Renewable and Sustainable Energy Reviews 197 (2024) 114423

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Can AI predict the impact of its implementation in greenhouse farming?

Siamak Hoseinzadeh *, Davide Astiaso Garcia

Department of Planning, Design, Technology of Architecture, Sapienza University of Rome, Via Flaminia 72, 00196, Rome, Italy

ARTICLE INFO

Keywords:
Artificial intelligence
Energy consumption
Greenhouse
Agriculture
Sustainability
CO2 emissions

ABSTRACT

The integration of Artificial Intelligence offers transformative solutions to modern-day challenges, especially in sectors like agriculture that are pivotal for human sustenance. This study underscores the profound impact of Artificial Intelligence in conditioned agricultural practices within greenhouses, based on data from an agricultural competition where teams optimized greenhouse performance using Artificial Intelligence-driven mechanisms. Results indicate that Artificial Intelligence-enhanced control strategies can drastically reduce energy consumption, particularly heating loads, without compromising crop yield, quality, or profitability. In some instances, performance even surpassed conventional methods. However, there are areas like Carbon Dioxide emissions and water usage where enhancements are still essential. Building on these insights, the study further ventures into AI's potential to predict greenhouse production outcomes. Through rigorous assessment of various machine learning models, the Radial Basis Function model exhibited commendable performance, achieving an Root Mean Squared Error of 0.8 and an R-squared value of 0.98 post-optimization. This establishes the feasibility of precisely forecasting greenhouse production rates in terms of kg/m^2. While this research predominantly centers on production volume, it lays a strong foundation for the predictive potential of AI in greenhouse operations and underlines the benefits of input optimization. It paves the way for future research focused on both the quality and quantity of greenhouse production.

1. Introduction

The persistent rise in the global population and the subsequent food demands have intensified the need for optimizing agricultural production. As available arable land diminishes, there has been a surge in the adoption of agricultural greenhouses to counter the overuse of natural resources. Within these controlled environments, the pivotal determinants for optimal plant growth encompass indoor air temperature, humidity, soil temperature, light intensity, and carbon dioxide concentration [1]. By fine-tuning these parameters, it is possible to create a sustainable solution that meets our escalating food needs while conserving natural resources. This, along with the unconscious use of chemicals in farming, creates mounting pressure on natural resources and harms the environment [2,3].

Agricultural producers consume non-renewable energy, leading to a decline in environmental quality [4]. Agriculture can significantly reduce and manage energy use [5]. Many environmental issues relating to food safety, climate change, carbon emissions, and waste created throughout the supply chain are raised worldwide due to inadequate ecological care [6]. Renewable energy is gaining significance as the world strives towards a more sustainable future by minimizing carbon emissions. Embracing renewable energy sources can diminish dependence on fossil fuels and curtail greenhouse gas emissions, thus aiding in mitigating the effects of climate change [7,8].

One of the practical solutions is to control the amount of energy consumption by the greenhouse structure [9]. Greenhouse farming can influence energy consumption, particularly water consumption [10]. The EU is estimated to have 405,000 ha of greenhouses, including glass and plastic-covered structures, as reported by the FAO [11]. Greenhouse farming plays a crucial role in energy consumption and decreasing reliance on non-renewable energy sources [12]. A review of energy consumption data for Spain, Greece, Italy, The Netherlands, and

Abbreviations: HVACD, Heating, Ventilation, Air-Conditioning, and Dehumidification; AI, Artificial Intelligence; ML, Machine Learning; RMSE, Root Mean Squared Error; RBF, Radial Basis Function; SVM, Support Vector Machine; GPR, Gaussian Process Regression; ANN, Artificial Neural Network; MAPE, Mean Absolute Percentage Error; TSSE, Total Sum of Squared Error; EF, Efficiency Factor; LM, Levenberg–Marquardt; CGF, Conjugate gradient backpropagation with Fletcher-Reeves updates; CGB, Conjugate gradient backpropagation with Polak-Ribiere updates.; BR, Bayesian Regularization backpropagation; FG, Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton backpropagation.

* Corresponding author.
E-mail addresses: siamak.hosseinzadeh@uniroma1.it, hoseinzadeh.siamak@gmail.com (S. Hoseinzadeh).

https://doi.org/10.1016/j.rser.2024.114423
Received 18 December 2023; Received in revised form 13 February 2024; Accepted 30 March 2024
Available online 6 April 2024

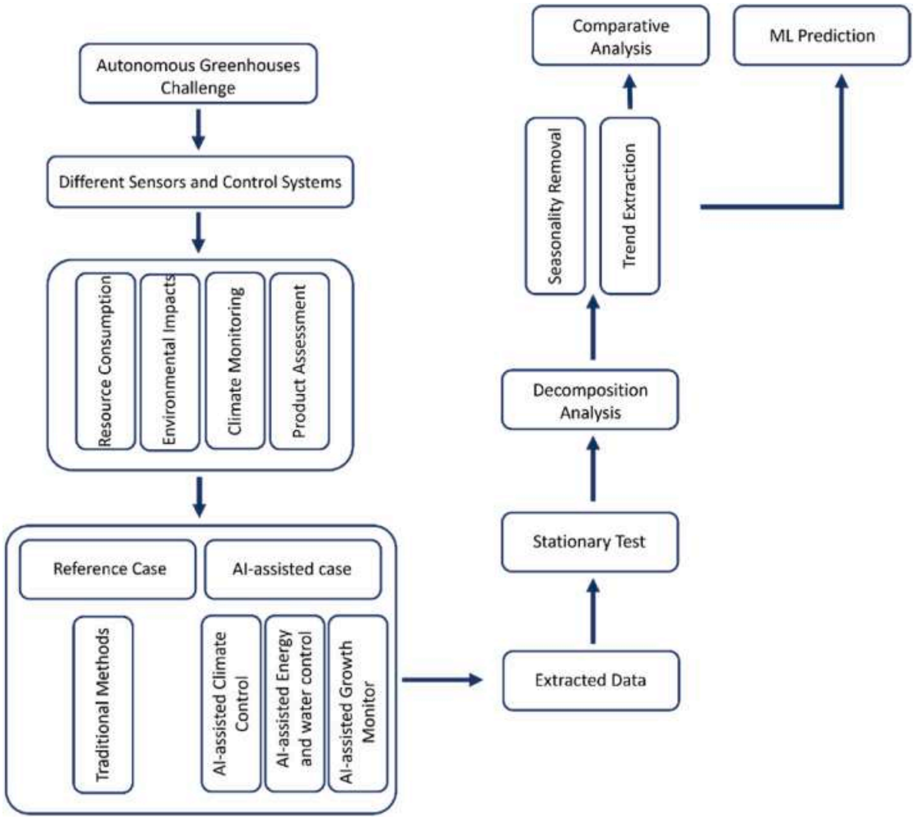


Fig. 1. Research workflow.

Germany, along with tomato production in the EU, was reported. The findings revealed that irrigation and fertilizer account for the highest energy usage, ranging from 1 to 19% and 1–27%, respectively [11]. Nonetheless, technological advancements in agriculture have aided in meeting consumers' ever-changing requirements [6]. Utilizing Artificial Intelligence (AI) efficiently decreases energy consumption in agriculture, particularly in greenhouse farming, while it can increase yields [13]. Furthermore, technologies such as AI, data-driven, machine learning, robots, and other systems can contribute optimization of greenhouse farming [14]. At the same time, it can reduce and control the amount of energy usage [15]. AI can be applied to control the environment, monitor crops, predict products, optimize resources, and automate processes [13,16]. The potential for AI to revolutionize greenhouse farming is significant, as it can enhance efficiency, minimize waste, increase crop yields, and ultimately lead to more sustainable and productive food systems [2,17,18].

Forecasting the efficiency of a greenhouse, especially regarding its indoor conditions or product quality, poses difficulties due to its reliance on several external elements [19]. In certain conditions, the natural environment might not favor ideal crop growth, given that factors such as temperature, relative humidity, the level of photosynthetically active radiation (PAR), and carbon dioxide concentrations influence plant progression [20]. The evolution of automation and AI has fueled the rise of intelligent greenhouses. These greenhouses incorporate tools and systems designed to boost product quantity and quality while curbing energy use [21]. The main objective of these instruments is to deploy smart control algorithms that regulate indoor climate factors, such as humidity, temperature, CO2, and light, to streamline and economize energy use.

Therefore, this study aims to explore two interconnected areas related to the application of AI in sustainable greenhouse farming. Firstly, the impact of using AI-based technologies on the performance of the greenhouse is studied. Secondly, the capability of machine learning algorithms for prediction of this impact is explored. Investigated data belongs to a case installed Autonomous Greenhouses International Challenge (AGIC) organized by Wageningen University & Research (WUR) in 2019. The objective of AGIC was to employ a data-driven approach and a data-management platform to materialize the concept of an autonomous greenhouse. The teams comprising AGIC were tasked with remotely producing a cherry tomato harvest within the duration of the test. WUR facilitated this project by providing each interdisciplinary team with a controlled greenhouse space and equipment at their agricultural research facility in the Dutch town of Bleiswijk. The project aimed to explore the application of machine learning techniques in greenhouse farming and its potential impact on crop cultivation. This experiment took over six months to prove AI control in 6 months. The data that will be analyzed in the following is taken from the results of all teams. They integrated AI methods to test software and sensors in greenhouse farming.

2. Objectives, hypotheses and contribution

2.1. Objectives and hypotheses

Building upon the outlined challenges and opportunities within the realm of sustainable greenhouse farming, this study sets forth a clear and comprehensive objective. At the heart of our investigation is the assessment of the applicability and efficacy of AI predictive models in transforming greenhouse agriculture. Specifically, the study aims to evaluate how AI-aided strategies can significantly enhance resource consumption optimization and production improvements within these controlled environments. The focus extends beyond mere technological application; we seek to understand the nuanced interplay between AI-driven interventions and their tangible impacts on agricultural efficiency and output.

By leveraging advanced AI algorithms, the study's objective is to dissect and analyze the multifaceted effects of these technologies on reducing energy and water use, while simultaneously boosting crop yield and quality. This dual-focus approach not only addresses the critical need for resource conservation but also aligns with the broader goals of sustainable agricultural practices. Through a detailed examination of AI's role in predictive analysis and operational optimization, our study endeavors to provide a holistic view of the potential benefits and limitations of integrating AI into greenhouse farming. This exploration is pivotal in laying the groundwork for future innovations and strategies that could redefine the standards of agricultural productivity and environmental stewardship.

The study hypothesizes that AI can fulfill dual objectives within sustainable greenhouse farming. Firstly, AI has the potential to optimize resource use, not only enhancing energy and water efficiency but also ensuring higher quality produce. Secondly, AI can serve as a tool for comprehensive impact assessments, evaluating the effectiveness of AI implementations from various perspectives. Such assessments could provide invaluable insights for decision-makers in the field, aiding in the strategic deployment of AI technologies to meet both economic and environmental goals. This hypothesis underscores the multifaceted value of AI in agriculture, suggesting a transformative role in advancing sustainable practices and informed decision-making.

2.2. Contribution and innovation

The study's innovation is deeply holistic, extending aI-driven approaches to not only enhance energy efficiency in greenhouse farming but also to significantly improve water conservation and reduce environmental impacts while improving the production in terms of quality and quantity. This comprehensive strategy ensures that the AI solutions proposed not only optimize resource use but also set new benchmarks for environmental sustainability in agriculture with higher standards for productions. By integrating considerations for crops, energy, water, and ecological well-being, this work pioneers a new era of eco-conscious agricultural practices, where technology serves as a bridge to more sustainable and responsible farming methodologies.

The contribution of this work is multifaceted, initially showcasing the efficacy of AI in enhancing greenhouse operations through advanced control strategies. It then extends to a novel secondary application of AI for conducting predictive impact assessments, enabling a detailed evaluation of the economic and environmental benefits versus the costs of AI adoption in agriculture. This dual application provides critical insights into the feasibility and long-term sustainability of integrating sophisticated AI technologies into agricultural practices, highlighting the potential for significant advancements in efficiency and eco-friendliness.

3. Materials and methods

This study initially examines the impact of AI technology on greenhouse farming. This case study involved multiple research teams conducting a real-life experiment using new technologies, such as AI, to cultivate plants in greenhouses. The data obtained from the challenge was used to evaluate the impact of various technologies in the study. Experimental research and field studies on plants (either cultivated or wild), including the collection of plant material, complies with relevant institutional, national, and international guidelines and legislation.

To analyze the performance of the different stages of data analysis were performed to preprocess the data. Then, the results were subjected to a comparative analysis to study the effectiveness of AI technology in improving the performance of greenhouse farming through different perspectives Finally, using ML algorithms, the capability of AI for prediction of this impact is assessed as shown in Fig. 1 to reach the main goal which was the impact prediction of AI in sustainable greenhouse farming.

"Impact prediction of AI" within this scholarly discourse signifies a methodical forecast of the comprehensive outcomes and benefits attributable to the deployment of AI-enhanced strategies within the domain of greenhouse agriculture. This prognostic evaluation extends to a thorough examination of a spectrum of critical factors, notably reductions in water and energy utilization, the diminution of environmental footprints through decreased CO2 emissions, and the amplification in both the caliber and volume of agricultural produce. The core objective of such predictive endeavors is to establish an analytical foundation for the execution of nuanced cost-benefit analyses, thereby furnishing policymakers and practitioners with the requisite empirical evidence to make well-informed decisions concerning the incorporation and execution of AI-based methodologies in greenhouse settings.

This anticipatory analytical process is pivotal in delineating the potential scope and scale of AI's transformative impact on greenhouse agriculture, offering a strategic vantage point from which stakeholders can evaluate the viability, efficiency, and sustainability implications of adopting AI-driven innovations. By systematically forecasting the multifaceted impacts of AI applications, this approach provides a robust platform for decision-makers to assess the strategic value and operational feasibility of integrating advanced AI technologies into greenhouse farming practices, thereby facilitating a data-driven, strategic decision-making paradigm that enhances the resilience, productivity, and environmental sustainability of agricultural systems.

To explain the research workflow, the "Autonomous Greenhouse Challenge" has been selected as a case study where different teams employed "Different Sensors and Control Systems". For each time "Resource Consumption", "Environmental Impacts", "Climate Monitoring", and "Product Assessment" analyses were conducted. Then in order to study the impacts of AI in greenhouse farming one "Reference case" confront some "AI-assisted cases" where different "AI-assisted climate control", "AI- assisted energy and water control", and "AI-assisted growth monitor" systems were deployed. After 6 months of this experiment "Extracted data" was subjected to "Stationary test" and "Decomposition analysis" to have "Seasonality removal" and "Trend extraction". Then, the outcomes were subjected to the "Comparative analysis" to fully understand the impacts of AI methods in greenhouse performance. On the final stage, the extracted and processed data, then used as the input for ML algorithm for prediction and according to standard procedure, after fine tuning the model performance assessment was conducted to study the accuracy of the model.

