CONSTRUCTION AND VALIDATION OF A PREDICTIVE MODEL FOR TOMATO ORGAN BIOMASS AT ORGAN SCALE BASED ON STACKING LEARNING
/
基于堆叠学习的番茄器官尺度的生物量预测模型的构建与验证

Qi LIU, Fengpei WANG, Gang LIU, Lian BAI, Wuping ZHANG*)
Software College of Shanxi Agricultural University/China
Tel: +86-15935664481; E-mail: zwping@126.com
DOI: https://doi.org/10.35633/inmateh-74-12

***Keywords:** Stacked Machine Learning; Tomato; Geometric Morphology Data; Linear Regression; Biomass Prediction Model; Organ-Scale*

ABSTRACT
In this study, a stacked machine learning algorithm was constructed with tomato organ biomass as the research object, taking the geometric morphology data of tomato organs as the variables, utilizing eight classical machine learning algorithms as the base-model, and applying the linear regression algorithm as the stacked meta-model. This algorithm was then utilized to establish a prediction model for tomato biomass at the organ scale, and the biomass models of tomato plant leaves and fruits at the organ scale were constructed. The model has R^2=0.86, MAE=0.49, and RMSE=0.81 in predicting leaves, and R^2=0.94, MAE=0.33, and RMSE=0.57 in predicting fruits. The model has practical applications in predicting tomato yield and supply, providing market information, and supporting agricultural investment decisions. It also helps to optimize agricultural production and management, guide industrial development and planning, and improve the efficiency and competitiveness of the agricultural sector.

摘要
本研究以番茄器官生物量为研究对象，将番茄器官的几何形态数据作为变量，利用八种经典机器学习算法作为基础模型，并应用线性回归算法作为堆叠元模型，构建了一种堆叠式机器学习算法。然后利用此算法建立器官尺度上的番茄生物量预测模型，并构建了器官尺度上番茄植株叶片和果实的生物量模型。该模型在预测叶片方面的R2=0.86，MAE=0.49，RMSE=0.81；在预测果实方面的R2=0.94，MAE=0.33，RMSE=0.57。该模型在预测番茄产量和供应、提供市场信息、支持农业投资决策等方面具有实际应用价值，还有助于优化农业生产和管理，指导产业发展和规划，提高农业部门的效率和竞争力。

INTRODUCTION
Tomato (*Solanum lycopersicum L.*) is one of the most important vegetable crops in the world, and it has become one of the largest vegetable crops in the world's cultivation area (*Li, 2013*). Some studies have shown that the dramatic changes in global climate in the 21st century will have a significant impact on biodiversity, which, to some extent, also indicates that the cultivation of high-yielding and stable crop varieties will face greater uncertainty and higher difficulty (*Damtew, 2017*). Therefore, it is necessary and important to obtain timely, rapid, and accurate information on crop growth and development and predict biomass.

Biomass is an important parameter for crop growth, which is directly related to the final yield, and the amount of biomass can effectively reflect the growth of crops (*Colomina et al., 2014*). Biomass modeling is the main method to estimate biomass, is an effective and relatively accurate investigation method (*Wang et al., 2008*), and has become a hot area of biomass research, rapid and accurate monitoring of biomass can be a timely understanding of crop growth and yield prediction, which is of great significance to the production and management of agriculture (*Chen et al., 2016*). However, the traditional methods for determining biomass are not only time-consuming, and slow, with large errors and low efficiency, but also cause damage to the crop during the measurement process, making it difficult to realize the measurement of biomass rapidly (*Liu et al., 2021*).

Domestic and international scholars have done a lot of research on biomass prediction modeling (*Liu et al., 2023*). Many scholars have used new science and technology, such as machine learning, in the production management of agriculture, to effectively realize the prediction and estimation of phenotypic indexes such as vegetation index, above-ground biomass, and chlorophyll content of crops with the help of new technology and new means, such as machine learning and artificial intelligence (*Fu et al., 2021*). *Wang* *et al., (2024*) constructed a biomass prediction model for larch in Xiaoxinganling using diameter at breast height (D) and tree height (H) as variables and found that the machine learning algorithm could predict biomass better than the traditional algorithm. However, most of the current studies have focused on biomass prediction at the individual scale, and not many prediction models have been developed at the organ scale. Moreover, it has been shown that there is a significant correlation between the morphological data of tomato organs and the amount of material produced by them (*Dong et al., 2007*). Therefore, in this study, a tomato organ biomass prediction model was developed at the organ scale based on machine learning techniques, taking tomato organ biomass as output parameters and tomato organ geometric morphology data as the input variables, to provide fast biomass prediction at the organ scale for tomato production, research, and breeding, to provide a theoretical basis and reference basis for tomato variety selection, cultivation management and production monitoring, and to provide experimental basis and scientific basis for tomato yield prediction and cost input.

