

Article

# Investigating Precise Decision-Making in Greenhouse Environments Based on Intelligent Optimization Algorithms

Zhenyi Zhu <sup>1,2</sup>, Chunguang Bi <sup>1</sup> and You Tang <sup>2,\*</sup>

<sup>1</sup> College of Information Technology, Jilin Agricultural University, Changchun 130118, China; zhuzhenyi888@jlau.edu.cn (Z.Z.); chunguangb@jlau.edu.cn (C.B.)

<sup>2</sup> Electrical and Information Engineering College, Jilin Agricultural Science and Technology University, Jilin 132101, China

\* Correspondence: tangyou@neau.edu.cn

**Abstract:** The precise control of a greenhouse environment is vital in production. Currently, environmental control in traditional greenhouse production relies on experience, making it challenging to accurately control it, leading to environmental stress, resource waste, and pollution. Hence, this paper proposes a decision-making greenhouse environment control strategy that employs an existing monitoring system and intelligent algorithms to enhance greenhouse productivity and reduce costs. Specifically, a model library is created based on machine learning algorithms, and an intelligent optimization algorithm is designed based on the Non-Dominated Sorting Genetic Algorithm III (NSGA-3) and an expert experience knowledge base. Then, optimal environmental decision-making solutions under different greenhouse environments are obtained by adjusting the greenhouse environmental parameters. Our method's effectiveness is verified through a simulated fertilization plan that was simulated for a real greenhouse tomato environment. The proposed optimization solution can reduce labor and time costs, enable accurate decision-making in the greenhouse environment, and enhance agricultural production efficiency.

**Keywords:** greenhouse environment; precise decision-making; machine learning; model library; tomato



**Citation:** Zhu, Z.; Bi, C.; Tang, Y. Investigating Precise Decision-Making in Greenhouse Environments Based on Intelligent Optimization Algorithms. *Processes* **2024**, *12*, 977. <https://doi.org/10.3390/pr12050977>

Academic Editor: Yinlun Huang

Received: 10 March 2024

Revised: 5 May 2024

Accepted: 8 May 2024

Published: 10 May 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The continuous development of sensors for greenhouse agriculture allows for the collection of valuable information, which, after analysis, can be employed to adjust the relevant greenhouse parameters, thereby increasing crop yields, optimizing greenhouse production processes, and reducing fertilizer waste in agricultural production. However, in addition to the continuous efforts to improve the terminal equipment, greenhouse-related intelligent algorithms must be developed to utilize the acquired data effectively. This has become feasible with the gradual improvement in machine learning and optimization algorithms in real-world applications. Precisely, combining different intelligent algorithms to construct decision-making and optimization models for smart greenhouse environments is possible.

This study develops an intelligent environmental decision-making optimization model based on an empirical knowledge base, aiming to overcome the problems caused by the difficulty in accurately controlling the crop growth environment in greenhouse agriculture. The designed model combines empirical knowledge with advanced machine learning and intelligent optimization algorithms and exploits technical guidance from different greenhouse experts, thereby providing an intelligent environmental decision-making optimization model. The proposed intelligent method helps improve greenhouse environmental decision-making accuracy, maximize agricultural production efficiency, and reduce the risk of resource waste and environmental pollution.

Traditional agricultural production management problems are solved using special equipment to collect data, e.g., combining remote sensing information with intelligent algorithms to regulate wheat planting and enhance productivity and management efficiency [1]. This study employed a fixed-wing UAV-based N-regulation algorithm to guide the in-season variable N topdressing at a farm scale. Although this method is novel, the UAV is limited by environmental conditions, such as strong winds, which may exceed its capabilities. Highly technical greenhouse agricultural production facilities use Internet of Things (IoT) devices to collect information to regulate crop planting and the production environment. For instance, combining big data analysis based on IoT devices with machine learning algorithms can improve the energy efficiency of pepper farms [2]. Specifically, the proposed method predicts the peak temperature energy of a farm based on a random forest (RF). Nevertheless, their work focuses on predicting the temperature rather than developing a complete control system. Additionally, external environmental factors and machine learning models have been used to predict temperature changes in a greenhouse [3]. Still, this study is limited in predicting indoor greenhouse temperature. Moreover, visual acquisition equipment can detect diseases in a greenhouse [4]. Although significant, automatically detecting diseases in a greenhouse is only a part of the crop monitoring process. Moreover, in [5], the authors develop a tomato maturity recognition model based on an improved You Only Look Once (YOLO) v5s-tomato-based algorithm. Nevertheless, their technique does not monitor any of the growth stages of the tomato. Thus, although this work is interesting, it focuses on the last stage of tomato production. Moreover, utilizing robotic arms can realize intelligent picking of fruits. Hence, collecting and analyzing greenhouse environmental information, combined with intelligent algorithm optimization, can provide precise environmental decision-making solutions to enhance efficiency and accuracy, promoting the further development of greenhouse agriculture.

Current research mainly employs IoT devices to predict the trend in greenhouse environmental indicators or combines different data types and algorithms to predict the elements required for plant growth, such as changes in leaf phosphorus concentration [6,7]. This work effectively estimates the P-concentration in rice leaves but is limited to the tillering stage of potted rice. Moreover, in [8], the authors identify the optimal environmental parameters for crops under specific conditions during greenhouse production. Specifically, they combine genetic algorithms with IoT devices to maximize crop yields while minimizing energy consumption. Although interesting, this study focuses on optimizing the setpoints of the proportional–integral–derivative (PID) controllers for energy efficiency rather than presenting a complete monitoring system.

Spurred by the current trends presented above, the objectives of this study are the following: (1) construct a precise decision-making model for the greenhouse environment, (2) optimize and analyze greenhouse environmental data by using machine learning technology and algorithms, (3) predict the environmental control plan for greenhouses and the growth status of crops, and (4) provide decision-making plans and control strategies to improve energy utilization efficiency in the greenhouse and the growth efficiency of crops. Moreover, the greenhouse environment decision-making model library and the crop precision decision-making optimization model constructed in this study rely on different data. The model library is built based on different characteristics, such as lighting, temperature, humidity, etc., and the knowledge base is established based on historical experience and expert knowledge.

The remainder of this paper is organized as follows. Section 2 introduces the process to collect and analyze greenhouse environmental information. Section 3 presents the proposed greenhouse environment decision-making model library and the optimization model. Section 4 verifies and analyzes the developed optimization model on tomato crops as an example. Finally, Section 5 summarizes this work.

## 2. Collection and Analysis of Greenhouse Environmental Information

### 2.1. Overview of Greenhouse Information

The experimental greenhouse used in this study is the solar greenhouse of the scientific research base of the Jilin Institute of Agricultural Science and Technology (43°96' N, 126°49' E). The greenhouse has a span of 10 m, a ridge height of 5.5 m, and a length of 60 m. Its transparent covering material is polyolefin (PO) film, and a layer of quilt is used for the external insulation cover. The northern wall is constructed from red brick, with no heating equipment in winter. The experimental tomato variety is Jingcai No. 8 Strawberry Persimmon (Beijing, China), According to expert experience, greenhouse tomatoes mainly undergo three growth stages: seedling stage, flowering and fruiting stage, and fruit enlargement stage. This study collected sensor data and meteorological data from three stages for a total of four months.

