchange, including transpiration and $CO_2$ uptake.

## Nitrates sub-model performances

For the N sub-model, different algorithms were used for each week: DNN for the first week, RF for the second, and SVM for the third to predict the amount of nitrates in lettuce crops. Table 6 presents the weekly performance of the model. As previously mentioned in Section "Machine learning sub-models", Nw$_{i-1}$ is not included as an input due to biological reasons. Instead, the predicted fresh weight from the FW sub-model is used as an input.

Compared to the other two sub-models, the nitrate model shows lower performance. This may be attributed to the nitrate dataset used for training, which had fewer collected data points due to the elimination of rows during the pre-processing phase. This elimination was necessary due to unreliable nitrate measurements conducted in the laboratory.

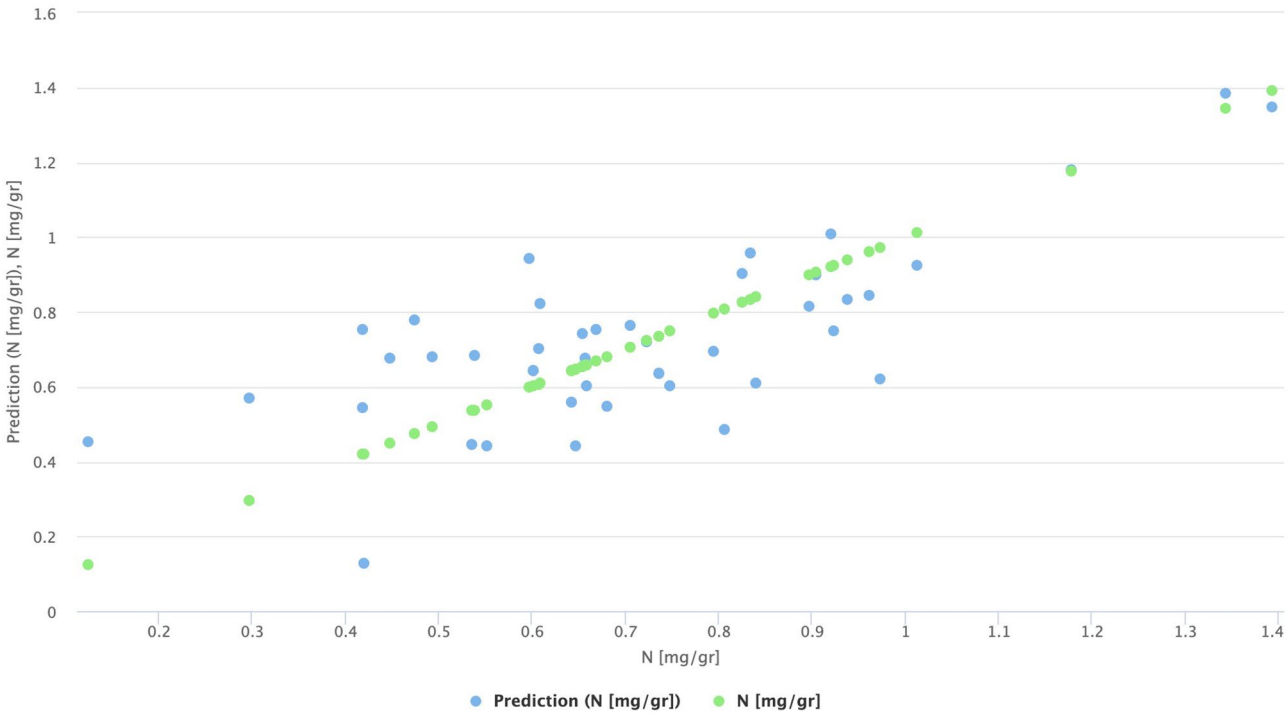



**Fig. 6**. Predicted Nitrate vs. Measured Nitrate at the third week.

| Week | MAPE | $R^2$ |
|---|---|---|
| Week 1 | 17.57% | 0.69 |
| Week 2 | 26.56% | 0.66 |
| Week 3 | 26.59% | 0.76 |

**Table 7**. Weekly performance of the water consumption sub-model.

Figure 6 compares the predicted nitrate from the model with the nitrate measured during the third week of the experiment.