4. Case study

In this section, detailed information about the case study is presented to have a comprehensive insight about it. Data related to the case study is open access under CC0 1.0 license and plants (either cultivated or wild), including the collection of plant material, complies with relevant institutional, national, and international guidelines and legislation. Therefore, the permission to collect cherry tomato. Have been granted according to the license and also information on the voucher specimen and who identified it is included in the data repository and 4TU. Federation is responsible for that.

4.1. Smart machines and control systems

5 compartments of greenhouse with the area of 96 m^2 and 76.8 m^2 of growing space was monitored during the experiment. The crop, a single tomato variety, was planted in the middle of December 2019. The substrate is particular slabs (Grodan's slab) used by each team. These slabs are exceptionally uniform in both quality and absorbency. This provides a considerable benefit for the roots and plant health by allowing for much more accurate regulation of the water content level and EC in the root zone. Also, the slabs do not emit any dirt, keeping the greenhouse spotless.

To give slab data and monitor crop growth, GroSens sensors were

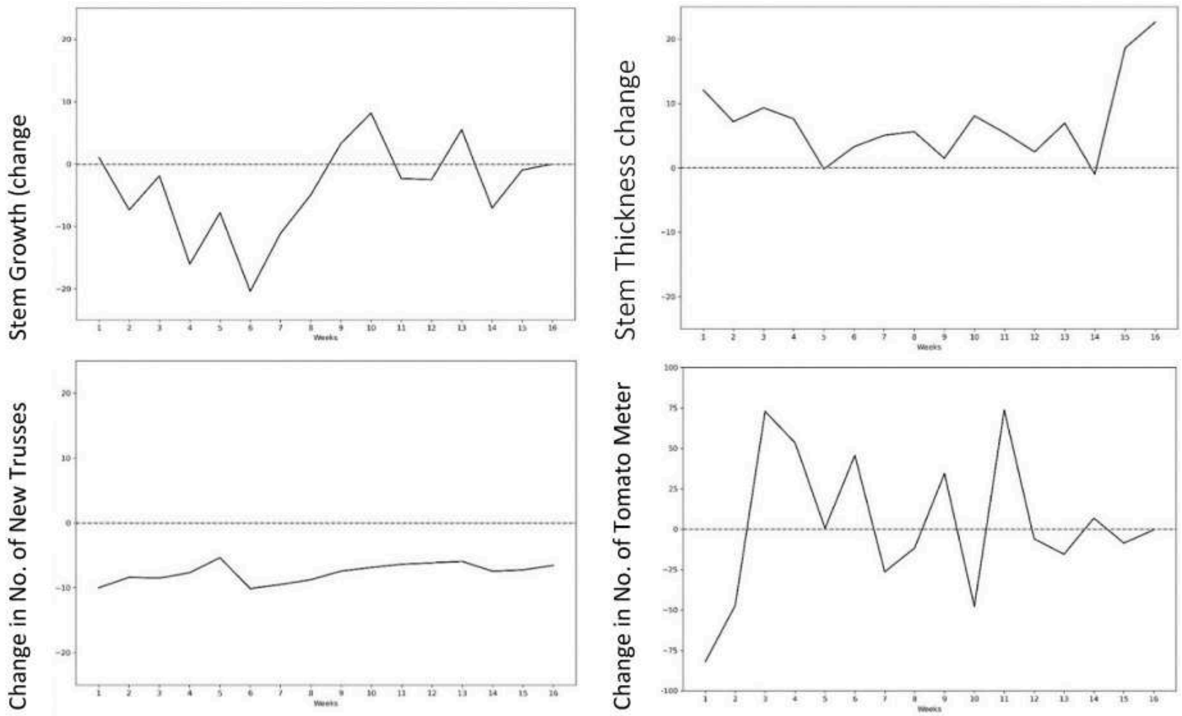


Fig. 2. Average AI-assisted greenhouse performance compared to the reference case.

fitted. It is an FDR-based sensor that measures the electrical characteristics of the slab alone as well as how much moisture is present in it. The greenhouse environment (temperature and relative humidity), and the plant profile, were acquired (the plant height, how many leaves, how many fruits per truss, and the stem thickness). Various sensors were utilized in this greenhouse to accomplish best the intended plant features, such as a sensor to assess leaf temperature.

4.2. Monitoring systems

The sensors used by the team not only measure the slab temperature and EC but also detect its weight to determine the water content, which is crucial for assessing the health of the plant, its foliage, and its fruits. However, GroSens plays a critical role in providing additional data that is equally important. By combining the weight data obtained from the slab with the information gathered by GroSens, the team can get a precise and comprehensive picture of the plant's condition. It is not merely a matter of acquiring more data but instead of obtaining more insightful and valuable data.

Data collected by sensors and webcams is transmitted almost continuously to servers. Every morning, this data is reviewed. The amount of irrigated water (L/m^2) is verified, and the drainage (L/m^2) is used to determine the drainage percentage. The trend of the slab water content (measured by the Grodan sensor) and slab weight (measured by the ioCrops sensor) is analyzed through a graph. The difference between the minimum and maximum values of slab water content and significance is evaluated to determine the necessary adjustments to the irrigation strategy, such as when to initiate or terminate irrigation, the amount and duration of irrigation, and the length of time between irrigation sessions.

4.3. Product and resources

An experiment was carried out using an unspecified variety of cherry tomato crop. On October 19, 2019, seedlings of the cv. "Axiany" (Axia Seeds, The Netherlands) were sown and then grafted onto Maxifort rootstock. The seedlings were planted in rock wool cubes and later transplanted to greenhouse compartments on December 16, 2019. Remote control of the experiment was assumed by the teams on December 20, 2019. The parameters which include Stem growth per week (m/week), Stem thickness (mm), umulative number of new set trusses on the stem (number/stem), Stem density (Stems/m^2), and Plant density (Plants/m^2) were checked weekly.

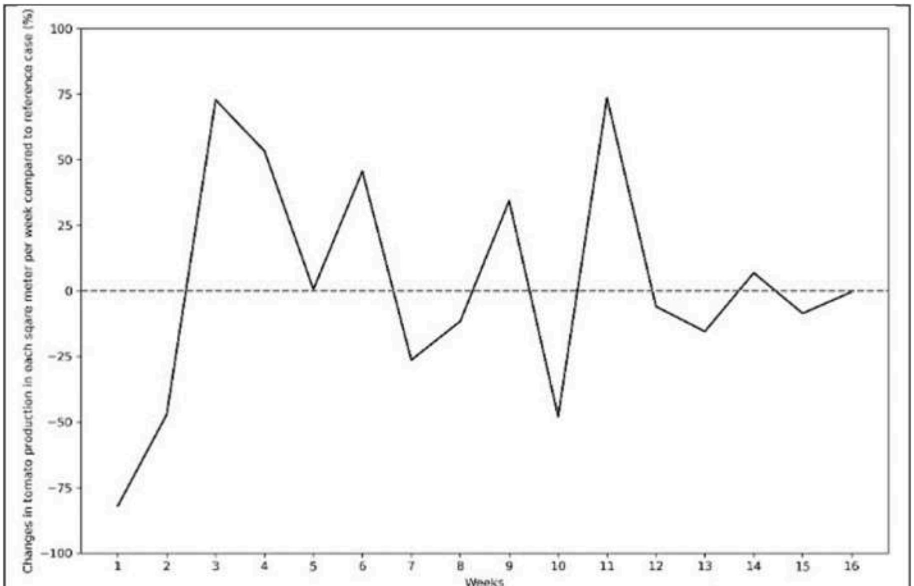
Resource energy consumption was calculated according to measured data for tomatoes: heat energy consumption (MJ/m^2) and electricity (kWh/kg) (artificial light) during pick-hours (7.00–23.00) and Electricity consumption (artificial light) during off-pick-hours kWh/m2, CO_2 emission (kg/m^2), drain water (L/m^2), irrigation water (L/m^2) was measured. These measurements are validated by by Wageningen University & Research (WUR) through actual metrics (resource metrics) after finishing the challenge.

5. Results and discussion

In this section, the results of analyses divided into two sections are presented. In the first section the impact of using AI technologies on the performance of greenhouse farming is studied. Secondly, the capability of ML algorithms for the prediction of this impact is assessed.

5.1. Impact of AI in sustainable greenhouse farming

To study the impact of AI technology on the performance of the greenhouse during the experiments, different teams tried to improve the performance of the greenhouse by applying different methods, algorithms, control systems to reach better results than the reference greenhouse which used the most common methods. The results are presented in this section which shows the average of team performance compared to the reference case. Indeed, irrespective of the applied method or control strategy or the machine learning algorithms, the critical point here is the fact that by using all these technologies what would be the outcomes. The real question here is that the implementation of these methods can lead to better results or not. At the first stage the quality of productions of AI-assisted greenhouse (all five team) where compared to the reference case.



Index	Average	Reference	Change	Index	Average	Reference	Change
1	0.06	0.32	**-82.06**	9	0.50	0.37	**34.43**
2	0.38	0.72	**-47.17**	10	0.47	0.91	**-47.96**
3	0.69	0.40	**72.90**	11	0.52	0.30	**73.67**
4	0.52	0.34	**53.41**	12	0.67	0.71	**-5.92**
5	0.49	0.49	**0.57**	13	0.56	0.66	**-15.48**
6	0.54	0.37	**45.68**	14	0.79	0.74	**6.89**
7	0.45	0.61	**-26.36**	15	0.74	0.81	**-8.57**
8	0.41	0.46	**-11.57**	16	0.64	0.64	**-0.38**

Fig. 3. Tomato weight per square meter of AI-assisted cases compared to the reference case.

In the context of the "Autonomous Greenhouse Challenge," it was ensured that the experimental setups, both for the reference case and those involving AI-assisted strategies, were uniformly standardized with respect to their physical dimensions. Each greenhouse was allocated an identical surface area, and the introduction of tomato plants into these environments was consistent across all experimental conditions. This uniformity was crucial to establish a level field for the comparative analyses between the AI-augmented cases and the reference scenario, eliminating any potential discrepancies in spatial or quantitative variables that could influence the outcomes. The variations in growth metrics observed, such as stem thickness, fruit yield per truss, and total production, were attributed to the strategic spatial arrangements of the tomato plants within the allocated areas. Decisions regarding the placement and spacing of plants, informed by data-driven insights from the challenge, were optimized to enhance environmental conditions conducive to plant growth. These included optimal light distribution, efficient air circulation, and effective resource utilization, all of which are fundamental to the healthy development of crops.

The differential outcomes in plant growth and yield across the various setups underscore the significant impact that AI-driven interventions can have on agricultural practices. By leveraging predictive analytics and data-driven insights, the challenge participants were able to make informed decisions that potentially improved the productivity and efficiency of greenhouse farming. This approach not only demonstrates the value of integrating AI into agricultural methodologies but also highlights the potential for such technologies to usher in advancements in sustainable and precision agriculture. Finally, since the size of tomatoes in the class A has the size range, and the fact that the number of tomatoes has not been specified in the public dataset, the choice of using stem thickness and length was the only available logical parameters that could reflect the growth parameters.

As it is depicted in Fig. 2, in terms of growth length of tomato stem during the whole experiment the reference case mainly showed a better performance. Nevertheless, in terms of steam thickness AI-assisted cases saw a meaningful better performance compared to the reference case. Having said that, there are other metrics that must be taken into account while assessing the product of each greenhouse. In here, the number of trusses was another metric that has been measured. Unexpectedly, none of team could reach a higher number of trusses compared to the reference case, while some teams during the experiments had higher number of tomatoes per square meters. This clearly means that the thicker stems have led to higher tomato per truss because while the average number of trusses for teams where less than the reference case, the total number of tomatoes were higher. The next step is to consider the density of stems and in order to do that at the same time with the number of tomatoes and their size, the weight of tomatoes per square meters could be an appropriate metric to consider.

The rationale for comparing the weight of tomatoes per square meter rather than the weight of the same number of tomatoes from each farm is multifaceted, taking into consideration both the methodological approach of the study and the nature of the dataset from the "Autonomous Greenhouse Challenge." Firstly, the classification of tomatoes into Class A, which includes tomatoes suitable for commercial trade within a specified size range, inherently minimizes the variability in tomato sizes across the farms. This classification ensures a relative uniformity in size, allowing for an equitable comparison of weight per square meter as an indicator of productivity. Furthermore, the study aims to assess the overall productivity of the greenhouse space, which is particularly relevant in agricultural practices where optimizing the use of limited space is crucial. By focusing on weight per square meter, the study

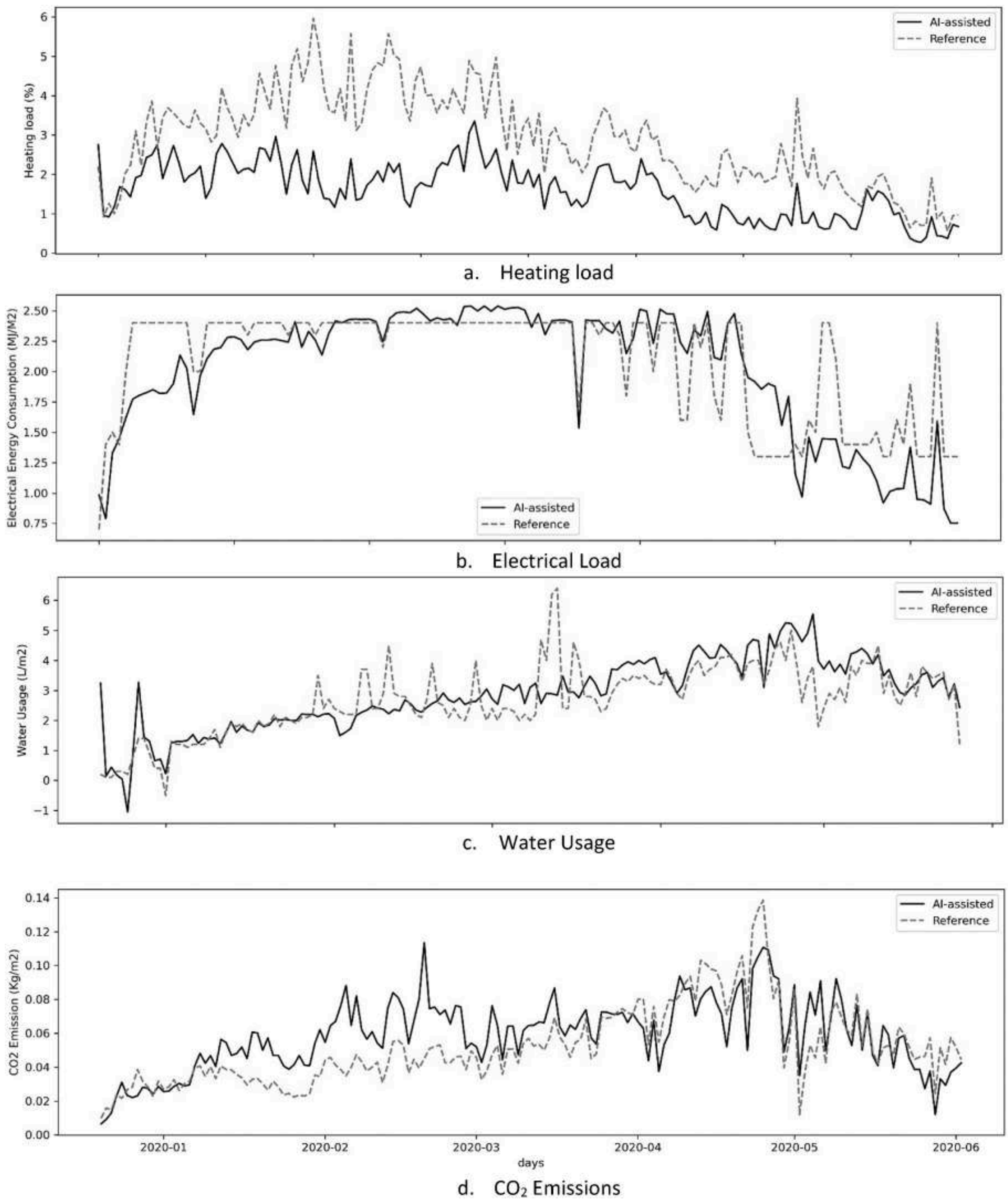


Fig. 4. Resource consumption and Environmental Impacts of cases.