### 2.2. Data Collection and Analysis

The data collected by the sensors mainly include soil salinity, photosynthetically active radiation, total radiation, leaf moisture, precipitation, greenhouse internal humidity, soil temperature, greenhouse temperature, wind direction, and wind speed. Meanwhile, day and night are included as an additional feature to consider the time factor more comprehensively. The data collection sensors used mainly include the soil comprehensive sensor model ZTS-3000-TR-\* (Shandong, China), which has a soil temperature range of  $-20\text{--}80\text{ }^{\circ}\text{C}$ , a resolution of  $0.01\text{ }^{\circ}\text{C}$ , and an accuracy of  $\pm 0.5\text{ }^{\circ}\text{C}$ . The soil moisture range is of  $0\text{--}100\%$ , with a resolution of  $0.01\%$  and an accuracy of  $\pm 2\%$ . The soil pH range is of  $0\text{--}14\text{ pH}$ , with a resolution of  $0.01\text{ pH}$  and an accuracy of  $\pm 0.3\text{ pH}$ . The soil N/P/K range is of  $0\text{--}1999\text{ mg/kg}$ , resolution  $0.1\text{ mg/kg}$ , accuracy  $\pm 2\%$ . The air temperature and humidity sensor EXTECH HT30 (Shanghai, China) has a humidity range of  $0\text{--}RH$ , temperature range of  $0\text{--}50\text{ }^{\circ}\text{C}$ , resolution  $0.1\% RH$ ,  $0.1\text{ }^{\circ}\text{C}$ . The photosynthetic effective radiation sensor LI-190R (Beijing, China) has a range of  $400\text{--}700\text{ nm}$ , an accuracy of  $\pm 5\%$ , and a resolution of  $0.1\%$ .

The data types are represented as floating point numbers, with an accuracy of two decimal places. Day and night information is identified as 0 and 1, respectively, and the growth stages are marked with integers. Table 1 presents some data samples. Overall, the data involve twelve features, which provide rich information and help develop comprehensive and accurate intelligent models for strategic plan prediction and optimization.

**Table 1.** Data samples.

Soil Moisture (%)	Soil Salinity ( $\mu\text{s/cm}$ )	Photosynthetically Active Radiation ( $\mu\text{mol/m}^2/\text{s}$ )	Total Radiation ( $\text{J/m}^2$ )	Greenhouse Temperature ( $^{\circ}\text{C}$ )	Leaf Moisture (%)	Precipitation (ml)	Relative Humidity (%)	Soil Temperature ( $^{\circ}\text{C}$ )	Temperature Outside the Greenhouse ( $^{\circ}\text{C}$ )	Wind Direction	Wind Speed (m/s)	Growth Stage	Day and Night
17.06	0.76	19	0.2	16.7	0	0	90.3	21.7	16.1	78.8	0.82	3	0
17.05	0.76	19	0.3	16.7	0	0	90.3	21.6	15.7	79	0.42	3	0
17.04	0.76	19	0.2	16.7	0	0	90.6	21.5	15.4	79	0.42	3	0
17.01	0.76	19	0.2	15.1	1	0	90.8	21.4	15.2	78.9	0.88	3	0
16.98	0.76	19	0.2	15.1	1	0	91.1	21.3	15	78.8	0.53	3	0
16.98	0.76	19	0.2	15.1	1	0	91.3	21.2	14.8	78.8	0.32	3	0

It is important to analyze whether the feature data have a practical impact on the model, enhancing its prediction accuracy. Thus, a feature permutation strategy randomly arranges the features and observes the changes in model performance. The advantage of this strategy is that the feature importance does not depend on the selected fitting model, demonstrating the relationship between the fitting target and the features.

The results in Table 2 highlight that four features affect the prediction results, and the most important feature is temperature, i.e., soil and greenhouse temperatures. Wind speed and total radiation are less important features, posing an inferior impact on the prediction results than temperature-related features. Therefore, for the following model setup, we discard the features of wind direction, precipitation, humidity inside the greenhouse, and temperature outside the greenhouse. The remaining eight features for the subsequent modeling include soil temperature, greenhouse temperature, photosynthetically active radiation, leaf moisture, soil salinity, wind speed, total radiation, and time (day or night).

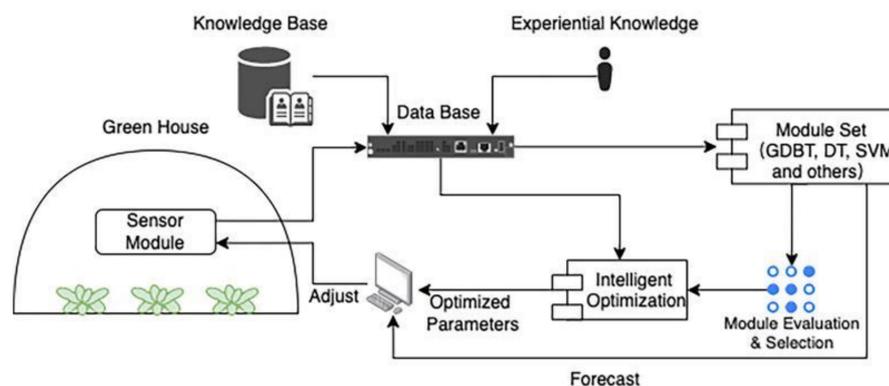
**Table 2.** Results on the feature importance ranking.

Soil Temperature	Greenhouse Temperature	Photosynthetically Active Radiation	Leaf Moisture	Soil Salinity	Wind Speed	Total Radiation	Wind Direction	Precipitation	Humidity inside the Greenhouse	Temperature Outside the Greenhouse
1.320	1.305	0.261	0.212	0.207	0.013	0.002	0	0	0	0

### 3. Construction and Optimization of Greenhouse Environmental Decision-Making Model

#### 3.1. Construction of Model Library Based on Machine Learning

Constructing an environmental prediction model for detecting crop growth stages can help optimize agricultural production. However, quantifying crop growth intuitively is challenging; therefore, this study uses a dataset on plant growth stages for modeling. The dataset is collected in the field and comprises plant growth stages annotated by experts. Although the proposed model is trained on that dataset, it can adapt to different crops and environmental conditions by configuring the parameters of the different crop growth stages, improving the model’s versatility. Figure 1 illustrates the model construction process.



**Figure 1.** Schematic diagram of model construction.

This study employs several machine learning algorithms to construct a model library, which is then used to establish the prediction model. Different models are trained for the same task to obtain better prediction results, and the one performing best on the test set is selected as the prediction model. Constructing a model library involves presetting

several machine learning algorithms as models trained on the actual greenhouse environmental data, setting several evaluation indicators for each model based on the accumulated collected data, and adaptively selecting the most suitable model as the learning model.