The nitrate sub-model demonstrates lower performance than the fresh weight and total leaf area sub-models. This may be due to the fact that the nitrate sub-model was trained with fewer data points than the other two machine learning models. Moreover, the model's inputs are restricted to environmental parameters, potentially overlooking other factors that might influence nitrate development in experimental cycles. For instance, proximity to LED lights could play a significant role but is not explicitly accounted for in the current model.

To enhance the interpretability of the proposed hybrid model, feature importance was examined based on the most interpretable algorithm used, Random Forest. In the case of nitrate content, the model indicates that higher fresh weight and increased light intensity are associated with reduced nitrate accumulation in the leaves, likely due to more efficient nitrogen assimilation during accelerated growth. Conversely, EC was identified as a positive driver of nitrate levels, as it reflects the concentration of nutrients, particularly nitrates, available in the solution.

## Water consumption sub-model performances

The following section presents the performance of the physics-based model used to estimate the plant's water consumption. Once $Et_c$ was calculated using formula (3) given in section "Physics-based sub-model" water consumption per plant was determined by multiplying $Et_c$ by the plant's covered area, which is estimated from the total leaf area. Thus, the predicted T-LA, used to estimate the total leaf area of the lettuce, is an input for the physics-based model of water consumption.

Table 7 shows the performance of the model by comparing the values estimated by the model with the actual values measured during the experimental activity.

The performance of this sub-model, when compared to those for fresh weight and leaf area, is worse. This may be due to the assumption that the system has no water inefficiencies, such as leaks in the irrigation system, leading to the conclusion that water consumption is solely from plant and soil transpiration. Moreover, the model's inputs are restricted to environmental parameters, potentially overlooking other factors that might influence plant growth and water consumption. For instance, proximity to LED lights or localized temperature increases could play a significant role but were not explicitly accounted for in the current model.

These findings align with similar studies in the literature, suggesting that the model's performance could be enhanced under certain conditions[49].

## Discussion

The hybrid model presented in this article allows for the weekly estimation of fresh lettuce weight based on the growing environmental conditions and the predicted fresh weight from the previous week. The model's fresh weight estimates are quite precise, with an $R^2$ ranging from 0.93 to 0.94 and a MAPE between 10.60% and 12.05%. Additionally, the model can estimate the total leaf area of lettuce with similarly high performance, showing an $R^2$ ranging from 0.81 to 0.93 and a MAPE between 9.17% and 13.80%. The hybrid model can also estimate the nitrate content in lettuce on a weekly basis, using the predicted fresh weight and the environmental conditions.

Finally, the hybrid model includes a physics-based sub-model to estimate weekly water consumption per plant, taking into account the growing environmental conditions and the surface area covered by the crop. As currently developed, the proposed hybrid model has been trained using data from an experimental setup involving lettuce cultivation in a controlled aeroponic environment. Thus, its current applicability is most relevant for growers operating in similar controlled conditions. Nevertheless, by accepting environmental parameters as inputs throughout the crop cycle, the model provides weekly predictions of key agronomic outputs including fresh weight, leaf area, nitrate content, and water consumption, supporting both predictive and prescriptive use. If environmental conditions are fixed, the model can predict outcomes; if adjustable, it can be used to optimize these parameters toward targeted productivity or quality goals. For instance, a grower cultivating lettuce in a controlled aeroponic system could use the model at the start of a production cycle to input expected environmental conditions (e.g., light intensity, temperature, electrical conductivity). If the model predicts excessive nitrate accumulation for a given week, the grower may decide to reduce nitrogen supply or increase light intensity in advance to maintain quality standards. Likewise, predictions of low fresh weight may prompt the grower to modify environmental conditions, such as increasing light intensity or temperature, to improve yield outcomes. In practical terms, a grower operating in a greenhouse or vertical farm could input environmental data for the upcoming growth cycle and use the model's output to inform decisions such as adjusting nutrient delivery to meet nitrate targets or optimizing environmental inputs to reduce water consumption without compromising yield.