Table 1
ADF results of resource data.

Resource	Type of case	ADF p-value	Stationary Data	Not Stationary data
Heating load	AI-assisted	0.26068		X
	Reference	0.34892		X
Electrical load	AI-assisted	0.91805		X
	Reference	0.15630		X
Water usage	AI-assisted	0.71840		X
	Reference	0.01908	X	
CO2 emission	AI-assisted	0.01486	X	
	Reference	0.12259		X

provides insights into how efficiently each square meter of greenhouse space is used to produce commercial-grade tomatoes. Additionally, the absence of specific data on the number of tomatoes produced in the "Autonomous Greenhouse Challenge" dataset necessitated an alternative approach. The weight of tomatoes per square meter offers a comprehensive metric that reflects the outcomes of strategic decisions regarding stem density and plant management practices influenced by AI-driven strategies.

This approach not only aligns with the objectives of maximizing greenhouse space utilization but also adheres to commercial agricultural standards, making the findings applicable to real-world agricultural operations. It captures the balance between the quantity and quality of the produce, crucial for commercial viability and sustainability in

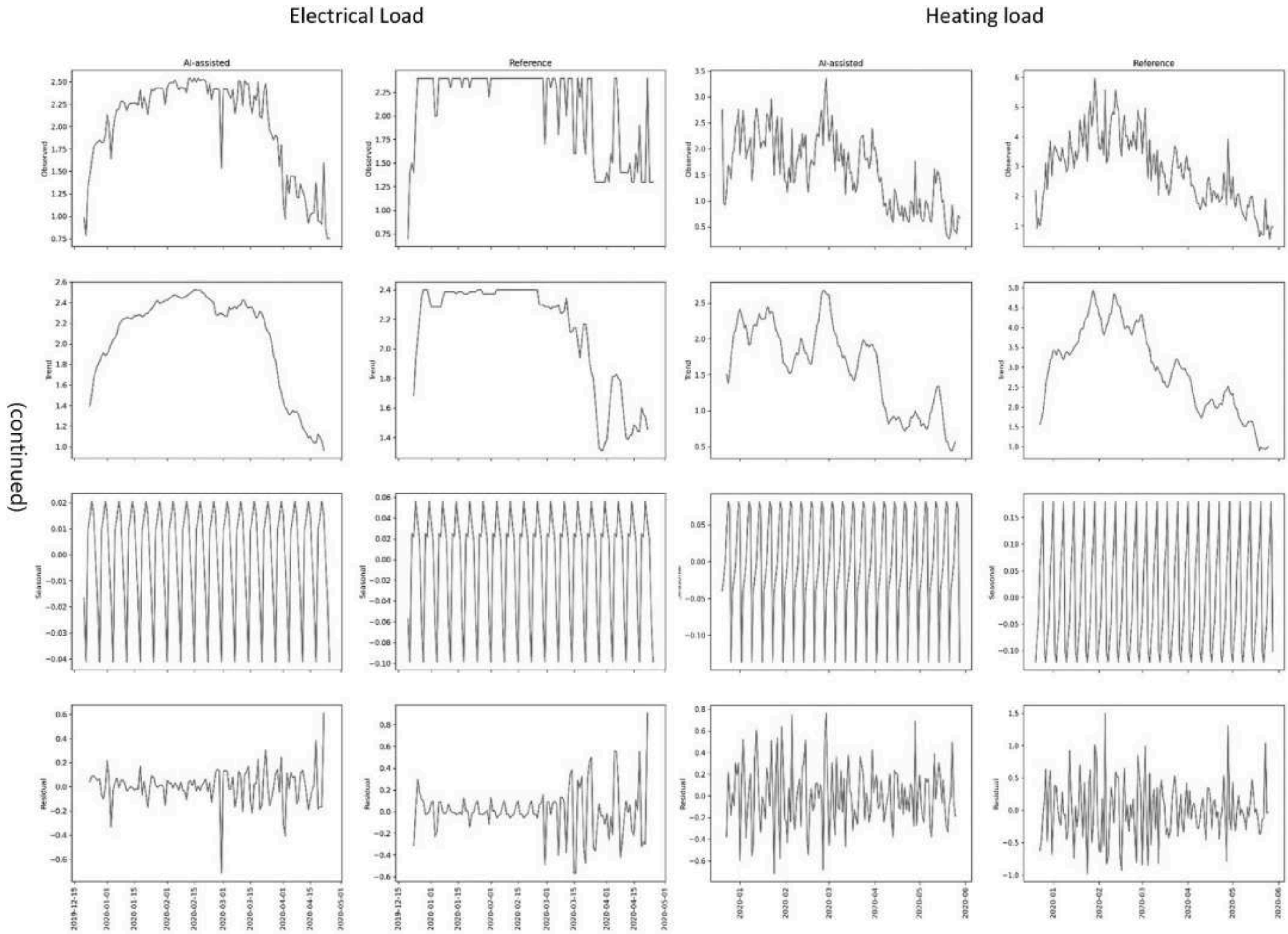


Fig. 5. Decomposition analysis of resource consumption and Environmental Impacts of cases.

greenhouse farming. Therefore, the decision to compare the weight of tomatoes per square meter, given the dataset limitations and the study's objectives, provides a holistic view of greenhouse productivity. This methodological choice is aligned with the overarching goals of precision agriculture and enhancing the commercial quality of agricultural output, within the constraints of available space and the commercial standards set for Class A tomatoes.

Fig. 3 shows that while on some stages of the experiment the products of AI-assisted cases where higher than the reference case, as the data is cumulative, at the end of the experiment the results were almost the same. It means that the wight of tomato products of reference case and the AI-assisted cases were the same. Considering the lower number of tomatoes in AI-assisted cases, it can be concluded that bigger and heavier tomatoes were achieved in AI-assisted greenhouses, but in general the total production weight is the same.

In the next stage of performance analysis of AI-assisted greenhouses compared to the reference case in terms of resources consumption were observed. Regarding heating load, by considering the extracted data from sensors it is clear that the average heating load of AI-assisted greenhouses is meaningfully lower than the average case that is shown in Fig. 4. This figure for water and electrical consumption shows very close figures. This means that AI technology showed a distinguishable advantage over the traditional methods for lowering the range of heating load in greenhouse farming. If the impacts of climate change in the form global warming is being taken into account here, it can be concluded that this energy load would have a lower importance. Having said that, extreme weather events as another possible consequence of climate change can add the significance of heating load. Therefore, there is no fixed and fast rule about the future of heating load in this respect. Regarding electrical load and water usage, in contrast to the expectation all sophisticated AL methods, control systems. Climate strategies could not reasonably add to the efficiency of the greenhouse. However, it should be noted that by monitoring cases with longer period of time the effectiveness of these strategies could be more analyzed and discussed.

Finally, in terms of environmental impacts of greenhouse, while at the beginning of the experiment it looed that the AI-assisted showed weaker performance in their environmental impacts, gradually the differences started to fade and the results became more similar.

The heating system utilized in both the reference and the experimental compartments comprised a rail pipe system at the floor level and an additional pipe heating system at crop height, predominantly powered by gas, although such systems can also operate on electricity. The preference for gas heating in greenhouse operations stems from its ability to deliver the significant thermal energy required to maintain temperatures conducive to optimal plant growth. Additionally, gas heating systems can contribute to increased CO2 levels within the greenhouse environment, fostering enhanced plant development. Gas heating is particularly prevalent in extensive or commercial greenhouse settings, where its efficiency and cost-effectiveness are paramount, especially in regions where gas is a more economically feasible option than electricity. The capability for independent control of these heating systems provides the necessary flexibility for precise thermal regulation,

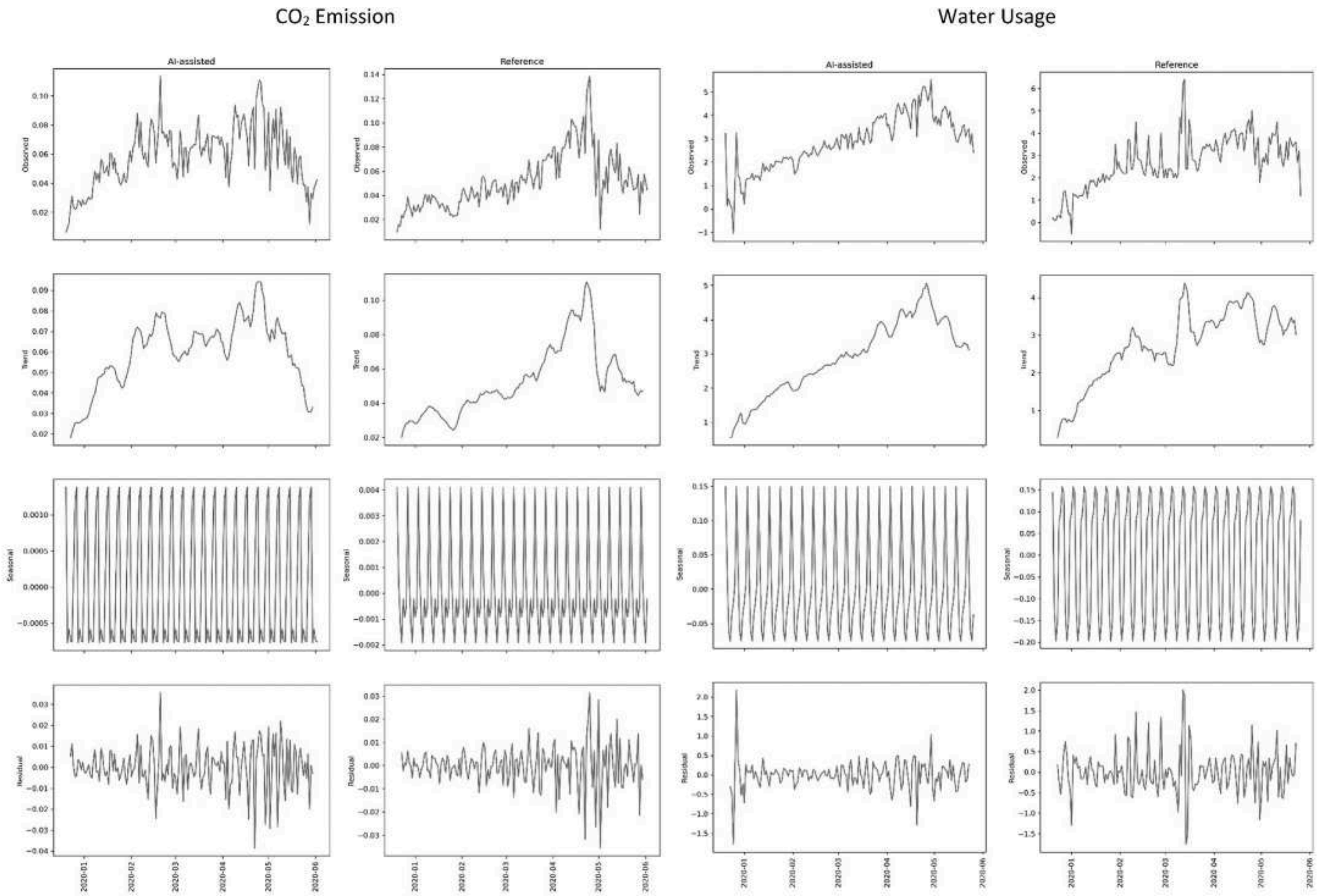


Fig. 5. (*continued*).

catering to the varying temperature requirements essential for different stages of crop growth.

Understanding the distinction between heating load and electricity consumption is essential in the context of greenhouse operations. The heating load is directly associated with the energy needed to sustain suitable temperature levels for plant cultivation, managed by the gas-powered heating systems. Conversely, electricity consumption within greenhouses is primarily linked to lighting requirements, with systems such as High-Pressure Sodium (HPS) and LED lights playing a crucial role in supplementing or replacing natural sunlight in controlled agricultural environments. These lighting systems represent a significant component of the total electrical load, independent of the energy used for heating.

Therefore, any fluctuations in electricity consumption within the greenhouse shown in Fig. 4 b are not directly correlated with the operation of the heating systemsdepicted in in Fig. 4 a. Since the heating mechanisms are primarily gas-powered, variations in the electrical load are more likely attributed to other energy-consuming systems, such as the lighting apparatus. This distinction is crucial for accurately interpreting energy usage patterns in greenhouse operations, where separate systems contribute to the overall energy footprint, each with its distinct energy source and function.

As the data here is a time series data, some preprocessing and time series analysis here can be helpful for better understanding the trend of the data. To deal with this type of data a common technique is decomposition analysis, which attempts to dissect a time series into its component parts. A pattern, a seasonal component, and a residual component are examples of these components. While the seasonal component captures the recurring patterns or cycles that occur within the time series, the trend component reflects the time series' long-term movement [22].The time series' random or erratic changes that cannot be explained by the trend or the seasonal components are represented by the residual component.

It would be useful to understand the structure of the data without seasonality and possible noises, perform better forecasting with higher reliability, and finally identify anomalies much easier in the trend. To perform the decomposition analysis, a deep understanding of the data is important, because depending on the type of time series an additive or multiplicative model should be chosen for decomposition model [23,24]. The seasonal and trend components of the series are combined together to create the observed values in an additive model, which makes the assumption that they remain stable over time. Alternatively said, regardless of the level of the series, the seasonal swings have a consistent amplitude. When seasonal fluctuations have a magnitude that is unrelated to trends or series levels, this kind of model is applicable.

In contrast, a multiplicative model presupposes that the observed values are the result of multiplying the seasonal and trend components of the series together rather than keeping them constant. In this instance, the level of the series determines whether the seasonal swings are higher or lower. When the size of seasonal swings is proportional to the level of the series, this kind of model is suitable.