To satisfy the computing resource requirements of the model library, this study increases the update interval of the calculations when updating the model. In actual greenhouse environment production, computing resources are idle in most cases. Thus, several models are calculated and evaluated during the idle periods to exploit these computing resources, thereby improving their utilization. The model is constructed using Pandas, a popular data processing library in the Python language, and Scikit-learn, a machine learning library.

Several machine learning models are evaluated as model libraries through simulation and comparison on the dataset, including three support vector machine models (SVM [9], linear-SVM [10], and nu-SVM [11]), two tree models (gradient boosting tree model [12] and decision tree model [13]), and three linear models (Bayes Ridge regression model [14], logistic regression model [15], and gradient descent model [16]).

The main reason for choosing these models is that they represent the classic models used in the current machine learning research. First, for the SVM model, it can distinguish different types of data by finding the hyperplane between the data. This method is also commonly used in the modeling process of agricultural machine learning models. In addition, the model also chooses two types of tree models, because these tree models have a good effect when addressing problems with complex features. The main principle of the algorithm is to use the tree structure to establish the interval parameters of different features. We chose two decision trees representing the model based on this idea and the optimized version of GDBT. In order to further enrich the model, we also chose some more classic regression models, including Bayesian Ridge regression model based on Bayesian thought, which can handle unbalanced data well. As a basic model for statistical calculation, the logistic model is also included in the model base. In addition, the linear regression model based on gradient decline is also used, which mainly uses gradient decline to optimize model parameters. These different models are currently widely used in the field of agricultural applications, but there is a lack of algorithms to optimize and combine these different models in an effective way. Therefore, this article integrates these different models using a model library to establish a greenhouse agriculture parameter prediction model for our method.

In terms of model parameter settings, the two SVM models use a penalty term parameter of 1.0, and it is obvious that the two models use different kernel functions; one is a linear kernel and the other is a nu kernel. The use of two types of kernels allows the model to encompass the learning ability of two different types of kernels. In the parameter settings of the two tree models, the minimum parameter splitting feature number is set to 2, and the minimum leaf feature number is set to 1 to ensure that each feature can be well learned. The learning rate of GDBT is set to 0.1, and in other settings such as linear model parameters and model parameter settings not mentioned, it is ensured that it is consistent with the parameter settings of classical models.

The parallel processing mechanism of Python is exploited for multi-threaded calculations to fully utilize computing resources and complete the management of different models. Meanwhile, the model library is abstracted into a Python class to write the corresponding functions to provide the corresponding responses, such as prediction, evaluation, and data processing (in Figure 2). Specifically, the collected feature data are first processed using Pandas and sent to the model library when the subsequent training process starts. The model library divides the feature data into a test set and a training set at a ratio of 2:8. Once training is completed, the trained models are evaluated using seven evaluation indicators implemented in Python, i.e., maximum error, mean absolute error, mean square error, root mean square error, mean square logarithmic error, median absolute error, and coefficient of determination. These indicators evaluate the model's prediction effect or quality, and the smaller the indicator value, the better the prediction effect. Then, by taking

these indicators as evaluation criteria, different models are voted on, and the model with the most votes is selected as the crop growth prediction model. Meanwhile, the model with the lowest training error is selected as optimal. When the performance of several models is relatively balanced, the coefficient of determination is taken as the reference index.

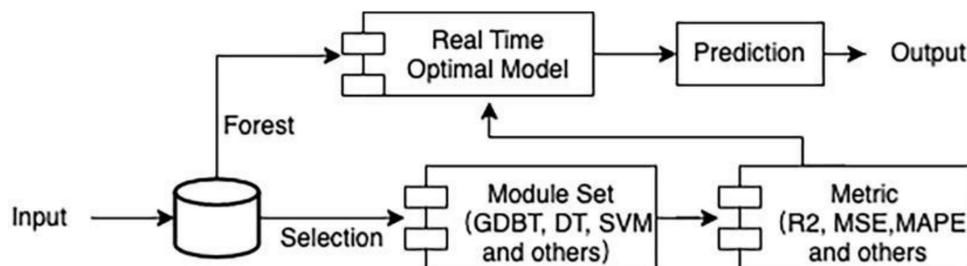


Figure 2. Model selection and prediction in the model library.

### 3.2. Construction of Greenhouse Decision-Making Optimization Model

The main method to investigate the intelligent decision-making process employs a crop growth prediction model and an intelligent optimization algorithm as the basic algorithm. It utilizes past experts’ experience and knowledge as the intelligent algorithm’s inspiration seeds. This strategy incorporates a precise decision-making optimization model based on the optimization algorithm for the greenhouse environment.

By continuously accumulating agricultural production knowledge, a suitable range of environmental parameters for different crop growth stages is obtained based on the growth and planting of crops. These parameters are summarized in the library presented in Table 3.

Table 3. Example of the empirical knowledge base.

Growth Stages	Daytime Temperature (°C)	Nighttime Temperature (°C)	Air Humidity (%)	Soil Moisture (%)	N/(mg·L <sup>-1</sup> )	P/(mg·L <sup>-1</sup> )	K/(mg·L <sup>-1</sup> )
Seedling growth stage	25–30	12–16	50–65	60–85	110–130	100–150	120–150
Flowering and fruit setting stage	25–28	12–18	60–65	70–80	120–130	70–90	140–230
Fruit expansion stage	28–30	16–20	70–75	80–90	120–180	60–80	170–220

Six parameters, including temperature, soil moisture, air humidity, nitrogen, phosphorus, and potassium, are divided based on some upper and lower limits, leading to 12 parameters in total. Meanwhile, a day and night mechanism is added. Then, the model library is used for prediction learning on the compiled parameters and the collected greenhouse environment information (Table 4).

Table 4. Example of a compiled knowledge base.

Growth Stage	Temperature Upper Limit (°C)	Temperature Lower Limit (°C)	Day and Night	Soil Moisture Upper Limit (%)	Soil Moisture Lower Limit (%)	Air Humidity Upper Limit (%)	Air Humidity Lower Limit (%)	Upper Limit N (mg·L <sup>-1</sup> )	N lower limit (mg·L <sup>-1</sup> )	P Upper Limit (mg·L <sup>-1</sup> )	P Lower Limit (mg·L <sup>-1</sup> )	K Upper Limit (mg·L <sup>-1</sup> )	K Lower Limit (mg·L <sup>-1</sup> )
0	30	25	0	65	50	85	60	130	110	150	100	150	120
1	28	25	0	65	60	80	70	130	120	90	70	230	140
2	30	28	0	75	70	90	80	180	120	80	60	220	170
0	16	12	1	65	50	85	60	130	110	150	100	150	120
1	18	12	1	65	60	80	70	130	120	90	70	230	140
2	20	16	1	75	70	90	80	180	120	80	60	220	170

Based on the same evaluation method, different machine learning algorithms are evaluated and voted on seven evaluation indicators to determine the optimal model that can effectively learn the appropriate indicators corresponding to environmental information, such as temperature, and realize accurate prediction and adjustment of the growth environment.