The estimation of nitrate content is less accurate than fresh weight and leaf area, with an $R^2$ ranging from 0.51 to 0.73 and a MAPE between 9.45% and 20.80%. These lower performances may be due to the nitrate sub-model being trained on a smaller dataset compared to the other two machine learning models. Similarly, the water consumption sub-model shows lower performance, with an $R^2$ ranging from 0.66 to 0.76 and a MAPE between 17.57% and 26.59%. This discrepancy may stem from the fact that the physics-based sub-model was originally designed for field cultivation, while the system under study operates under soilless conditions. Indeed, the Penman-Monteith model was applied in this study under controlled aeroponic conditions using a set of adapted assumptions. In this context, some environmental parameters such as wind speed and net radiation may have an altered role, and total water consumption may include components beyond pure transpiration, such as system inefficiencies. These aspects should be considered when interpreting the model's outputs in soilless, enclosed systems.

Another limitation is the narrow range of environmental conditions used in training. This restricts the model's ability to generalize to suboptimal or extreme scenarios, limiting its robustness under variable climate conditions. Also, both data collection and model prediction are currently conducted on a weekly basis; while suitable for medium-term decision-making, a finer temporal resolution could improve the responsiveness of the model in real-world applications.

The model can be retrained and adapted to suit other soilless controlled environment systems with similar characteristics. However, extension to fundamentally different cultivation systems or open-field agriculture would require more extensive retraining and validation. Expanding the training data, particularly for nitrate content and water consumption, would likely improve the robustness of those sub-models. With sufficient data availability, replacing the physics-based irrigation model with a fully data-driven sub-model may also enhance accuracy.

In future developments, the model can be trained using data collected under more extreme environmental conditions to better capture plant stress responses and improve its decision-support capabilities under challenging climates. Additionally, including secondary input parameters, such as proximity to LED lighting or microclimatic variations, may improve estimation performance for variables sensitive to spatial heterogeneity within the growing environment. Moreover, since the model currently predicts crop growth over a three-week cycle, it would be interesting in future work to train it on longer growth periods. In such extended cycles, incorporating longer temporal dependencies could be particularly valuable, as it would allow the model to capture cumulative physiological effects that develop over time. Additionally, integrating multi-step prediction frameworks and more advanced time-series architectures (e.g., LSTM or GRU) could further improve the model's capacity to capture longer-term dependencies and enhance its utility in forward-looking agricultural planning.

The model's architecture supports the integration of additional output variables. For instance, developing a sub-model to estimate root parameters would be particularly valuable in soilless systems, where root health is critical. Predicting energy consumption could also provide a more complete picture of resource efficiency, helping to guide sustainability strategies in controlled-environment agriculture.

Lastly, although the current model version has been trained exclusively on lettuce, retraining it on data from other crops would allow its extension to different plant species, including fruit-bearing and aromatic plants. This would considerably broaden the model's applicability and value across diverse agricultural systems.

## Conclusion

The hybrid model developed in this study demonstrates a promising approach for supporting precision agriculture in controlled soilless environments. By combining machine learning with a physics-based sub-model, it enables weekly predictions of key agronomic parameters, fresh weight, total leaf area, nitrate content, and water consumption, based on current environmental conditions and previous predictions. The model achieves high predictive accuracy for fresh weight ($R^2$ = 0.93–0.94, MAPE = 10.60–12.05%) and leaf area ($R^2$ = 0.81–0.93, MAPE = 9.17–13.80%), while nitrate content and water consumption show lower yet acceptable performance (nitrate: $R^2$ = 0.51–0.73, MAPE = 9.45–20.80%; water: $R^2$ = 0.66–0.76, MAPE = 17.57–26.59%).

Beyond predictive accuracy, the model provides actionable insights. It allows growers to improve environmental inputs such as temperature, light intensity, and electrical conductivity to achieve productivity or quality goals. For instance, the model can alert growers to potential nitrate accumulation or low biomass early in the cycle, enabling timely interventions. Feature importance analysis reinforces the physiological plausibility of the model, identifying light intensity, temperature, and relative humidity as key factors driving plant growth.

The hybrid model also supports prescriptive decision-making. By simulating how changes in environmental conditions affect crop outcomes it enables scenario planning and proactive management. Although trained on data from a controlled aeroponic system, the model architecture is flexible and can be retrained for other soilless systems or crops. Nonetheless, its generalizability is currently limited by the narrow range of environmental conditions in the training data and the weekly resolution of predictions, which may limit responsiveness in rapidly changing scenarios.