One way to choose between additive and multiplicative model is checking if the data is stationary. The statistical characteristics of the time series, such as mean and variance, are consistent over time if the data is stationary. Use of an additive model for time series decomposition in these situations is frequently preferable than a multiplicative model. Therefore, checking if the data is stationary is an integral part of decomposition analysis. In several disciplines, including economics,

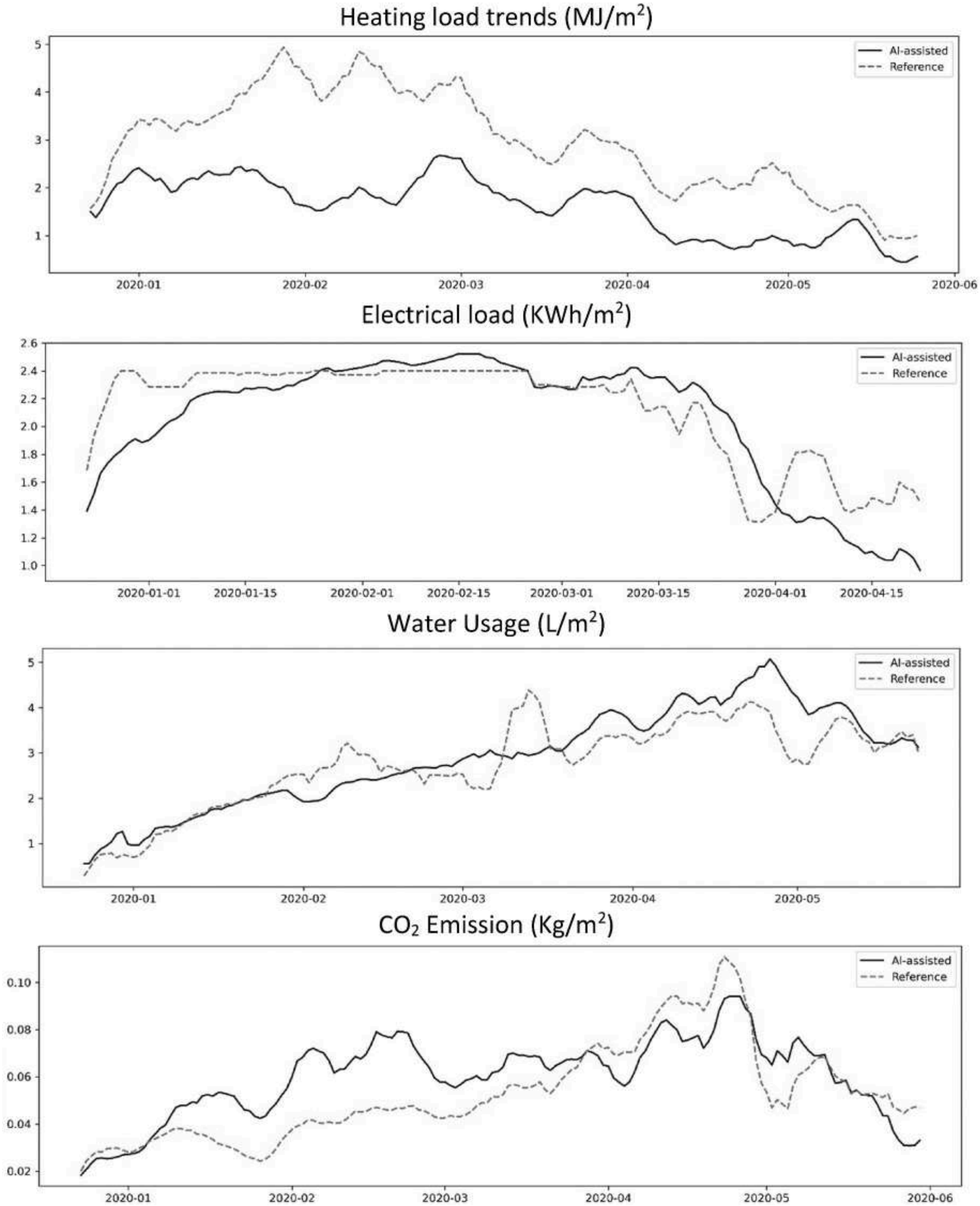


Fig. 6. Trends of resource consumption and environmental impacts of greenhouses after seasonality removal.

finance, and engineering, the Adfuller stationary test is a crucial tool for determining if time series data are stationary [25]. The results of ADF test for the resource data are presented in Table 1.

After choosing an appropriate model for each data, the decomposition analysis was performed. The results of decomposition analysis in Fig. 5 for the resource consumption and environmental impact variables are presented.

Results of decomposition analysis showed that the performance of the greenhouses, both AI-assisted ones and the reference one, were heavily seasonal. In other words, the behavior of cases in terms of resource consumption and environmental impacts had a fixed pattern over time which is visible in seasonal section of Fig. 6. After identifying seasonality in the data and removing the seasonality and white errors, the remaining ones would be the actual trends.

Trends revealed that the most influence of AI technology was on the heating load which was notable lower for AI-assisted cases and during the whole experiment time it was always lower than the reference case. Having said that, it is important to note that the difference between heating load of AI-assisted greenhouses and the reference case started with a high amount and ended with the lower amount. This means that there is a possibility that if the experiments continue, there might be situations where the heating load of the reference case happens to be

Table 2
Input and output of the dataset.

Variable	Description	Unit	Interval	dataset	Input/ Output
Tair	Greenhouse Air temperature	°C	5 min	Climate condition	Input
Rhair	Greenhouse relative humidity	%	5 min		
CO2air	CO2 greenhouse	ppm	5 min		
HumDef	Greenhouse humidity deficit	g/m^3	5 min		
VentLee	Leeward vents opening	%	5 min		
Ventwind	Windward vents opening	%	5 min		
AssimLight	HPS lamps status (on-off)	%	5 min		
EnScr	Energy curtain opening	%	5 min		
BlackScr	Blackout curtain opening	%	5 min		
PipeLow	Rail pipe Temperature (Lower circuit)	°C	5 min		
PipeGrow	Crop pipe Temperature (Growth circuit)	°C	5 min		
co2_dos	CO2 dosing	kg/ha hour	5 min		
Tot_PAR	Total inside PAR (Sun + HPS + LED)	μmol/ m^2 s	5 min		
ot_PAR_Lamps	PAR sum from HPS and LED lamps	μmol/ m^2 s	5 min		
EC_drain_PC	Drain EC	dS/m	5 min		
pH_drain_PC	Drain pH	[−]	5 min		
Water_sup	Cumulative number of minutes of irrigation in a day	minutes	5 min		
Cum_irr	Cumulative number of litres of irrigation in a day	L/m^2 day	5 min		

Variable	Description	Unit	Interval	dataset	Input/ Output
irr_PH	pH of irrigation water	[−]	bi-weekly	Lab analysis	Input
irr_EC	EC of irrigation water	[dS/m]	bi-weekly		
irr_NH4	Ammonium concentration in irrigation water	[mmol/ l]	bi-weekly		
irr_K	Potassium concentration in irrigation water	[mmol/ l]	bi-weekly		
rr_Na	Sodium concentration in irrigation water	[mmol/ l]	bi-weekly		
irr_Ca	Calcium concentration in irrigation water	[mmol/ l]	bi-weekly		
irr_Mg	Magnesium concentration in irrigation water	[mmol/ l]	bi-weekly		
irr_Si	Silicon concentration in irrigation water	[mmol/ l]	bi-weekly		
irr_NO3	Nitrate concentration in irrigation water	[mmol/ l]	bi-weekly		

Table 2 (*continued*)

Variable	Description	Unit	Interval	dataset	Input/ Output
irr_Cl	Chlorine concentration in irrigation water	[mmol/ l]	bi-weekly		
irr_SO4	Sulphate concentration in irrigation water	[mmol/ l]	bi-weekly		
irr_HCO3	Bicarbonate Ion concentration in irrigation water	[mmol/ l]	bi-weekly		
irr_PO4	Phosphate concentration in irrigation water	[mmol/ l]	bi-weekly		
irr_Fe	Iron concentration in irrigation water	[mmol/ l]	bi-weekly		
irr_Mn	Manganese concentration in irrigation water	[mmol/ l]	bi-weekly		
irr_Zn	Zinc concentration in irrigation water	[mmol/ l]	bi-weekly		
irr_B	Boron concentration in irrigation water	[mmol/ l]	bi-weekly		
irr_Cu	Copper concentration in irrigation water	[mmol/ l]	bi-weekly		
irr_Mo	Molybdenum concentration in irrigation water	[mmol/ l]	bi-weekly		
drain_PH	pH of drainage water	[−]	bi-weekly		
drain_EC	EC of drainage water	[dS/m]	bi-weekly		
drain_NH4	Ammonium concentration in drainage water	[mmol/ l]	bi-weekly		
drain_K	Potassium concentration in drainage water	[mmol/ l]	bi-weekly		
drain_Na	Sodium concentration in drainage water	[mmol/ l]	bi-weekly		
drain_Ca	Calcium concentration in drainage water	[mmol/ l]	bi-weekly		
drain_Mg	Magnesium concentration in drainage water	[mmol/ l]	bi-weekly		
drain_Si	Silicon concentration in drainage water	[mmol/ l]	bi-weekly		
drain_NO3	Nitrate concentration in drainage water	[mmol/ l]	bi-weekly		
drain_Cl	Chlorine concentration in drainage water	[mmol/ l]	bi-weekly		
drain_SO4	Sulphate concentration in drainage water	[mmol/ l]	bi-weekly		
drain_HCO3	Bicarbonate Ion concentration in drainage water	[mmol/ l]	bi-weekly		
drain_PO4	Phosphate concentration in drainage water	[mmol/ l]	bi-weekly		
drain_Fe	Iron concentration in drainage water	[mmol/ l]	bi-weekly		
drain_Mn	Manganese concentration in drainage water	[mmol/ l]	bi-weekly		
drain_Zn	Zinc concentration in drainage water	[mmol/ l]	bi-weekly		

(*continued on next page*)

Table 2 (*continued*)

Variable	Description	Unit	Interval	dataset	Input/ Output
drain_B	Boron concentration in drainage water	[mmol/ l]	bi-weekly		
drain_Cu	Copper concentration in drainage water	[mmol/ l]	bi-weekly		
drain_Mo	Molybdenum concentration in drainage water	[mmol/ l]	bi-weekly		
Prod	Total tomato Production quality	kg/m^2	at date (harvest)	Production	Output

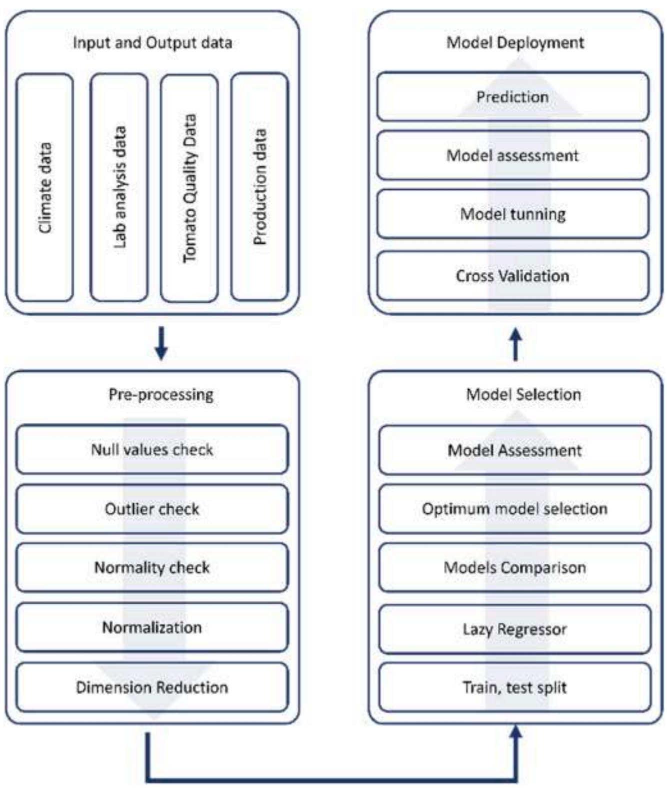


Fig. 7. ML deployment stages.

lower than AI-assisted. Surely, this is just a hunch and it needs deeper study to discuss it. Regarding electricity consumption, AI-assisted cases had almost similar behavior to the reference case, with some fluctuations. This similarity continued to happen for water usage as well where even in most days the water usage of the reference case was lower than the average of AI-assisted greenhouses.

Environmental impacts of AI-assisted greenhouses, however, had a different behavior where up to the middle of the experiment AI-assisted cases were weaker than the reference case but in the second half of the experiment, the average of their performance was quite close to the reference case.

5.2. *AI impact prediction*

5.2.1. *Pre-processing*

In the next stage of analysis, pre-processed data from the case study has been used to train ML algorithms in order to predict the performance of the AI-aided greenhouse in terms of the production that it can yield. In this process, 2 sets of data were used as the input to predict the prediction. Input datasets belong to indoor climate condition, and lab analysis of the irrigation and drain samples shown in Table 2).

There are some tasks that must be done before model deployment as depicted in Fig. 7. After importing different dataset and joining them with respect to the time series and the time interval of each dataset that are different. In pre-processing stage after checking and removing null values and outliers, normality check performed which results in an additional stage of normalization with MinMax scaler.

5.2.2. *Feature engineering*

Upon completing the data normalization process, we embarked on an in-depth analysis of the dataset's intricate inter-relationships. This was crucial in our pursuit to condense its dimensions and make the data more manageable for subsequent analysis. This comprehensive study involved examining the correlations between various data variables. We identified and eliminated those variables that demonstrated an exceedingly strong correlation, which is evident in Figs. 8–9. A specific correlation threshold was set at 0.75, and any variables that exceeded this limit were judiciously excluded, ensuring a streamlined and more manageable dataset.

To further refine the data, we undertook a meticulous sensitivity analysis. In this step, we compared each variable against the output variables. Variables that displayed correlations lower than 0.6 were deemed less influential and were subsequently removed to maintain only the most impactful data. From our dedicated efforts, we observed notable reductions in variable counts. Specifically, from the climate dataset, out of an initial 26 variables, only 6 were retained for in-depth subsequent analysis and for the eventual model deployment. Similarly, in the case of the laboratory analysis dataset, by adhering to the same correlation criteria, we saw a reduction from 38 variables down to a concise set of 8.

Such rigorous data refinement ensures that the input data for our machine learning algorithms remains not only precise and effective but also limited in scope. This approach holds significant advantages. A limited set of impactful variables not only boosts the algorithms' efficiency but also substantially reduces computational demands. This strategy paves the way for quicker processing times and ensures that computational resources are judiciously utilized, leading to significant cost savings.

Fig. 8 illustrates the interconnectedness of climate-related variables within the dataset through a correlation matrix, offering insights into their interdependent behaviors. This matrix facilitates a comparative analysis of different variables to discern any relationships, whether direct or inverse. For example, the correlation between the status of lamps (AssimLight) and CO2 levels (co2_vip) suggests a synchronous relationship; an increase in lamp activity typically coincides with a rise in CO2 levels, indicating that these variables behave in tandem. Conversely, a robust inverse relationship is observed between the relative humidity inside the greenhouse (Rhair) and the status of the ventilation systems (t_ventlee_vip, t_ventwind_vip). This indicates that enhanced ventilation is likely to decrease the relative humidity, aligning with expectations of greenhouse climate control.