For the optimization model, the decision function is based on the prediction model of the six indicators obtained after training using the model library. The upper and lower limits of the various indicators in the knowledge base are used as the upper and lower limits of the decision variables. The main goal of optimization is to control the greenhouse environment based on reasonable decision-making solutions. Using nitrogen, phosphorus, and potassium as examples, the optimization goal is to minimize the soil's nitrogen, phosphorus, and potassium contents. For this approach, a multi-objective optimization problem with a minimization objective of constrained continuous decision variables is employed, formulated as follows:

$$\begin{aligned} \min \{z_1 = f(x), z_2 = f(x), \dots, z_r = f(x)\} \\ \text{s.t. } g_i(x) \leq 0, i = 1, 2, \dots, m, \end{aligned}$$

where  $f_k(x), k = 1, 2, \dots, r$  represents the prediction function of different indicators obtained by learning the machine learning model, used as the objective function of the optimization problem.  $Z_k$  is the predicted indicator, the value of  $r$  is determined by the number of indicators in the knowledge base used,  $g_i(x)$  denotes the upper and lower limits after transformation from the indicators in the knowledge base, and  $m$  represents the number of corresponding constraints in the knowledge base. Then, the optimization model is used to optimize and solve this problem.

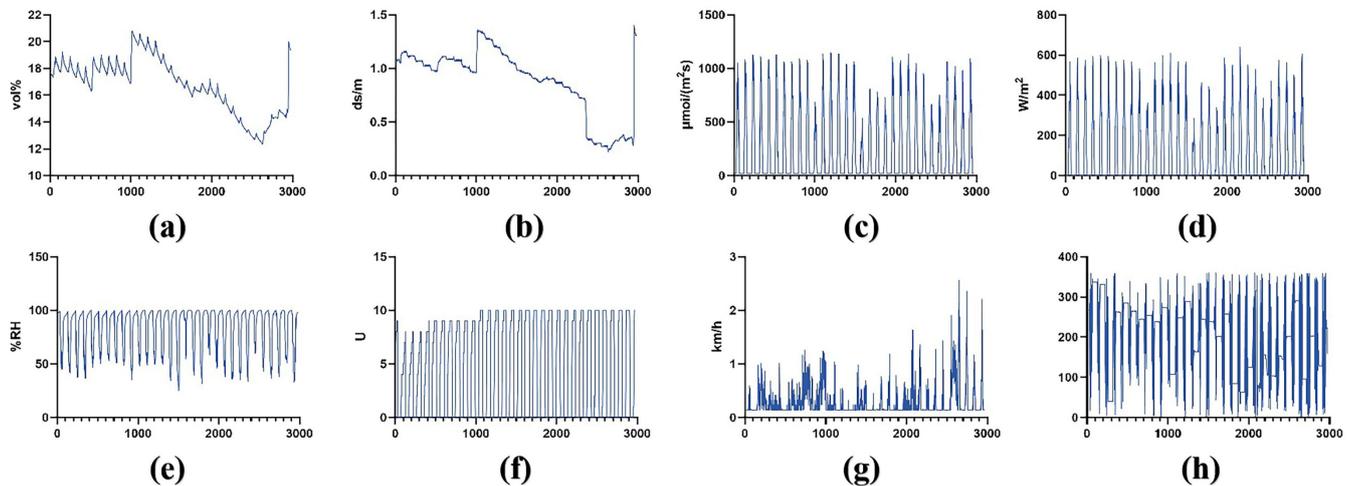
This article establishes a multi-objective optimization problem to minimize greenhouse input, which includes several indicators, shown in Table 3, except for the growth stage. The changes in these indicators in the knowledge base determine the main cost of plant growth, and the changes in these indicators require different cost inputs, such as increasing the heat source in the greenhouse and increasing fertilizer application, which will cause cost increases. At the same time, due to the mutual influence of greenhouse factors, such as humidity affecting fertility, different factors need to be comprehensively considered. Based on the knowledge base for learning modeling, the first step is to enable the prediction model to provide the optimal value of a single indicator in the current greenhouse environment. At this time, the correlation between different indicators is not considered. Using the optimization model based on the single indicator prediction model can obtain its global optimum, that is, considering the optimal solution of different greenhouse indicators. Obviously, using the optimization model can further help decision-making in the greenhouse environment rather than local adjustment of a certain indicator. Therefore, the target variable for modeling this optimization model is the optimal indicator summarized from prior experience in the knowledge base, which means that the optimal corresponding indicator under the current parameters can be obtained.

This study compares four optimization models from the intelligent optimization field, namely the Non-Dominated Sorting Genetic Algorithm II (NSGA-2) [17], NSGA-3 [18], push and pull search Multi-Objective Evolutionary Optimization Decomposition (PPS MOEAD DE) [19], and Reference Vector-Guided Evolutionary Algorithm (RVEA) [20], and the most appropriate is selected as the appropriate model for solving the above problems. Before deploying the intelligent decision-making model, the corresponding optimization model parameters must be adjusted on the collected data. Considering the equipment's performance limit, the number of populations and the maximum number of optimization iterations are set to 50 and 2000, respectively.

## 4. Verification and Analysis

### 4.1. Tomato Greenhouse Environmental Data Processing and Analysis

The collected dataset contains 23,432 records, covering four months. It has six duplicate data records, and proportionally negligible data are available. Figure 3 depicts the variation curves of each feature over time.