Several areas for improvement have been identified. Enhancing data coverage, particularly for nitrate content and water consumption, would likely improve the performance of the corresponding sub-models. Replacing the physics-based irrigation sub-model with a fully data-driven alternative could also yield better results, especially in soilless environments where traditional assumptions may not fully apply. Expanding training data to include more extreme environmental conditions would improve robustness under stress scenarios, while incorporating microclimatic variables or spatial heterogeneity (e.g., proximity to light sources) could boost predictive precision.

Future developments could also focus on extending the model to longer crop cycles or with sampling at closer time intervals (e.g., daily instead of weekly), using time-series architectures to capture time-dependent physiological dynamics. Additional outputs, such as root development or energy use, would offer a more holistic view of crop performance and resource efficiency. Finally, adapting the model to different crops, including fruit-bearing or aromatic plants, would significantly broaden its applicability and utility in modern agricultural systems.

## Data availability

The data supporting this study are available upon request from the corresponding author.

Received: 20 March 2025; Accepted: 15 May 2025
Published online: 02 July 2025

## References

1. Division, P. United Nations. Department of Economic and Social Affairs, World Population Prospects. Summary of Results, 2024. Accessed: Feb. 26, 2025. [Online]. Available: https://population.un.org/wpp/assets/Files/WPP2024_Summary-of-Results.pdf
2. Malhi, G. S., Kaur, M. & Kaushik, P. Impact of climate change on agriculture and its mitigation strategies: A review, Feb. 01, *MDPI*. (2021). https://doi.org/10.3390/su13031318
3. Kumar, A., Singh, S., Gaurav, A. K., Srivastava, S. & Verma, J. P. Plant Growth-Promoting Bacteria: Biological Tools for the Mitigation of Salinity Stress in Plants, Jul. 07, *Frontiers Media S.A.* (2020). https://doi.org/10.3389/fmicb.2020.01216
4. Rahman, S. U. et al. Jul., The interactive effect of pH variation and cadmium stress on wheat (Triticum aestivum L.) growth, physiological and biochemical parameters, *PLoS One*, vol. 16, no. 7 July, (2021). https://doi.org/10.1371/journal.pone.0253798
5. Young, M. D., Ros, G. H. & de Vries, W. Impacts of agronomic measures on crop, soil, and environmental indicators: A review and synthesis of meta-analysis. *Oct. 01 2021 Elsevier B V* https://doi.org/10.1016/j.agee.2021.107551
6. Lal, R. Soil carbon sequestration to mitigate climate change. *Nov* https://doi.org/10.1016/j.geoderma.2004.01.032 (2004).
7. Lask, J. et al. A parsimonious model for calculating the greenhouse gas emissions of miscanthus cultivation using current commercial practice in the United Kingdom, *GCB Bioenergy*, vol. 13, no. 7, pp. 1087–1098, Jul. (2021). https://doi.org/10.1111/gcbb.12840
8. Tudi, M. et al. Agriculture development, pesticide application and its impact on the environment. *Feb 01 2021 MDPI AG* https://doi.org/10.3390/ijerph18031112
9. Food and Agriculture Organization of the United Nations. *The state of food and agriculture. Moving forward on food loss and waste reduction*. (2019).
10. Awouda, A. M. M., Fasciolo, B., Bruno, G. & Razza, V. Cyber-physical system framework for efficient management of indoor farming production, in Contemporary Developments in Agricultural Cyber-Physical Systems, IGI Global, 66–86. doi: https://doi.org/10.4018/978-1-6684-7879-0.ch004. (2023).
11. Cowan, N. et al. CEA Systems: the Means to Achieve Future Food Security and Environmental Sustainability? Jun. 15, *Frontiers Media S.A.* (2022). https://doi.org/10.3389/fsufs.2022.891256
12. Fasciolo, B., Awouda, A., Bruno, G. & Lombardi, F. A smart aeroponic system for sustainable indoor farming, in *Procedia CIRP*, Elsevier B.V., pp. 636–641. (2023). https://doi.org/10.1016/j.procir.2023.02.107
13. Eldridge, B. M. et al. Getting to the roots of aeroponic indoor farming, *New Phytologist*, vol. 228, no. 4, pp. 1183–1192, Nov. (2020). https://doi.org/10.1111/nph.16780
14. Rufí-Salís, M. et al. Recirculating water and nutrients in urban agriculture: an opportunity towards environmental sustainability and water use efficiency? *J. Clean. Prod.* **261** https://doi.org/10.1016/j.jclepro.2020.121213 (Jul. 2020).
15. Kasilingam, S. et al. Physics-based and data-driven hybrid modeling in manufacturing: a review. *Prod. Manuf. Res.* **12** (1). https://doi.org/10.1080/21693277.2024.2305358 (2024).
16. Wang, J., Li, Y., Gao, R. X. & Zhang, F. Hybrid physics-based and data-driven models for smart manufacturing: modelling, simulation, and explainability. *J. Manuf. Syst.* **63**, 381–391. https://doi.org/10.1016/j.jmsy.2022.04.004 (Apr. 2022).