Furthermore, accompanying density plots augment the correlation matrix by providing visual cues about the nature of the relationships between variables. They help in distinguishing whether the relationships are linear, suggesting a consistent rate of change between variables, or non-linear, which could imply more complex interactions that may vary under different conditions. This level of detail is particularly crucial for cause-effect studies, where understanding the nature of the relationship between variables is essential for drawing accurate conclusions about their interactions and impacts within the greenhouse environment. The combined use of correlation matrices and density plots thereby equips researchers with a nuanced understanding necessary for constructing and refining predictive models, ensuring that these models are based on accurate representations of the underlying data relationships.

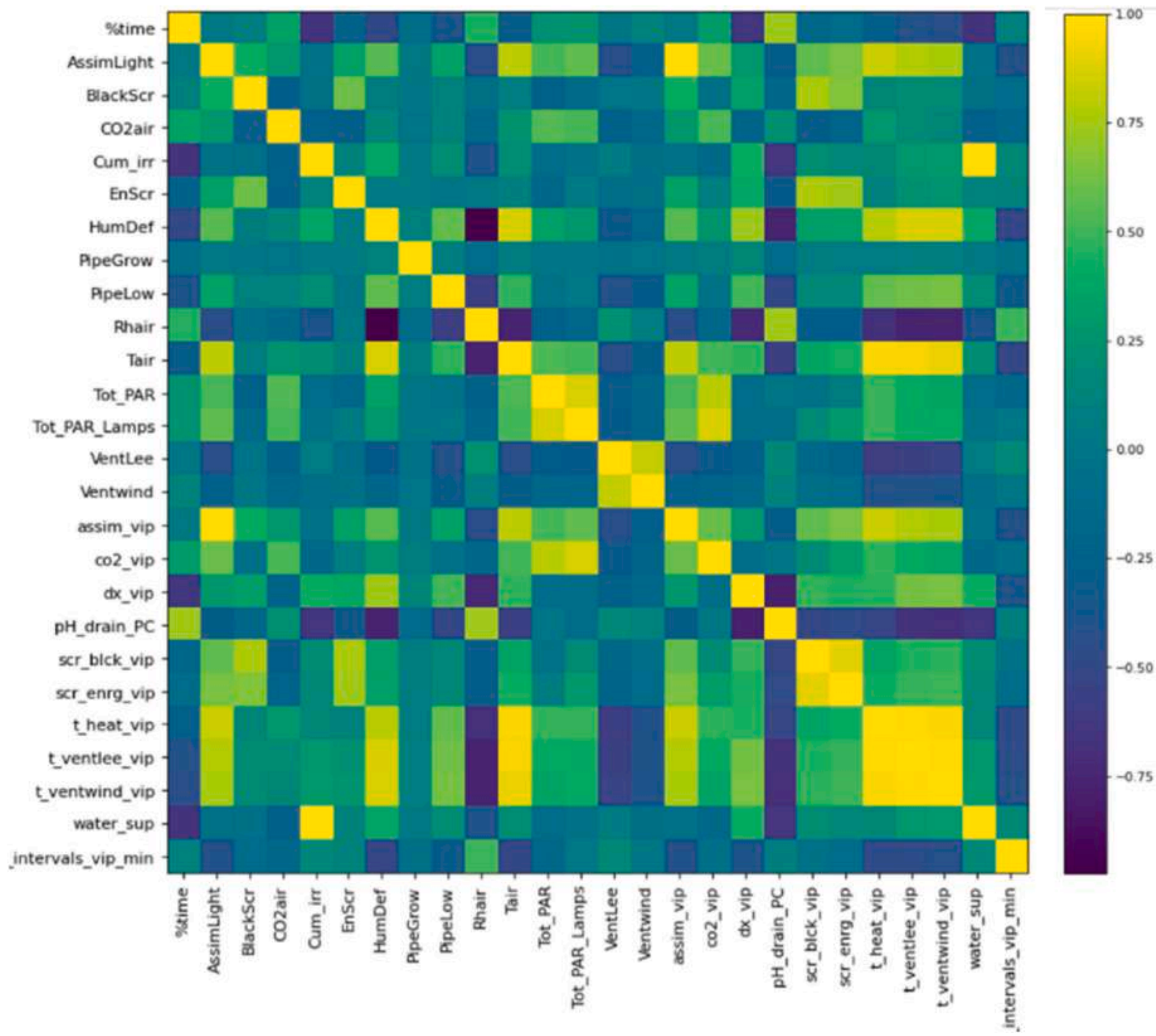


Fig. 8. Climate data correlation matrix and density plot.

Upon examining the laboratory data through the lens of correlation coefficients and density plots, it is evident that a subset of variables exhibits substantial positive and negative correlations. This statistical interdependence in Fig. 9, implies that during the feature engineering phase, such variables may be redundant and could thus be considered for removal to enhance model efficiency and manageability. The principle of dimensionality reduction comes into play here, advocating for a focus on variables that demonstrate varied behaviors to avoid multicollinearity, which could otherwise distort the predictive model's outcomes. For groups of variables that manifest strong correlations, the selection of a single representative variable for each group is suggested. This representative would encapsulate the shared information content of the group, thereby streamlining the dataset for further analysis and AI implementation. The aim is to distill the dataset to its most informative elements, eschewing superfluous data that does not contribute to the predictive power of the AI algorithms.

In the process of feature selection, density plots serve a dual purpose. First, they help confirm the linearity or non-linearity of the relationships between variables. Second, they assist in visualizing the distribution and overlap of data points, further informing which variables may be redundant. By implementing this rigorous approach to data refinement, the number of variables can be effectively reduced, easing the computational load and enhancing the clarity of the dataset. This meticulous method of variable categorization and reduction is critical for developing a robust AI framework capable of delivering accurate and insightful predictions.

Through an in-depth sensitivity analysis, we systematically assessed the features considering both their correlation with the intended outcome and their intricate inter-relationships. This rigorous process allowed us to distill our selections meticulously. From the vast array of variables in the climate cluster, only 6 were deemed paramount and thus retained. Similarly, from the laboratory analysis data, a selection of 8 stood out as especially relevant, as illustrated in Fig. 10. Consequently, our model has been refined and streamlined to incorporate only these 14 judiciously chosen input variables, all aimed at predicting a singular, focused output. This optimized approach not only ensures a more targeted predictive trajectory but also potentially enhances the model's overall efficiency and accuracy.

5.2.3. Model selection

Finally, in order to employ ML algorithms different traditional regressions models have been used at a basic level to compare and have an initial insight about selecting top models for further analyses. 'Therefore, through Lazy Regressor package for python different models have been tested with default settings and hyper parameters. Results of the top 10 models are depicted in Fig. 11 where there are three models with R-squared higher than 0.4 and RMSE less than two. These models are the optimum models for further analyses in this study.

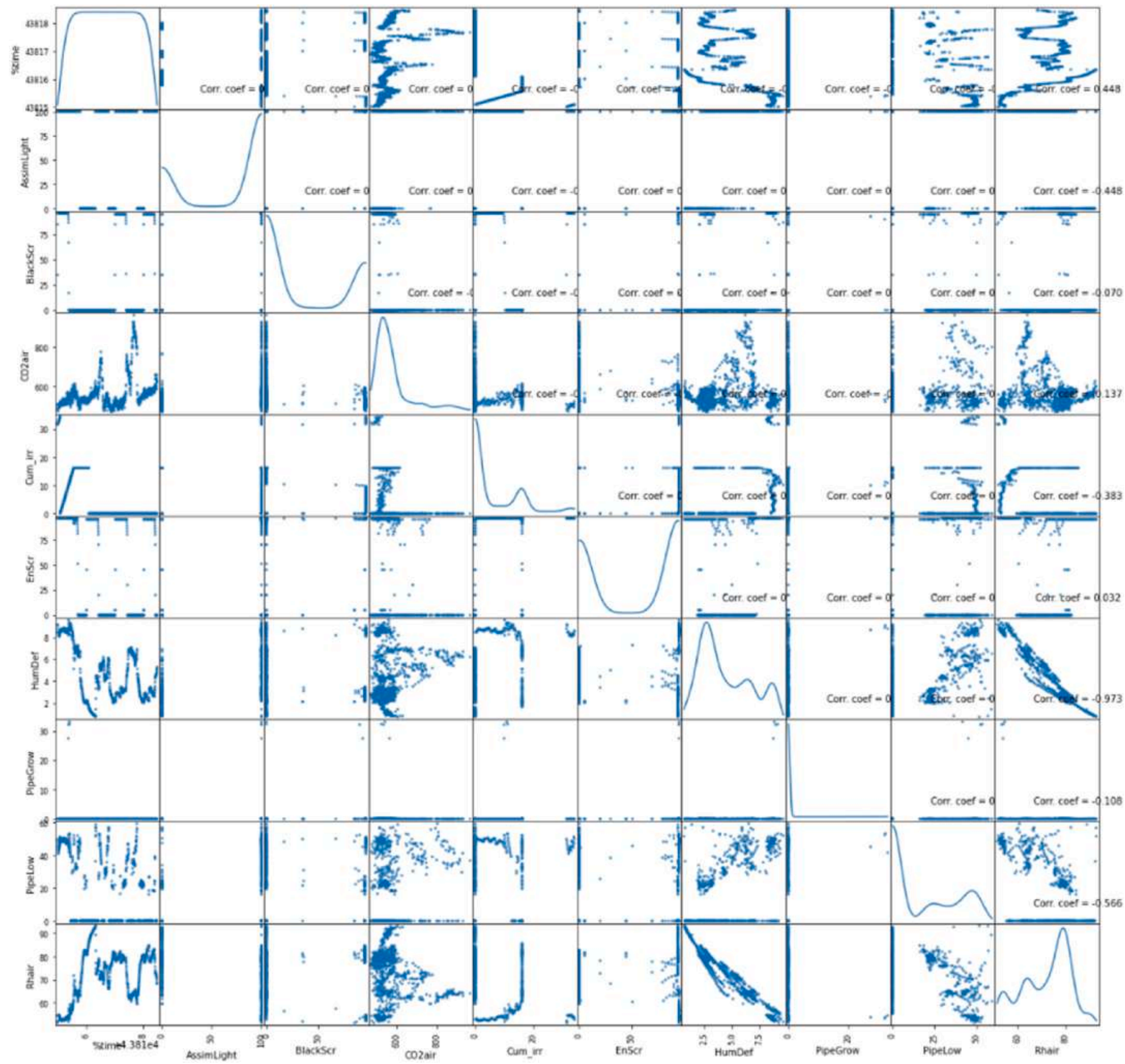


Fig. 8. (*continued*).

In this part in order to have a deeper insight about what are the optimum models, each of them are briefly introduced.

- **Radial Basis Function:** The Artificial Neural Network (ANN) model, recognized for its predictive prowess, generally consists of at least three distinct layers. The inaugural layer, termed the input layer, corresponds in size to the total number of model inputs. Within this framework, every input possesses a corresponding weight. The subsequent hidden layer is populated with several neurons. Their collective presence amplifies the efficacy of the ANN by maintaining an adequate neuron count within. Matching the network's output, the number of neurons in the final, output layer is set. Since in this work the final goal is to predict the product weight per square meter of the greenhouse a singular neuron is designated for the output layer. Conversely, the Radial Basis Function (RBF) approach sees each neuron in its hidden layer governed by a distinct nonlinear activation function. During the RBF network's training phase, the bias component is harnessed to guide the network towards a global minimum [26]. Here, a singular hidden layer model has been used while crafting the RBF model. The RBF's was optimized by altering the neuron count in the hidden layer between three and thrity five, culminating in the selection of the most optimal setup.
- **Support Vector Machine:** Support Vector Machines (SVMs) have gained considerable recognition for their effectiveness in solving classification problems, where the aim is to categorize data into distinct classes. While they are primarily known for this role, their application extends into regression analysis as well, albeit less commonly. In the context of regression, these models go by the name of Support Vector Regression (SVR). SVR models aim to predict continuous outcomes as opposed to discrete classes, and they share many of the same foundational principles with their classification counterparts. Despite being less documented, SVRs are gaining

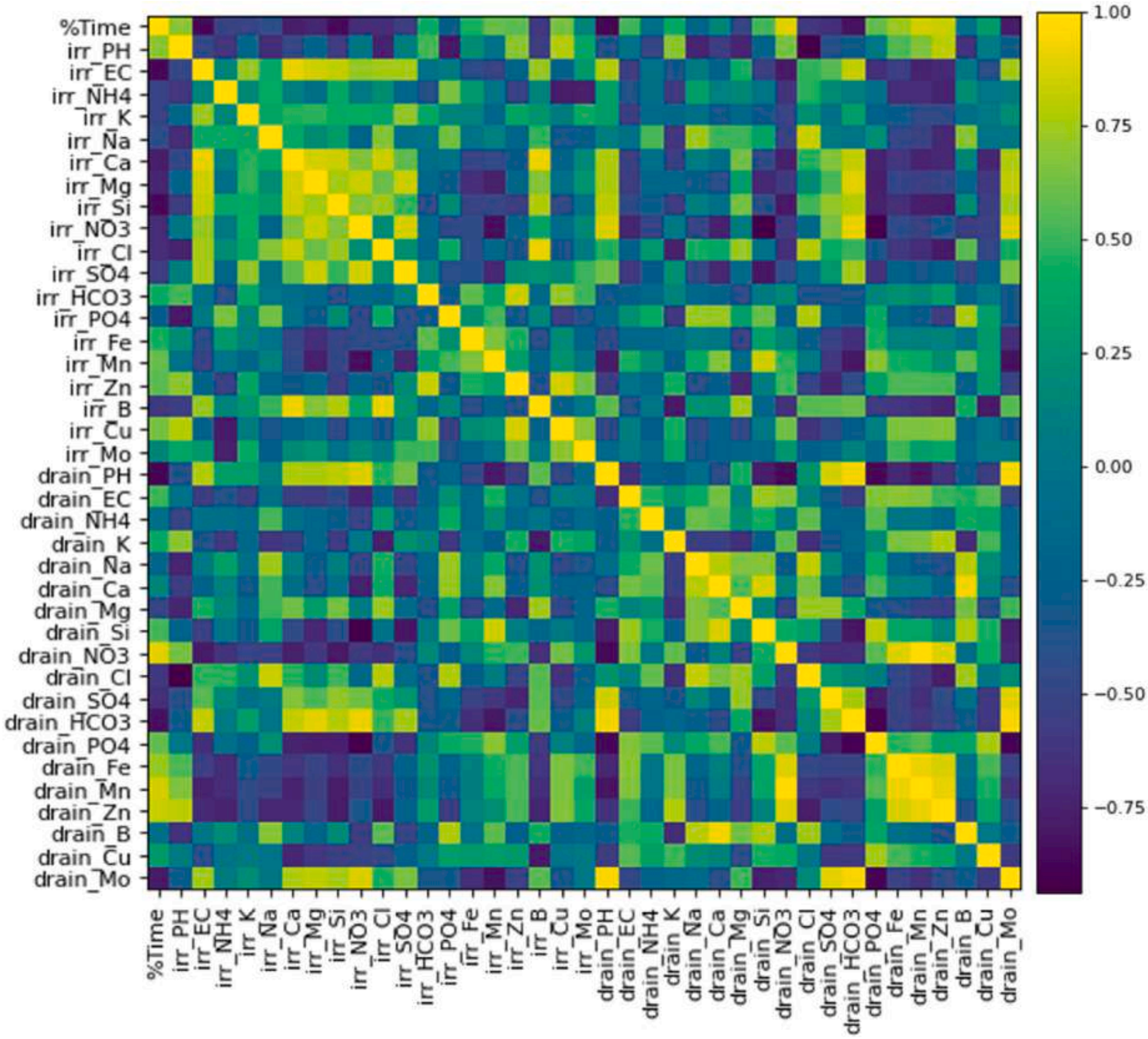


Fig. 9. Lab analysis data correlation matrix and density plot.

traction for their ability to handle complex, high-dimensional data in predictive modeling. SVR seeks to minimize prediction error by identifying the optimal hyperplane and narrowing the gap between predicted and observed values. In the equation provided, reducing the value of 'w' is equivalent to maximizing the margin.