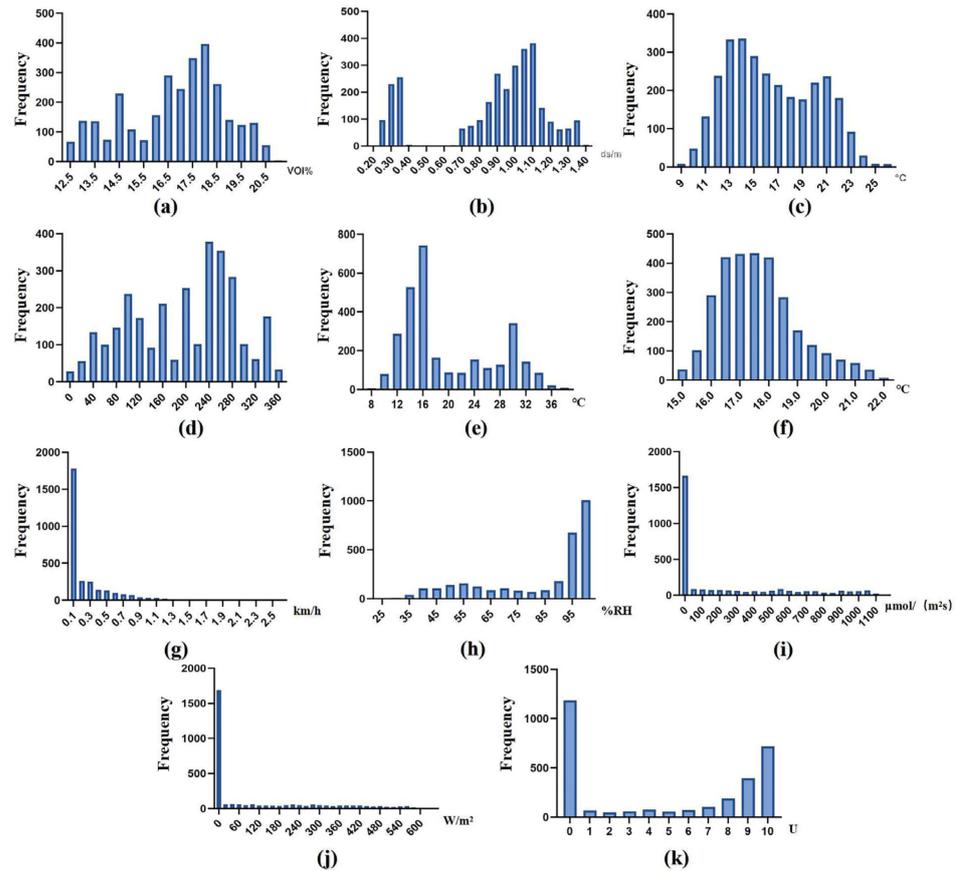


**Figure 3.** Variation in different features over time: (a) soil moisture, (b) soil salinity, (c) photosynthetically active radiation, (d) total radiation, (e) humidity inside the greenhouse, (f) leaf dampness, (g) wind speed, and (h) wind direction.

Figure 3 highlights that the greenhouse environmental data exhibit certain patterns over time, with one data type presenting similar changing trends. For instance, the changing trends of soil moisture and salinity are similar. This characteristic change is consistent with the general understanding that salt concentration is lower when there is a significant amount of water in the soil. This is closely related to the concentration of salts required by plants. Therefore, it indicates that these two factors play an important role in regulating basic elements for plant growth.

Another data type presents periodic changes, e.g., radiation, humidity, and wind speed, which similarly change over time in small periods but have differences in long ones. This suggests that this data type is sensitive to the trend of time changes. The main reason for this periodic change is that light heavily affects this factor, as, during the day, light intensity and temperature are high, leading to more water vapor evaporation. In contrast, the temperature decreases at night due to the decline in light, leading to a decline in temperature and humidity. These findings prove the necessity of adding the time factor (day and night) to the parameter prediction model as it improves the fitting ability of the prediction model to identify features.

Figure 4 reveals that the data distribution has three common characteristics: normal, multimodal, and long tail. Moreover, different feature data types lead to different data amplitudes. For example, the data range of effective light radiation ranges from 18 to 1592  $\mu\text{mol}/\text{m}^2\text{s}$ , whereas the humidity ranges from 9.4 to 24.04 vol%. A typical solution to eliminate this difference is normalizing the data. After experimental comparison, it was found that the data training results after normalization were almost consistent with those without normalization. It should be noted that this study does not adopt normalization processing to reduce the consumption of computing resources.



**Figure 4.** Data distribution histogram: (a) soil moisture, (b) salinity, (c) temperature inside the greenhouse, (d) wind direction, (e) internal temperature of equipment, (f) soil temperature, (g) wind speed, (h) humidity inside the greenhouse, (i) photosynthetically active radiation, (j) total radiation, and (k) leaf dampness.

#### 4.2. Tomato Growth Environment Prediction Model and Verification

As mentioned in Section 3, the model will first learn and predict the growth stages of different crop labels. Then, the predicted growth stages of this model will be associated with the knowledge base information that also includes the growth stages to establish a connection between the model and the knowledge base. Then, the data will be learned from the associated knowledge base model to achieve the adaptation parameter prediction model from the machine learning model to the knowledge base in the greenhouse environment. Then, NSGA-3 will be used to optimize the multi-objective solving problem established. As multiple knowledge base parameters are considered in this optimization problem, a recommended decision indicator for greenhouse environment regulation can be obtained. The following first evaluates the results of the prediction model on tomato greenhouse data to demonstrate its actual effectiveness.

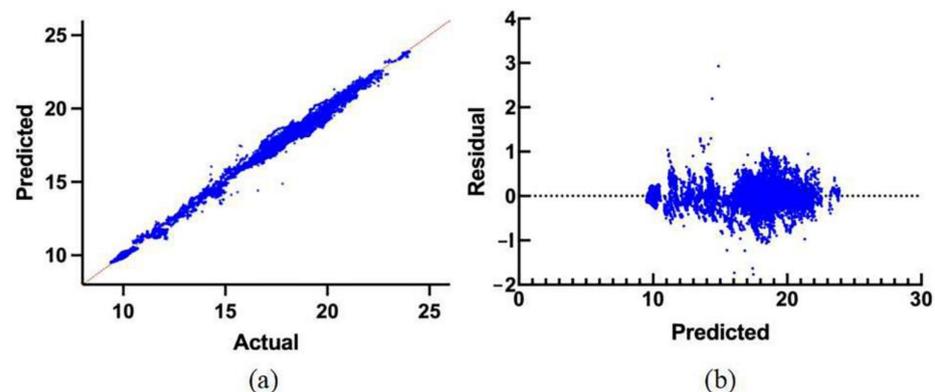
Once the collected greenhouse data are processed and analyzed, they are inputted into the model library to construct a tomato growth environment prediction model, where the model input data of different algorithms in the model library are the same. The optimal prediction model is then selected by comparing the learning capabilities of different models under different environmental parameters. Taking the model learning results of soil moisture and temperature as examples, the two indicators obtained the best values on the gradient boosting tree model and the decision tree model, respectively. Meanwhile, to demonstrate the learning and process of the model, a five-fold cross-validation of the model effect is conducted. Next, the soil moisture and temperature factors are taken as examples to present the five-fold cross-validation results of the model.

As listed in Table 5, the model exhibits different cross-validation effects on different data after using cross-validation to divide the learning data. However, the results are relatively close. Therefore, it is considered that the model obtains more accurate learning results for different data and better results on various indicators for subsequent optimization model modeling.