17. Cohen, A. R. et al. Dynamically Controlled Environment Agriculture: Integrating Machine Learning and Mechanistic and Physiological Models for Sustainable Food Cultivation, Jan. 14, *American Chemical Society*. (2022). https://doi.org/10.1021/acsestengg.1c00269
18. Gertphol, P., Chulaka & Changmai, T. Predictive models for lettuce quality from internet of Things-based hydroponic farm., (2018).
19. Mokhtar, A. et al. Using machine learning models to predict hydroponically grown lettuce yield. *Front. Plant. Sci.* **13** https://doi.org/10.3389/fpls.2022.706042 (Mar. 2022).
20. Ojo, M. O., Zahid, A. & Masabni, J. G. Estimating hydroponic lettuce phenotypic parameters for efficient resource allocation. *Comput. Electron. Agric.* **218** https://doi.org/10.1016/j.compag.2024.108642 (Mar. 2024).
21. Chang, C. L., Chung, S. C., Fu, W. L. & Huang, C. C. Artificial intelligence approaches to predict growth, harvest day, and quality of lettuce (Lactuca sativa L.) in a IoT-enabled greenhouse system, *Biosyst Eng*, vol. 212, pp. 77–105, Dec. (2021). https://doi.org/10.1016/j.biosystemseng.2021.09.015
22. Asy'ari, M. Z., Aten, J. F. C. & Prasetyo, D. Growth predictions of lettuce in hydroponic farm using autoregressive integrated moving average model, *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 6, pp. 3562–3570, Dec. (2023). https://doi.org/10.11591/eei.v12i6.4820
23. Metin, A., Kasif, A. & Catal, C. Temporal fusion transformer-based prediction in aquaponics. *J. Supercomput.* **79** (17), 19934–19958 (2023).
24. Elsherbiny, O. et al. Advancing lettuce physiological state recognition in IoT aeroponic systems: A meta-learning-driven data fusion approach. *Eur. J. Agron.* **161**, 127387 (2024).
25. Taha, M. F. et al. Deep Learning-Enabled dynamic model for nutrient status detection of aquaponically grown plants. *Agronomy*, **14**, 10, (2024).
26. Rajendiran, G. & Rethnaraj, J. Optimizing lettuce crop yield prediction in an indoor aeroponic vertical farming system using IoT-Integrated machine learning regression models. *Revue d'Intelligence Artificielle*, **38**, 3, (2024).
27. Shovic, J. C. & Shovic, J. C. *Raspberry Pi IoT Projects* (Springer, 2016).
28. Pennisi, G. et al. Optimal photoperiod for indoor cultivation of leafy vegetables and herbs. *Eur. J. Hortic. Sci.* **85** (5), 329–338 (2020).
29. Brechner, M., Both, A. J. & Staff, C. E. A. Hydroponic lettuce handbook. *Cornell Controlled Environ. Agric.* **834**, 504–509 (1996).
30. Chen, X. et al. Growth and nutritional properties of lettuce affected by different alternating intervals of red and blue LED irradiation. *Sci. Hortic.* **223**, 44–52 (2017).
31. Barla, S. A., Salachas, G. & Abeliotis, K. Assessment of the greenhouse gas emissions from aeroponic lettuce cultivation in Greece. *EuroMediterr J. Environ. Integr.* **5**, 1–7 (2020).
32. Méndez-Guzmán, H. A. et al. Iot-based monitoring system applied to aeroponics greenhouse, *Sensors*, vol. 22, no. 15, p. 5646, (2022).
33. Ragaveena, S., Shirly, A., Edward & Surendran, U. Smart controlled environment agriculture methods: A holistic review. *Rev. Environ. Sci. Biotechnol.* **20** (4), 887–913 (2021).
34. Hosseini, H. et al. Nutrient use in vertical farming: optimal electrical conductivity of nutrient solution for growth of lettuce and Basil in hydroponic cultivation. *Horticulturae* **7** (9), 283 (2021).
35. Tunio, M. H. et al. Influence of atomization nozzles and spraying intervals on growth, biomass yield, and nutrient uptake of butterhead lettuce under aeroponics system. *Agronomy* **11** (1), 97 (2021).