- **Gaussian Process Regressor:** This model is a non-parametric statistical method designed to model intricate data distributions, particularly when dealing with noisy or incomplete data. GPR operates by forming a prior distribution over potential functions fitting the data and then updating this as new observations emerge. Within GPR, this distribution is illustrated through a Gaussian process, characterized by a set of jointly Gaussian distributed random variables [27]. The covariance among data points indicates their similarity and guides the prediction process. Important parameters like length scale and amplitude dictate the characteristics of these functions [28]. Once the data is observed, GPR produces a posterior distribution, which then facilitates predictions and uncertainty evaluations for unfamiliar data points. GPR's strengths lie in its capability to handle nonlinear relationships, its adaptive nature concerning covariance functions, and its provision for uncertainty predictions.

5.2.4. Model deployment

Upon narrowing down to three optimal models using the outcomes from the lazy regressor, the next step involved a meticulous manual setup for each of these models. The authors took charge of every aspect of training and evaluation to gain a preliminary insight into the model's performance, ensuring a more informed decision for the final model selection. Once these three models were fully operational, a comprehensive evaluation was conducted. This assessment utilized key performance metrics such as MAPE, RMSE, TSSE, and EF, details of which are elaborated upon as follows.

- **MAPE (Mean Absolute Percentage Error):** It measures the prediction accuracy in forecasting methods. It calculates the average percentage error between the actual and the predicted values. Therefore, a lower MAPE value indicates better fit of the data by the model.

$$\text{MAPE} = \frac{100\%}{\text{n}} \sum \frac{|\text{Actural} - \text{Forecast}|}{|\text{Actural}|}$$

- **RMSE (Root Mean Square Error):** It is a frequently used measure of the differences between values predicted by a model and the values

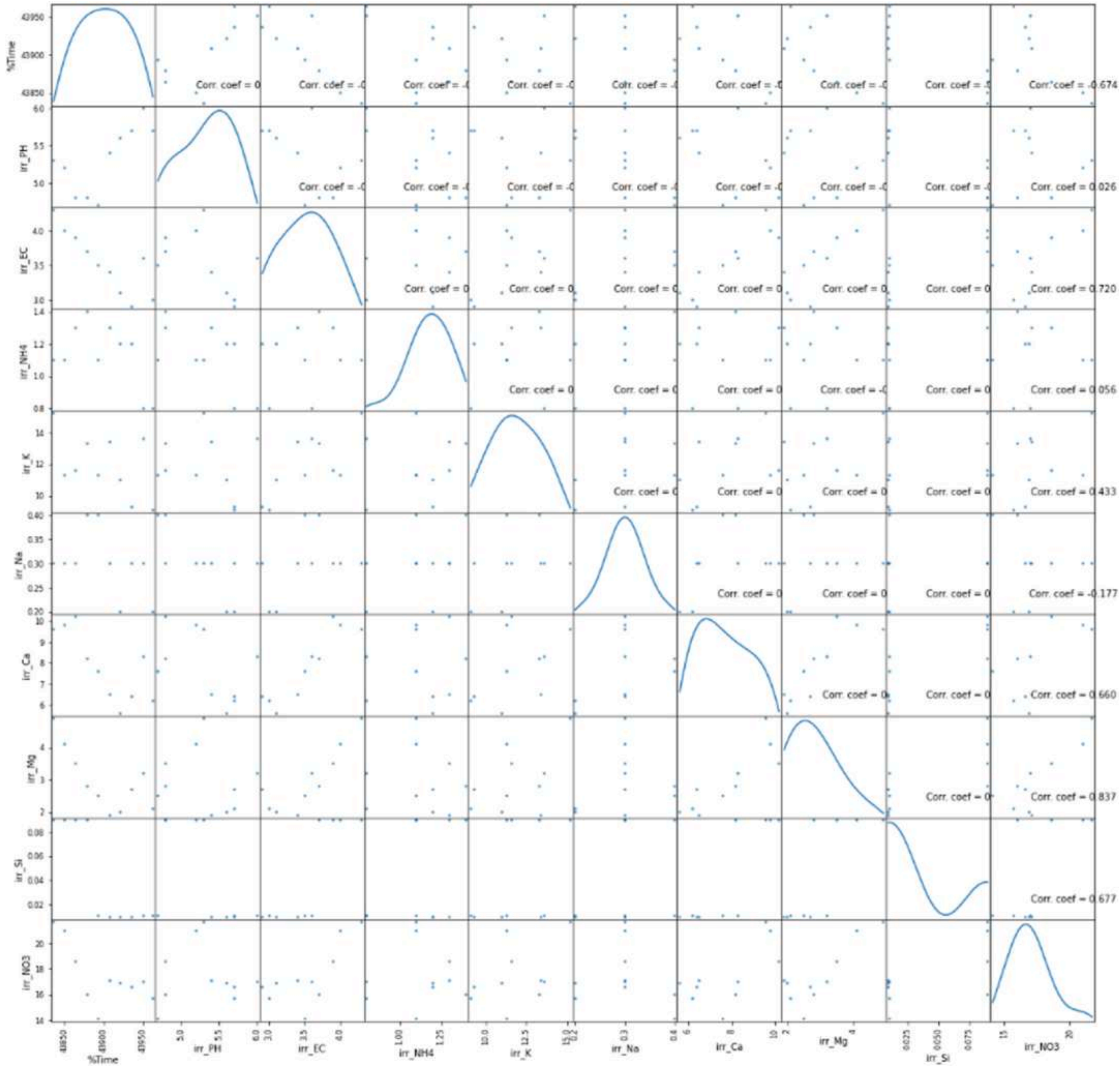


Fig. 9. (*continued*).

observed. Therefore, a lower RMSE value indicates a better fit of the data.

$$RMSE = \sqrt{\frac{\mathbf{1}}{\mathbf{n}}\sum (Actual - Forecast)^{\mathbf{2}}}$$

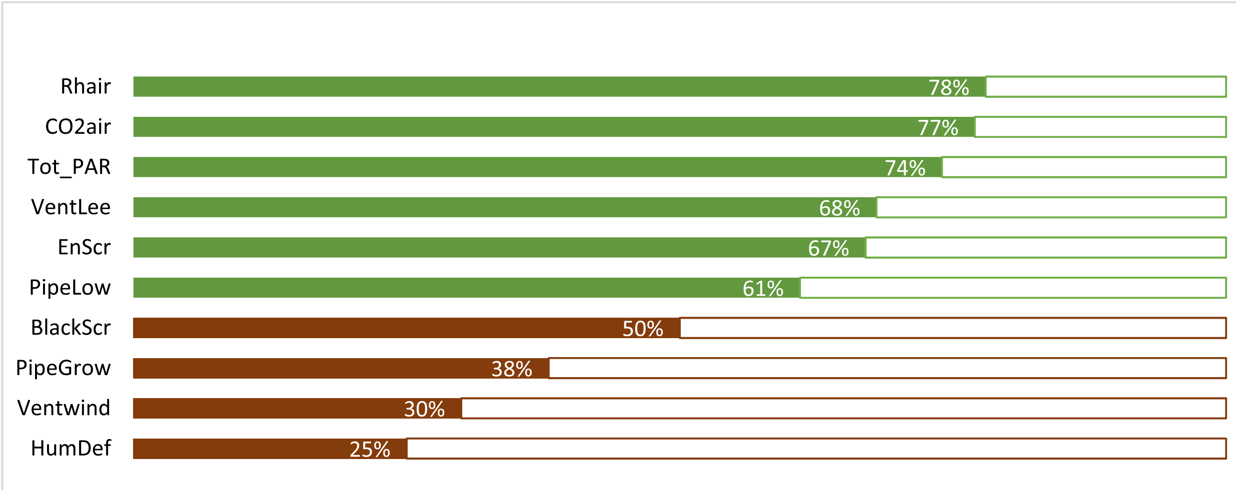
- **TSSE (Total Sum of Squared Error):** It measures the total discrepancy between the predicted and actual values, squared. It is the basis for many statistical tests and measures in regression. Therefore, a lower TSSE indicates a model that better fits the data.

$$TSSE = \sum (Actual - Forecast)^{\mathbf{2}}$$

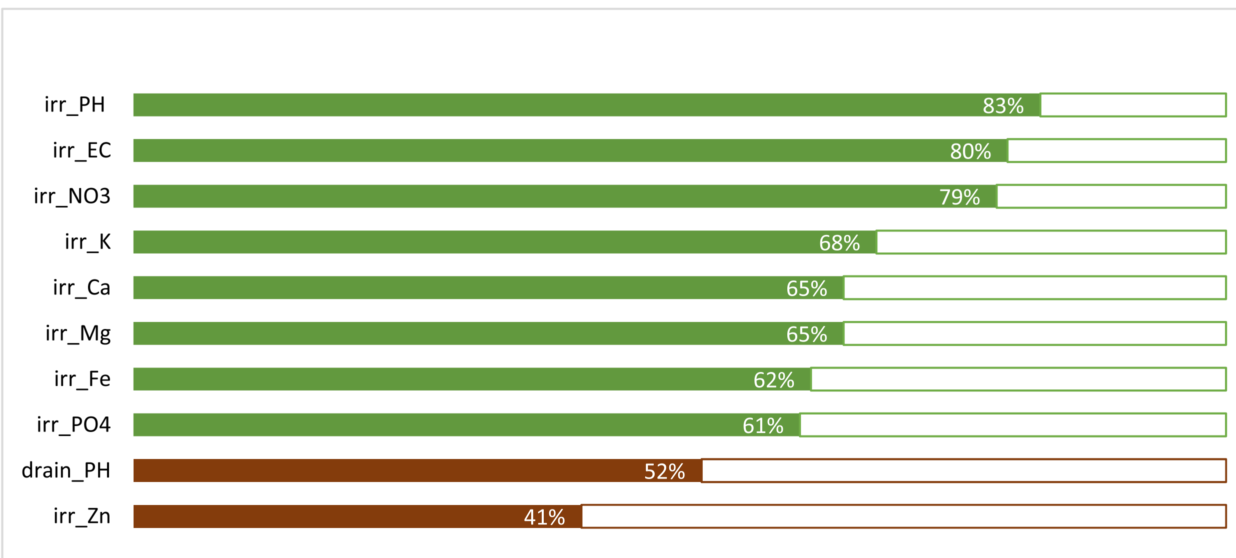
- **EF (Efficiency Factor):** It is not as standard as the other metrics, and its definition might vary across disciplines. Generally, in the context of hydrological modeling, it is also known as the Nash-Sutcliffe efficiency coefficient. It determines the relative magnitude of the residual variance compared to the measured data variance. An EF of 1 indicates perfect predictions, an EF of 0 indicates that the model predictions are as accurate as the mean of the observed data, and an EF less than 0 indicates that the mean of the observed data is a better predictor than the model.

$$EF = \mathbf{1} - \frac{\sum (Actual - Forecast)^{\mathbf{2}}}{\sum (Actual - Forecasted\ Mean\ )^{\mathbf{2}}}$$

Results, showed the better performance of RBF model with 0.80 RMSE followed by with 1 and GPR with 1.3 in Table 3. From this result, it can be undertood that firstly, the order of the most optimum models are still the same in terms of the most important metric (RMSE). Secondly, RBF and GPR showed better results than lazy regressor, while



a



b

Fig. 10. Sensitivity analysis of input variables a. climate data, b. lab analysis data.

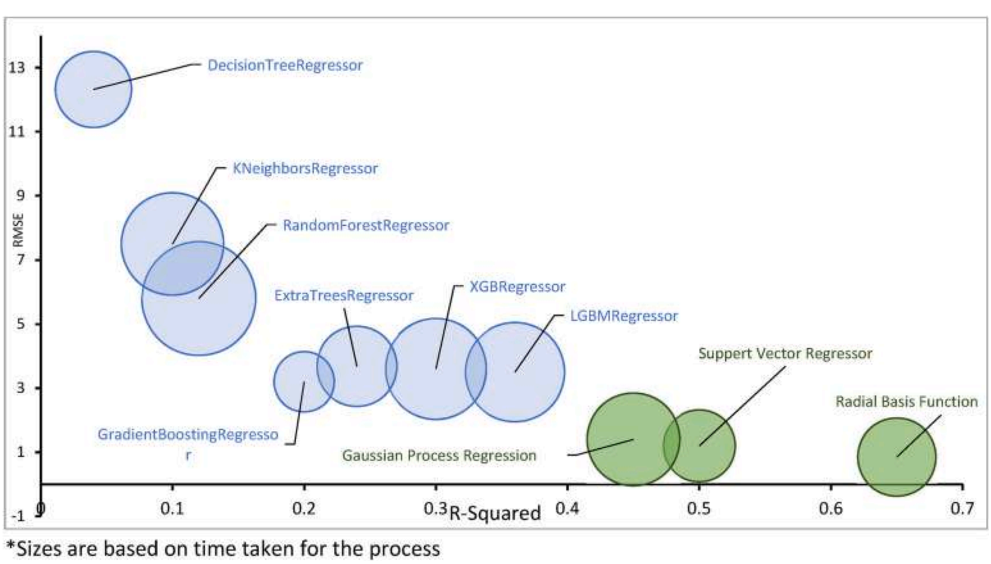


Fig. 11. Initial model comparison result based on lazy regressor.

Table 3
Selected model performance.