**Table 5.** Five-fold cross-validation results of the soil moisture factor on the gradient boosting tree model.

Fitting Time (s)	Prediction Time (s)	max_Error	mae	mse	rmse	r2	mgd	mape
2.634	0.014	2.369	0.579	0.675	0.821	0.892	0.003	0.042
2.908	0.01	1.26	0.485	0.323	0.569	0.975	0.002	0.034
2.676	0.01	2.399	0.339	0.221	0.47	0.902	0.001	0.017
2.938	0.01	2.291	0.516	0.459	0.678	0.903	0.002	0.033
2.861	0.009	2.162	0.568	0.475	0.689	0.836	0.001	0.03

As shown in Figure 5, the comparison between the model prediction results and the actual values reveals that the model prediction values are relatively close to the actual values, except for a few points. Meanwhile, the residual diagram reveals no evident morphological result between the model prediction results and the residual values, indicating that the model completes the prediction well.

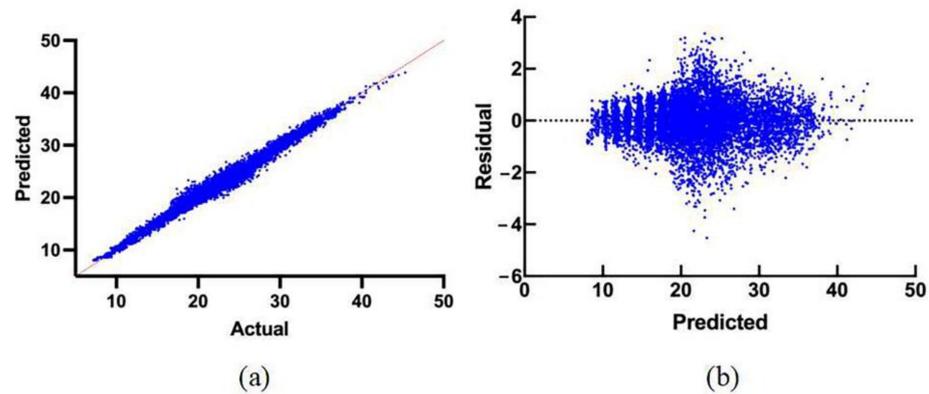


**Figure 5.** (a) Predicted and true soil moisture values and (b) comparison of predicted values and residuals.

Similar to the soil moisture factor test results, the cross-validation results present certain differences but are still relatively close (Table 6). Therefore, the learning results of the model for different segmented data are more accurate, and the results for each indicator are better. Hence, the model obtained by learning the temperature factor can also be used for subsequent optimization model modeling. The comparison results in Figure 6 demonstrate that the model has achieved good predictions.

**Table 6.** Five-fold cross-validation results of the temperature factor on the decision tree.

Fitting Time (s)	Prediction Time (s)	max_Error	Mae	mse	rmse	r2	mgd	mape
0.118	0.005	5.5	0.768	0.991	0.995	0.974	0.002	0.033
0.115	0.005	10.2	0.958	1.909	1.381	0.951	0.004	0.043
0.14	0.005	8.9	1.037	2.435	1.56	0.898	0.006	0.054
0.121	0.005	6.9	1.056	2.395	1.547	0.823	0.006	0.062
0.12	0.005	8.8	1.187	2.965	1.722	0.929	0.005	0.051



**Figure 6.** (a) Predicted and true values of the temperature factor and (b) comparison of predicted values and residuals.

The appropriate model is selected as the greenhouse environment feature prediction model by comprehensively considering various indicators. Different models show different effects on different features. Thus, we can determine the appropriate model for greenhouse environmental data prediction by comparing these combinations. Table 7 reports the experimental results during the model selection process.

**Table 7.** Model selection after comparison of cross-validation results.

Features	Models
Growth stage	Gradient boosting tree
Temperature	Gradient boosting tree
Soil moisture	Decision tree
Air humidity	Decision tree
N	Decision tree
P	Decision tree
K	Gradient boosting tree
Conductivity	Decision tree

Note: N, P, K represent N fertilizer, P fertilizer and K fertilizer, respectively.

There are two main greenhouse environmental feature prediction models: the decision tree model and the gradient boosting tree model. These two models are similar as they are based on tree models. Compared with the decision tree model, the gradient boosting tree integrates the prediction results of data on multiple small partitioned datasets. Meanwhile, according to the above data analysis, the greenhouse environmental data exhibit certain characteristics of cyclical changes over time. In contrast, the tree model can find the dividing points of the data features. Therefore, the performance of these two models makes them more advantageous than other models. Moreover, the five-fold cross-validation results of the temperature factor and soil moisture factor mentioned above reveal that the results of the two models are similar and consistent with the analysis results.

From the results, it can be seen that the models trained through the model library mainly use GDBT and decision tree models in processing greenhouse tomato environmental data. This indicates that, for different features, suitable models are different. Although in the same environment, compared to traditional manual experimental decision-making, using the model library can help select the most suitable model in real time based on the data changes corresponding to environmental changes. This enables the modeling process to use real-time detection data, eliminating the dependence on pre-collected data and the time cost required for selecting different models, and can ensure that its prediction results are always in an optimal state at the model end. This is what traditional machine learning tasks lack in the modeling process.

### 4.3. Tomato Environmental Decision-Making Optimization Model and Analysis

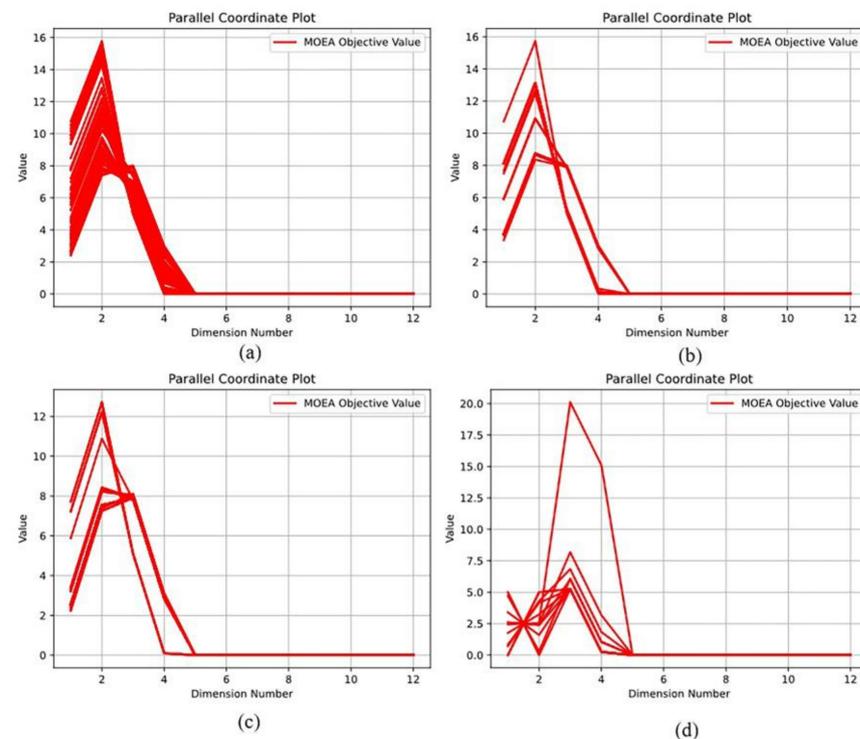
Generally, the greater the population iterations, the more representative the optimization results. However, there is a trade-off between the number of populations and the number of iterations. To guarantee that the model optimization structure is sufficiently representative and considering the model effect, training time, and equipment performance limitations, the number of populations is set to 50, and the maximum number of optimization iterations is 2000.