36. Wang, J., Lu, W., Tong, Y. & Yang, Q. Leaf morphology, photosynthetic performance, chlorophyll fluorescence, stomatal development of lettuce (Lactuca sativa L.) exposed to different ratios of red light to blue light. *Front. Plant. Sci.* **7**, 250 (2016).
37. Pennisi, G. et al. Optimal light intensity for sustainable water and energy use in indoor cultivation of lettuce and Basil under red and blue leds. *Sci. Hortic.* **272**, 109508 (2020).
38. Albornoz, F., Lieth, J. H. & González-Fuentes, J. A. Effect of different day and night nutrient solution concentrations on growth, photosynthesis, and leaf NO3-content of aeroponically grown lettuce. *Chil. J. Agric. Res.* **74** (2), 240–245 (2014).
39. BSI Standards Publication Foodstuffs. *BSI Standards Publication Foodstuffs-Determination of nitrate and/or nitrite content Part 2: HPLC/IC method for the determination of nitrate content of vegetables and vegetable products*. (2018).
40. Fasciolo, B. et al. An evaluation of research interests in vertical farming through the analysis of KPIs adopted in the literature. *Feb 01 2024 Multidisciplinary Digit. Publishing Inst. (MDPI)* https://doi.org/10.3390/su16041371
41. li Chen, X. et al. Growth and nutritional properties of lettuce affected by different alternating intervals of red and blue LED irradiation, *Sci Hortic*, vol. 223, pp. 44–52, Sep. (2017). https://doi.org/10.1016/j.scienta.2017.04.037
42. Anjana, S. U. & Iqbal, M. Nitrate accumulation in plants, factors affecting the process, and human health implications. A review. *Agron. Sustain. Dev.* **27** (1), 45–57. https://doi.org/10.1051/agro:2006021 (2007).
43. Spyromitros-Xioufis, E., Tsoumakas, G., Groves, W. & Vlahavas, I. Multi-target regression via input space expansion: treating targets as inputs. *Mach. Learn.* **104**, 55–98 (2016).
44. Bhatla, S. C. & Lal, M. A. *Plant Physiology, Development and Metabolism* (Springer Nature, 2023).
45. Kaminishi, A. & Kita, N. Seasonal change of nitrate and oxalate concentration in relation to the growth rate of spinach cultivars, (2006).
46. Hastie, T., Tibshirani, R. & Friedman, J. The elements of statistical learning, *Citeseer*. (2009).
47. (Rick, R. G. *G.) Allen and Food and Agriculture Organization of the United Nations., Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements* (Food and Agriculture Organization of the United Nations, 1998).
48. Rapid Miner.
49. Hua, D., Hao, X., Zhang, Y. & Qin, J. Uncertainty assessment of potential evapotranspiration in arid areas, as estimated by the Penman-Monteith method. *J. Arid Land.* **12** (1), 166–180. https://doi.org/10.1007/s40333-020-0093-7 (2020).


## Acknowledgements

This work was supported by the Agritech National Research Center and received funding from the European Union Next-GenerationEU (Piano nazionale di ripresa e resilienza (PNRR) – Missione 4 componente 2, investimenti 1.4 – D.D. 1032 17/06/2022, CN00000022) and PNRR – Decreto Ministeriale n. 1061. This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

## Author contributions

Conceptualization B.F., G.B. ; methodology B.F, N.G. ; formal analysis B.F.,; investigation B.F., N.G. ; writing—original draft B.F, N.G.; writing—review and editing G.B., P.C., B.F, N.G ; supervision G.B.,P.C.; fundings G.B.,P.C.

## Declarations

### Competing interests
The authors declare no competing interests.

### Additional information
**Correspondence** and requests for materials should be addressed to B.F.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025