Model	Train				Test			
	RMSE	MAPE	TSSE	EF	RMSE	MAPE	TSSE	EF
RBF	0.80	1.21	350	0.98	0.83	1.30	230	0.97
SVM	1.00	2.37	2041	0.95	2.60	3.74	1473	0.92
GPR	1.30	1.80	920	0.97	4.20	4.39	1639	0.90

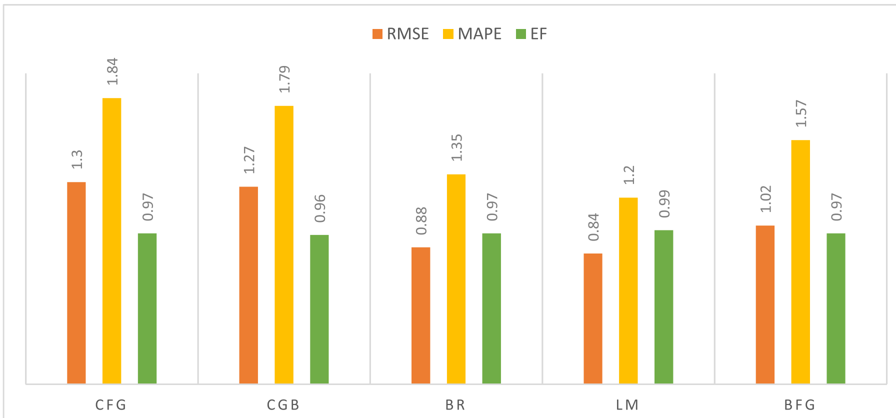


Fig. 12. Performance of RBF model with different training algorithms.

GPR showed even worse results than basic models comparison.

5.2.5. Final model tunning and evaluation

To finalize a single model as the preferred choice and further refine it, an in-depth comparison becomes imperative. The evaluation results, as presented in Table 2, highlight that the RBF model excels in terms of assessment metrics on the training dataset. However, the margin of superiority is not overwhelmingly vast, suggesting that with adjustments, other models might surpass the RBF. Viewing it from another angle, both the SVM and GPR models displayed suboptimal performances when subjected to the rain dataset. This suggests that while their performance on the training dataset is commendable, their efficacy diminishes considerably with test data. An overfitted model, such as these, can be deemed unreliable for predicting data points not encompassed within the training dataset. Consequently, given its consistent performance across training and test datasets, only the RBF model will advance to subsequent stages of analysis and refinement. To optimize RBF model there are three areas that can have impacts on the performance of the model which are the focus of this part.

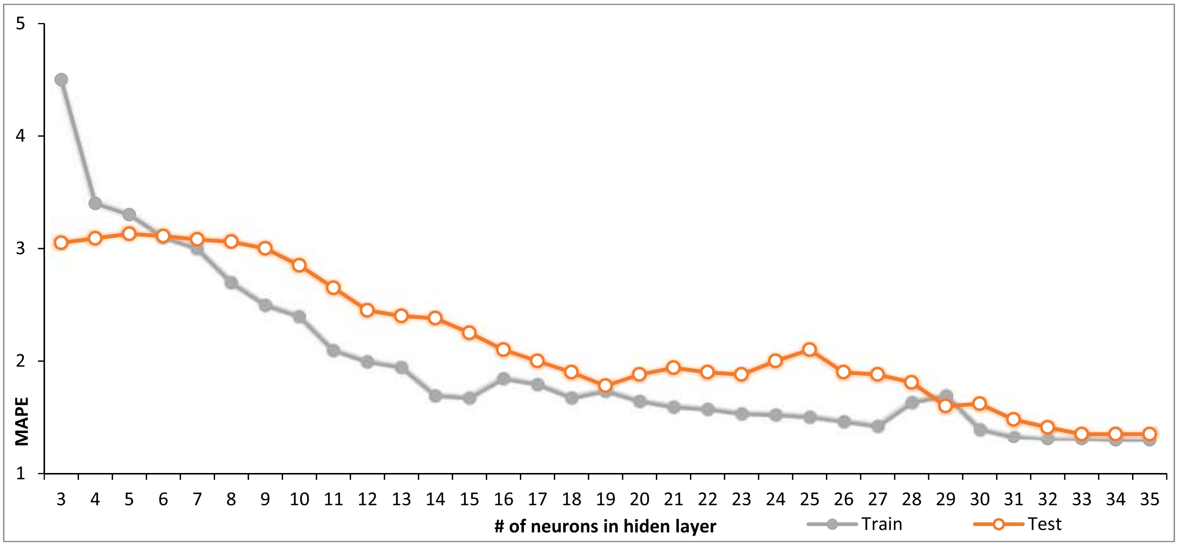
Firstly, choosing the optimal training algorithm for the RBF neural network model is crucial for its accuracy and performance. While various training algorithms exist, each comes with its own advantages and pitfalls in terms of convergence speed, computational demands, and potential for overfitting. The backpropagation algorithm, often employed for RBF neural networks, adjusts weights and biases iteratively but might converge slowly and risk landing in local minima. In contrast, the Levenberg-Marquardt algorithm adjusts its learning rate depending on the error surface's curvature, leading to quicker convergence. In this research, 5 top training algorithms for the RBF neural network based on the background studies were assessed (CFG, CGB, BR, LM, BFG) at Fig. 12.

As it is depicted in Fig. 12, the LM algorithm emerged as the most effective, outperforming others in accuracy metrics. Its faster convergence and reduced overfitting risks make it a favored choice for many. This algorithm is anticipated to enhance the RBF neural network's predictions for greenhouse indoor temperatures, especially with extensive datasets.

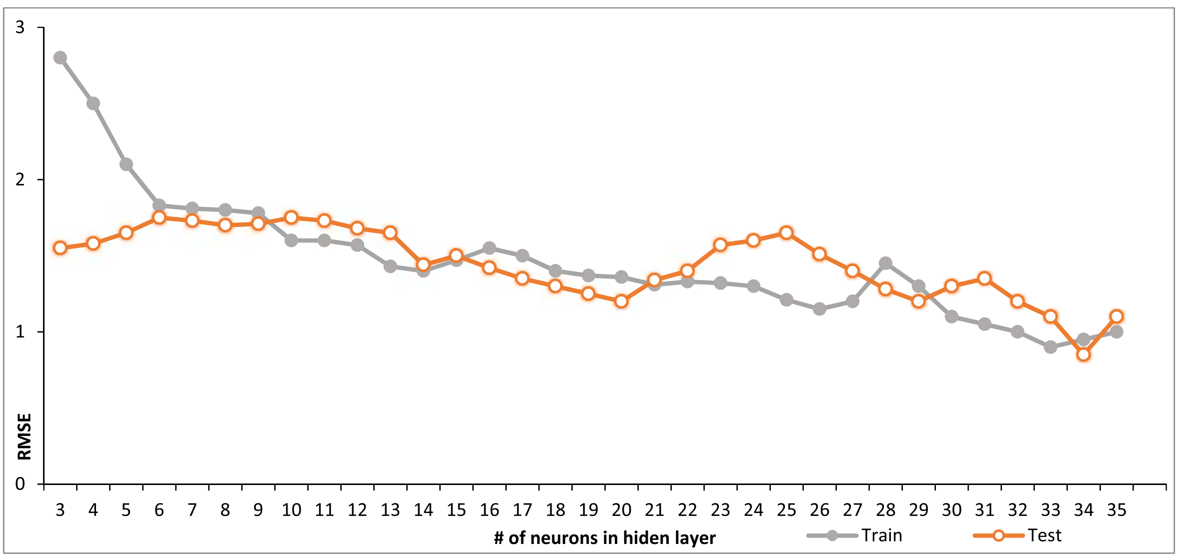
The next stage of optimization, hidden layer was subjected to optimization as it is crucial in converting input data into a space more amenable to linear analysis. It deals with nonlinear input patterns by mapping them to a higher dimension via the hidden layer. Cover's theorem suggests that the nonlinear patterns can be made linearly separable in this higher dimension. Therefore, having more neurons in the hidden layer than input neurons enhance the model's ability to understand nonlinear relationships. However, the optimal neuron count in the hidden layer varies based on data complexity. Too few might lead to under-fitting, causing the model to be oversimplified, while too many might cause overfitting, where the model over-learns from the training data and struggles with new data. Through a sensitivity analysis, where the RBF model is trained with varied neuron counts, one can determine the ideal number to avoid both overfitting and under-fitting. In this study, the impact of varying neuron counts in the hidden layer, from three to thirty five, on predicting the performance of the greenhouse in terms of its production was examined.

In the analysis depicted by Fig. 13, a critical observation was made regarding the configuration of the neural network. It was discerned that when thirty four neurons were utilized in the hidden layer, the model's performance reached an optimal level. This specific setup resulted in a MAPE of approximately 1.3% and an RMSE of roughly 0.9, which was consistent across both the training and testing datasets. Such findings underscore the significance of appropriately configuring the hidden layer, as it can greatly influence the accuracy and reliability of predictions made by the model.

Finally, the last stage of fine-tuning the RBF model is the spread factor optimization as an essential component in the RBF model whichholds significant influence over both its accuracy and overall efficiency. It specifically sets the width of the RBF kernel function. This width, in turn, plays a pivotal role in shaping the overlap of the RBF functions and how the input data is spatially represented. To pinpoint the most suitable spread factor for optimal model performance, a meticulous sensitivity analysis is necessary. This involves experimenting with different values of the spread factor and then methodically evaluating the results of the model performance against various statistical metrics in Fig. 14. It is worth noting that the ideal spread factor can vary based on the inherent characteristics of the input data and the intricacy of the relationships between input and output variables. For instance, in



a



b

Fig. 13. Performance of RBF model with different number of neurons a. MAPE, b. RMSE.

situations dealing with complex, nonlinear systems, a smaller spread factor might be preferable. Conversely, simpler systems might benefit from a larger spread factor. In the context of this study, the explored range for the spread factor was between 0.1 and 1.

Upon fine-tuning the spread factor to a value of 0.25 for the examined RBF model, a distinct improvement in the model's precision was observed. This advancement in accuracy was underscored by the MAPE and RMSE metrics, which yielded values of approximately 1.33 and 0.76, respectively. It is noteworthy that these metrics remained consistent across both training and test datasets, further reinforcing the effectiveness of the optimization. This suggests that the model's performance was considerably bolstered by the specific adjustments made to the spread factor.

Following the comprehensive process of adjustments and calibrations, the model was primed for its ultimate performance evaluation, specifically to analyze the discrepancy between the predicted and actual outcomes. The difference in performance of the final model, both pre and post-refinement, underscores the efficacy of the optimization efforts. This significant enhancement in the model's proficiency was achieved through systematic fine-tuning across three critical stages: the training algorithm, the count of neurons, and the spread factor, as illustrated in Fig. 15. A deeper dive into the results reveals that the residuals, which represent the variance between the actual and forecasted values, have notably diminished. In addition to the RMSE metric discussed in the preceding section, the heightened R-squared value, which rose from 0.92 to 0.98, serves as a testament to the model's augmented accuracy and the success of the refinement process.

6. Challenges

6.1. Limitations

The study's reliance on third-party data from the "Autonomous Greenhouse Challenge" presents a notable limitation due to the lack of control over data quality and scope. This constraint might affect the predictive models' performance and the AI-aided strategies' effectiveness, as they were designed based on the available, unmodifiable data. The variance in the effectiveness of these AI strategies across different studies underscores the need for further investigation into optimization and control strategies that are specifically tailored to individual greenhouse environments, ensuring more precise and applicable results.

Beyond the study's reliance on third-party data from the

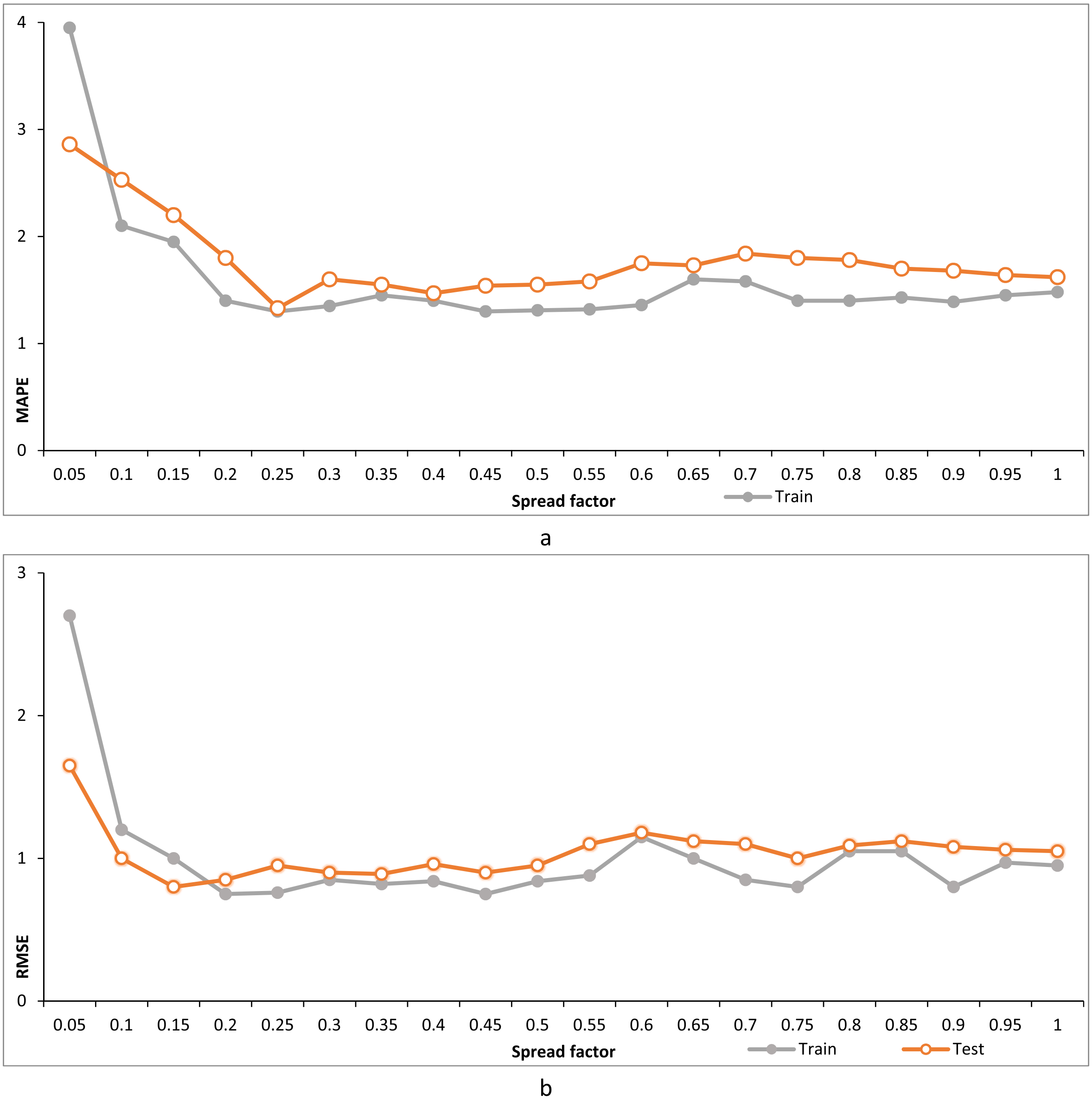


Fig. 14. Performance of RBF model with different values for spread factor a. MAPE, b. RMSE.