Moreover, to better initialize the model, we employ a value within the data range of the empirical knowledge base. As shown in Table 8, this is because combining expert knowledge and historical experience helps improve the performance and convergence speed of the model and can guide the model to find the optimal solution.

**Table 8.** Initial values of the model.

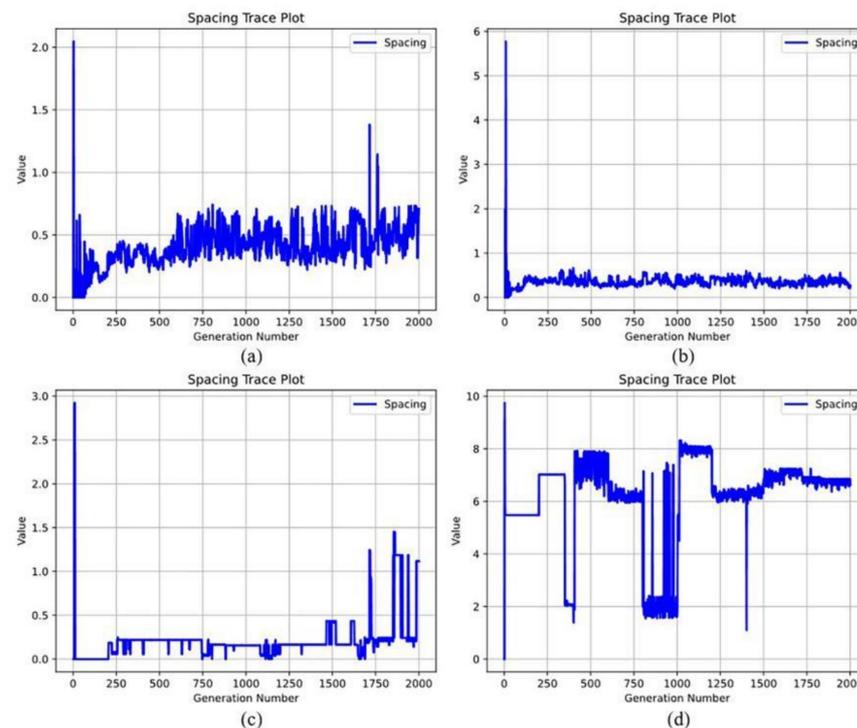
Features	Temperature (°C)	Soil Moisture (%)	Air Humidity (%)	N/(mg·L <sup>-1</sup> )	P/(mg·L <sup>-1</sup> )	K/(mg·L <sup>-1</sup> )	Conductivity/(ms·cm <sup>-1</sup> )
Values	25	60	80	120	90	140	1.5

Taking the fertilization decision-making objective as an example to construct an optimization model, and given the constraints and objective function of the greenhouse environment decision-making optimization model, Figure 7 illustrates the number of Pareto solutions obtained by executing the model using NSGA-2, NSGA-3, and PPS MOEAD DE, demonstrating similar results. Although the values of the first five variables differ, they have a similar trend. Meanwhile, consistent results are obtained for the remaining factors across the three models. Comparing the optimization results of the RVEA model reveals large differences among the results, and the optimization results are rather different from those of the other models.



**Figure 7.** Horizontal comparison of the number of Pareto solutions of the four models. Different horizontal lines represent different Pareto solutions. (a) NSGA-2, (b) NSGA-3, (c) PPS MOEAD DE, and (d) RVEA.

Figure 8 infers that, as the number of iterations evolves, the average value of the NSGA-3 model target space decreases rapidly, then increases, and stabilizes with gradual fluctuations. According to these results, the decision-making results of the greenhouse environment optimization decision-making model gradually stabilize, and the results of the NSGA-3 optimization model converge. However, the results of the other three models still fluctuate in the later stages of evolution. This analysis indicates that the optimization results of the NSGA-3 model on greenhouse environmental features can fully represent the optimal decision-making results in the greenhouse environment.



**Figure 8.** Changes in the target space value with the increase in the number of iterations (a) NSGA-2, (b) NSGA-3, (c) PPS MOEAD DE, and (d) RVEA.

After several experiments, the average optimization time of the model is 2 min and 31 s, with a model time complexity of  $O(MN^2)$ , where  $M = 12$  based on the number of objective functions. Since the greenhouse environment in which the model is evaluated is adjusted periodically and the optimal solution is obtained through optimization, the execution speed meets the needs of greenhouse environment optimization decisions.

Although there are certain differences in the optimization parameters of different optimization models, it can be seen that the overall results obtained by different optimization models are relatively similar. This means that different methods have been used to find more approximate solutions, indicating that the similarity of the results obtained by using the model library to learn real-time data and the optimization model for the optimization process is relatively robust. At the same time, in this section, by comparing various cutting-edge optimization methods, it is ensured that the optimization model chosen in the end can effectively solve the optimization decision of tomato greenhouse environmental parameters in this paper.

## 5. Discussion and Conclusions

This study utilizes real-time multi-node distributed sensor data to monitor and collect greenhouse environmental data. The processing and analysis of the collected feature information provide an accurate basis for precise decision-making in the greenhouse environment, enhancing greenhouse productivity while reducing costs. Meanwhile, this

study searched relevant textbooks and the literature, collected historical production data, conducted expert consultation, and performed regression analysis on relevant data. Based on this, a knowledge base of characteristic parameter ranges in a standard growth environment was designed for greenhouse crops (such as tomatoes) during each growth period. Then, the knowledge base and greenhouse environmental data were integrated to predict decision-making solutions, allowing the model to be widely applied to the environmental decision-making needs of different greenhouse crops. This provides an effective data basis and a direction for optimization of the decision-making solution for the prediction model.

The main feature factors of greenhouse crop growth are determined through correlation feature analysis, and a greenhouse crop environmental decision-making model library is constructed based on the data. Combined with the knowledge base parameters of various crops at different growth stages, the model library can be applied to different crops and select the optimal prediction model. The model obtains appealing results in predicting greenhouse environmental changes under different influencing parameters. Based on this model, intelligent parameter decision-making suggestions can be provided for greenhouses to improve the efficiency of greenhouse energy utilization and crop growth. The closeness between the predicted and real values validates the model's robustness, and its precise decision-making plan is implemented and verified in an actual production environment. Moreover, based on the model library, an intelligent optimization model is constructed by taking the fertilization decision-making target as an example. The model can utilize local experience data as initial values for optimization. By comparing different models and value changes in the target space, NSGA-3 is determined as the greenhouse fertilization decision-making optimization model. It should be noted that the optimization results are consistent with the experience results, indicating that the model can meet the needs of intelligent optimization decision-making.

The model established in this article can adaptively select machine learning models in tomato greenhouses based on real-time monitoring data, learn changes in greenhouse environment data, and use the model obtained in this way as a component of the optimization method objective function. Combined with the knowledge base established through expert prior knowledge to obtain the optimal greenhouse environment parameter settings in the current environment, the cost of building machine learning models in tomato greenhouse agriculture can be reduced, and the corresponding adaptive ability of the model can be improved. At the same time, this model base, real-time data, knowledge base, and optimization methods combined in multiple aspects can evidently be generalized to different forms of agricultural planting, providing a paradigm for modeling other greenhouse crop environments.

Meanwhile, during the research process, we found that the proposed model still has certain shortcomings. For example, in the process of optimizing greenhouse parameters, the prediction model and optimization model have a high dependence on the expert knowledge base. If the parameters of the knowledge base cannot represent the corresponding indicators required for different planting stages of crops well, such as ammonia nitrogen concentration, the suggested parameter results will also have deviations. Therefore, in future research, further optimization processing of the knowledge base information will be carried out to improve the robustness of the integrated knowledge base model proposed in this paper.

This study successfully constructs a greenhouse crop environmental decision-making model library and an intelligent decision-making optimization model for greenhouse fertilization. These models can provide effective decision support for the environmental management of smart greenhouses. Considering the complex coupling relationships of various environmental characteristics in the greenhouse environment, future work will combine the coupling relationships of various environmental characteristics of the greenhouse, such as the mutual effects of light, temperature, and humidity, and accumulate various greenhouse crop growth data and environmental data to enrich the knowledge base. Meanwhile, we will focus on connecting with intelligent greenhouse environmental control equipment and verifying and continuously improving precise decision-making

plans in the actual production environment, further enhancing the model's practicality and reliability.

**Author Contributions:** Y.T. and C.B. conceived and designed the research. Z.Z. conducted experiments and analyzed the data. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Key R&D projects of Jilin Provincial Department of Science and Technology of China, grant number 20200402117NC.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author Tang Y., upon reasonable request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Jiang, J.; Wu, Y.; Liu, Q.; Liu, Y.; Cao, Q.; Tian, Y.; Zhu, Y.; Cao, W.X.; Liu, X.J. Developing an efficiency and energy-saving nitrogen management strategy for winter wheat based on the UAV multispectral imagery and machine learning algorithm. *Precis. Agric.* **2023**, *24*, 2019–2043. [[CrossRef](#)]
- Venkatesan, S.K.; Lim, J.; Ko, H.; Cho, Y.Y. A Machine Learning Based Model for Energy Usage Peak Prediction in Smart Farms. *Electronics* **2022**, *11*, 218. [[CrossRef](#)]
- Bolandnazar, D.; Sadriani, H.; Rohami, A.; Marinillo, F.; Taki, A. Application of Artificial Intelligence for Modeling the Internal Environment Condition of Polyethylene Greenhouses. *Agriculture* **2023**, *13*, 1583. [[CrossRef](#)]
- Zhang, L.X.; Tian, X.; Li, Y.X.; Chen, Y.Q.; Chen, Y.Y.; Ma, J.C. Estimation of Disease Severity for Downy Mildew of Greenhouse Cucumber Based on Visible Spectral and Machine Learning. *Spectrosc. Spectr. Anal.* **2020**, *40*, 227–232.
- Li, R.Z.; Ji, Z.J.; Hu, S.K.; Huang, X.D.; Yang, J.L.; Li, W.F. Tomato Maturity Recognition Model Based on Improved YOLOv5 in Greenhouse. *Agronomy* **2023**, *13*, 603. [[CrossRef](#)]
- Raza, A.; Shahid, M.A.; Safdar, M.; Tariq, M.A.R.; Zaman, M.; Hassan, M.U. Exploring the Impact of Digital Farming on Agricultural Engineering Practices. *Biol. Life Sci. Forum.* **2023**, *27*, 10.
- Zhang, Y.; Wang, T.; Li, Z.; Wang, T.L.; Cao, N. Based on machine learning algorithms for estimating leaf phosphorus concentration of rice using optimized spectral indices and continuous wavelet transform. *Front. Plant Sci.* **2023**, *14*, 1185915. [[CrossRef](#)] [[PubMed](#)]
- Abbood, H.M.; Nouri, N.M.; Riahi, M.; Riahi, M.; Alaghrband, S.H. An intelligent monitoring model for greenhouse microclimate based on RBF Neural Network for optimal setpoint detection. *J. Process Control.* **2023**, *129*, 103037. [[CrossRef](#)]
- Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
- Ladický, L.; Torr, P.H.S. Locally Linear Support Vector Machines. In Proceedings of the 28th International Conference on Machine Learning (ICML-11), Bellevue, WA, USA, 28 June–2 July 2011.
- Scholkopf, B.; Smola, A.J.; Williamson, R.C.; Bartlett, P.L. New Support Vector Algorithms. *Neural Comput.* **2000**, *12*, 1207–1245. [[CrossRef](#)] [[PubMed](#)]
- Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
- Kosmider, C. Induction of Decision Trees. In Proceedings of the IIS'2002 Symposium on Intelligent Information Systems, Sopot, Poland, 3–6 June 2002; Springer: Berlin/Heidelberg, Germany, 2002.
- Hamrani, A.; Akbarzadeh, A.; Madramootoo, C.A. Machine learning for predicting greenhouse gas emissions from agricultural soils. *Sci. Total Environ.* **2020**, *741*, 140338. [[CrossRef](#)] [[PubMed](#)]
- Zhu, J.; Chen, Y.; Li, Z.; Duan, W.; Fang, G.; Wang, C.; He, G.; Wei, W. Using Film-Mulched Drip Irrigation to Improve the Irrigation Water Productivity of Cotton in the Tarim River Basin, Central Asia. *Remote Sens.* **2023**, *15*, 4615. [[CrossRef](#)]
- Kiefer, J.; Wolfowitz, J. Stochastic estimation of the maximum of a regression function. *Ann. Math. Stat.* **1952**, *3*, 462–466. [[CrossRef](#)]
- Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [[CrossRef](#)]
- Deb, K.; Jain, H. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems with Box Constraints. *IEEE Trans. Evol. Comput.* **2014**, *18*, 577–601. [[CrossRef](#)]
- Fan, Z.; Li, W.; Cai, X.; Li, H.; Wei, C.; Zhang, Q.; Deb, K.; Goodman, E. Push and pull search for solving constrained multi-objective optimization problems. *Swarm Evol. Comput.* **2019**, *44*, 665–679. [[CrossRef](#)]
- Cheng, R.; Jin, Y.; Olhofer, M.; Sendhoff, B. A Reference Vector Guided Evolutionary Algorithm for Many-Objective Optimization. *IEEE Trans. Evol. Comput.* **2016**, *20*, 773–791. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.