"Autonomous Greenhouse Challenge," several other constraints merit attention. The findings, while significant, may not be universally applicable due to the unique nature of each greenhouse environment, including variations in climate, crop types, and technological infrastructure. The complexity of the AI models used poses another challenge, as their deployment requires substantial computational resources and expertise, potentially limiting their use in less advanced agricultural settings. Despite efforts to streamline the dataset, the interpretability of the remaining variables and their influence on model predictions remains a concern, which is crucial for practical applications. Integrating AI technologies with existing greenhouse systems could be particularly challenging, especially in operations that are technologically outdated. The study also does not fully explore the broader environmental and ethical implications of AI integration in agriculture, such as the potential impact on labor markets and the ecological consequences of increased technological reliance. The long-term sustainability of AI-driven strategies, their effects on soil health, crop diversity, and the ecological balance within and around greenhouses, is another area that remains uncertain. Furthermore, the dynamic nature of agricultural environments, subject to changing weather patterns, pest populations, and crop varieties, may challenge the adaptability and resilience of AI systems without continuous updates and adaptations.

These considerations underscore the complexity of implementing AI in agriculture and highlight the need for comprehensive future research to ensure that technological advancements contribute positively to sustainable and resilient farming practices.

6.2. *Potential sources of errors and dealing with them*

Potential errors in this analysis may arise from inaccuracies in greenhouse data collection, sensor inaccuracies for environmental and crop monitoring, and biases from competition participants. The study also notes the need for improvements in managing CO2 emissions and water use, indicating possible limitations in the current AI optimization strategies for greenhouse operations.

Enhancing the foundational assumptions of this analysis could extend to a more thorough process of calibration and validation for the sensors gathering data, ensuring the precision and dependability of their measurements. This step is crucial for establishing a robust dataset that reflects true environmental conditions and crop growth metrics.

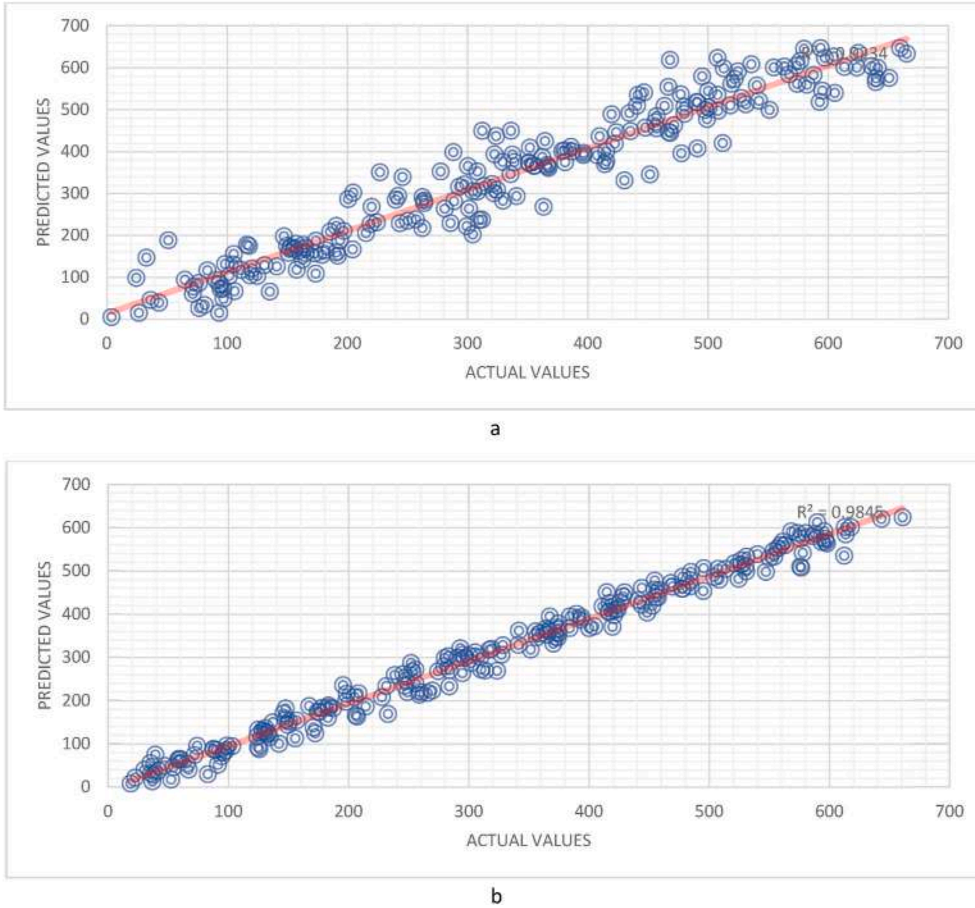


Fig. 15. Performance of RBF model a. Before fine-tuning, b. after fine-tuning.

Furthermore, broadening the scope of participation in the agricultural competition to include a diverse range of independent teams or researchers could significantly reduce potential biases. Such inclusivity would not only enrich the dataset with a variety of approaches and insights but also bolster the robustness and generalizability of the research findings, providing a more comprehensive understanding of the effectiveness of AI-driven strategies in different agricultural contexts. The potential errors and inaccuracies could skew the study's results and conclusions, possibly leading to an overestimation of AI's efficacy in enhancing greenhouse efficiency, particularly in energy reduction and crop yield improvement. Such discrepancies might impact decision-making by policymakers and scientists, especially in setting and achieving climate-related goals. Therefore, it's crucial for these stakeholders to critically assess the study's assumptions and limitations to ensure informed and accurate decisions based on the research outcomes.

The presence of errors could lead to overestimated benefits of AI in greenhouse enhancements. Addressing this requires critical evaluation of the data and methodologies used. Potential solutions include applying regularization techniques to the predictive model to prevent overfitting, normalizing input data to ensure consistent scale across variables, and removing outliers to mitigate their undue influence on the model's performance. These steps can help refine the accuracy of the findings, offering a more reliable basis for decision-making by policymakers and scientists in the context of climate change and agricultural policies.

7. Conclusion

In conclusion, this study has highlighted the transformative potential of artificial intelligence (AI) in refining agricultural practices, particularly within the controlled environments of greenhouses. Our exploration was anchored in a robust dataset derived from an agricultural competition that leveraged AI to optimize greenhouse operations. The empirical evidence presented herein underscores the efficacy of AI in reducing energy consumption, notably heating, thereby contributing to more sustainable agricultural practices without sacrificing crop yield, quality, or financial gain. Nonetheless, it is evident that the integration of AI in managing other critical aspects such as CO2 emissions and water usage requires further advancement. This research provides a foundational understanding of AI's benefits in greenhouse farming and paves the way for future innovations to address these remaining challenges.

This study extends beyond the operational optimization of greenhouses, venturing into the realm of predictive analytics. It examines the capability of artificial intelligence (AI) to forecast the outcomes of its integration within greenhouse operations. A critical assessment of various machine learning (ML) models culminated in the identification of the Radial Basis Function (RBF) model as particularly efficacious following meticulous optimization. The model achieved a notable Root Mean Square Error (RMSE) of 0.8 and an R-squared value of 0.98, demonstrating a high level of accuracy in predicting greenhouse production quantified in kg/m2. This breakthrough underscores the remarkable potential of AI in not only facilitating day-to-day greenhouse management but also in forecasting production outcomes with a high degree of precision.

The study primarily focuses on the quantitative aspects of production, showcasing the effectiveness of machine learning (ML) in forecasting outcomes in AI-enhanced greenhouse operations. The findings highlight the critical role of input variable optimization in maximizing output, which bears significant implications for engineering design by emphasizing the importance of integrating precise control systems and sensors in greenhouse infrastructure. Additionally, the study's emphasis on output optimization aligns with regulatory standards and energy policies aimed at promoting sustainable agricultural practices, potentially informing policy development and regulatory frameworks to support eco-efficient farming. From an energy systems perspective, the insights gained could guide the design and implementation of energy-efficient solutions in greenhouse management, contributing to broader energy conservation efforts. Financially, the optimization of inputs versus outputs underlines the economic benefits of adopting AI in agriculture, which could influence investment decisions and financial planning in the agricultural sector. Lastly, the study's focus on efficient production through AI integration touches upon Environmental, Social, and Governance (ESG) criteria by potentially reducing environmental impacts and supporting sustainable agricultural practices, thus contributing to the broader ESG goals of minimizing ecological footprints and fostering responsible resource management.

Future studies should focus on conducting comparative analyses across various case studies to validate and broaden the applicability of the findings. Exploring AI's effectiveness in different environmental and operational settings can deepen understanding of its adaptability. There's also potential in combining AI with technologies like IoT and edge computing for improved control in greenhouse farming. Furthermore, merging AI with environmental science could advance sustainable agricultural practices.

Funding

This work belongs to the "Water-Energy-Food-Environment Nexus" project financed by the "SAPIEXCELLENCE 2022 - BE-FOR-ERC" Fund.

CRediT author statement

Siamak Hoseinzadeh: Term, Data Curation, Conceptualization, Formal Analysis, Methodology, Software, Investigation, Validation, Visualization, Writing-Original draft, Writing - review & editing, Resources, Project administration, Supervision, Supervision, Funding acquisition.

Davide Astiaso Garcia: Visualization, Methodology, Writing - review & editing, Resources, Project administration, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Jie H, Khan I, Alharthi M, Zafar MW, Saeed A. Sustainable energy policy, socio-economic development, and ecological footprint: the economic significance of natural resources, population growth, and industrial development. Util Pol 2023;81:101490.
[2] Mok WK, Tan YX, Chen WN. Technology innovations for food security in Singapore: A case study of future food systems for an increasingly natural resource-scarce world. Trends in food science & technology 2020;102:155–68.
[3] Heikkinen AM. Climate change, power, and vulnerabilities in the Peruvian Highlands. 2021/08/21 Reg Environ Change 2021;21(3):82.
[4] Raihan A, Tuspekova A. The nexus between economic growth, renewable energy use, agricultural land expansion, and carbon emissions: new insights from Peru. 2022/06/16/ Energy Nexus 2022;6:100067.
[5] Ruan J, et al. A life cycle framework of green IoT-based agriculture and its finance, operation, and management issues. IEEE Commun Mag 2019;57(3):90–6.
[6] Türkten H, Ceyhan V. Environmental efficiency in greenhouse tomato production using soilless farming technology. 2023/04/20 J Clean Prod 2023;398:136482.
[7] Ramirez Camargo L, et al. Pathway to a land-neutral expansion of Brazilian renewable fuel production. 2022/06/07 Nat Commun 2022;13(1):3157.
[8] Panos E, Kannan R, Hirschberg S, Kober T. An assessment of energy system transformation pathways to achieve net-zero carbon dioxide emissions in Switzerland. 2023/05/08 Communications Earth & Environment 2023;4(1):157.
[9] Ghoulem M, El Moueddeb K, Nehdi E, Boukhanouf R, Calautit JK. Greenhouse design and cooling technologies for sustainable food cultivation in hot climates: review of current practice and future status. Biosyst Eng 2019;183:121–50.
[10] Benyezza H, Bouhedda M, Rebouh S. Zoning irrigation smart system based on fuzzy control technology and IoT for water and energy saving. J Clean Prod 2021;302:127001.
[11] Paris B, et al. Energy use in greenhouses in the EU: a review recommending energy efficiency measures and renewable energy sources adoption. Appl Sci 2022;12(10):5150.
[12] Usman M, Makhdum MSA. What abates ecological footprint in BRICS-T region? Exploring the influence of renewable energy, non-renewable energy, agriculture, forest area and financial development. Renew Energy 2021;179:12–28.
[13] Javaid M, Haleem A, Khan IH, Suman R. Understanding the potential applications of artificial intelligence in agriculture sector. Advanced Agrochem 2023;2(1):15–30.
[14] Ahmad T, Madonski R, Zhang D, Huang C, Mujeeb A. Data-driven probabilistic machine learning in sustainable smart energy/smart energy systems: key developments, challenges, and future research opportunities in the context of smart grid paradigm. Renew Sustain Energy Rev 2022;160:112128.
[15] Chen W-H, You F. Semiclosed greenhouse climate control under uncertainty via machine learning and data-driven robust model predictive control. IEEE Trans Control Syst Technol 2021;30(3):1186–97.
[16] Ben Ayed R, Hanana M. Artificial intelligence to improve the food and agriculture sector. J Food Qual 2021;2021:1–7.
[17] Bahn RA, Yehya AAK, Zurayk R. Digitalization for sustainable agri-food systems: potential, status, and risks for the MENA region. Sustainability 2021;13(6):3223.
[18] Wijerathna-Yapa A, Pathirana R. Sustainable agro-food systems for addressing climate change and food security. Agriculture 2022;12(10):1554.
[19] Azizpanah A, Fathi R, Taki M. Eco-energy and environmental evaluation of cantaloupe production by life cycle assessment method. 2023/01/01 Environ Sci Pollut Control Ser 2023;30(1):1854–70.
[20] Taki M, Ajabshirchi Y, Ranjbar SF, Rohani A, Matloobi M. Modeling and experimental validation of heat transfer and energy consumption in an innovative greenhouse structure. 2016/09/01/ Information Processing in Agriculture 2016;3(3):157–74.
[21] Parajuli S, Narayan Bhattarai T, Gorjian S, Vithanage M, Raj Paudel S. Assessment of potential renewable energy alternatives for a typical greenhouse aquaponics in Himalayan Region of Nepal. 2023/08/15/ Appl Energy 2023;344:121270.
[22] S. F. Stefenon, L. O. Seman, V. C. Mariani, and L. D. Coelho, "Aggregating prophet and seasonal trend decomposition for time series forecasting of Italian electricity spot prices," Energies, vol. 16, no. 3. doi: 10.3390/en16031371.
[23] Prema V, Rao KU. Development of statistical time series models for solar power prediction. 2015/11/01/ Renew Energy 2015;83:100–9.
[24] Panigrahi S, Behera HS. A hybrid ETS–ANN model for time series forecasting. 2017/11/01/ Eng Appl Artif Intell 2017;66:49–59.
[25] S. Carta, A. Medda, A. Pili, D. Reforgiato Recupero, and R. Saia, "Forecasting E-commerce products prices by combining an autoregressive integrated moving average (ARIMA) model and google trends data," Future Internet, vol. 11, no. 1. doi: 10.3390/fi11010005.
[26] Pu L, Li Y, Gao P, Zhang H, Hu J. A photosynthetic rate prediction model using improved RBF neural network. 2022/06/10 Sci Rep 2022;12(1):9563.
[27] S. Wang, J. Gong, H. Gao, W. Liu, and Z. Feng, "Gaussian process regression and cooperation search algorithm for forecasting nonstationary runoff time series," Water, vol. 15, no. 11. doi: 10.3390/w15112111.
[28] S. S. Ghosh, S. Dey, N. Bhogapurapu, S. Homayouni, A. Bhattacharya, and H. McNairn, "Gaussian process regression model for crop biophysical parameter retrieval from multi-polarized C-band SAR data," Rem Sens, vol. 14, no. 4. doi: 10.3390/rs14040934.