MATERIALS AND METHODS

Test Material and Test Site

The test tomato variety is YOUCUI8850, which is an infinite-growth large-fruited variety of tomato, a variety of medium-early maturity; the fruit is nearly round, turns red and bright in colour after maturity, has high hardness, a moderate size, and continuous fruit-setting ability; the weight of a single fruit is 190 g - 260 g; storage and transportation resistance, long shelf-life; resistance to the tomato yellowing curculio virus, blight, and tobacco mosaic virus, etc. The developmental stages of tomato fruit, specifically the green maturity and complete ripening phases, are shown in Fig. 1 and Fig. 2.

Fig. 1 - Green maturity of tomato fruits

Fig. 2 - Complete ripening of tomato fruits

The test site is located in Datong City, Shanxi Province, Yanggao County, West Lijia Soap Village, Tomato Industry Research Institute test base (longitude 113°40′42″, latitude 40°09′50″, elevation 1125 m) to carry out the test site for the meso-thermal temperate continental semiarid monsoon climate, the average annual temperature of 7.1°C, the temperature difference between day and night is obvious; the average annual number of hours of sunshine is 2,691.4 h; the average annual precipitation is 364.9mm; the annual frost-free period is 161 days; the soil is mainly loamy and of medium fertility. The test site is shown in Fig. 3.

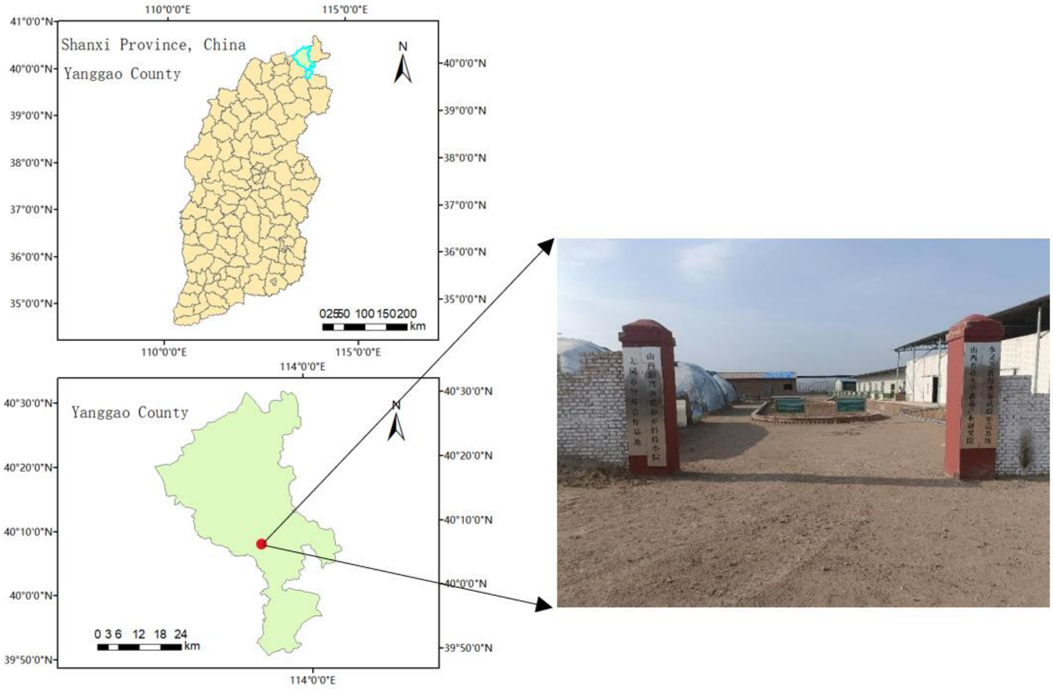

Fig. 3 - Visualization of test site locations

The experimental greenhouse was located in the north-south direction, with a length of 70 m, a span of 12 m, and a ridge height of 8.6 m. The planting ridge was in the shape of a trapezoid, with a length of 10 m, a width of 110 cm at the base of the ridge, a width of 60 cm at the surface of the ridge, a height of 40 cm, and a spacing of 50 cm between neighbouring ridges; the ridge was planted with two rows, two rows staggered, with a spacing of 50 cm between the rows and a spacing of 40 cm between the plants. The layout of tomato planting inside the greenhouse is shown in Fig. 4.

Fig. 4 - Inside the greenhouse tomato planting scene

The experiment started on May 15, 2023 when the tomato seedlings were planted and continued till August 13, 2023. The trial was managed routinely during the trial period.

The experimental environment used in this study: CPU: Intel(R) Core (TM) i7-13700K@3.40GHZ, 64GB of running memory, Operating System: Windows 11, GPU: NVIDIA GeForce RTX 4080 with 16GB video memory. The machine learning framework is scikit-learn 1.0.2, and the programming language is Python 3.7.0.

Data Acquisition and Processing

One week after planting, six YOUCUI8850 plants with consistent growth were selected for observation. Additionally, six plants with similar growth potential and morphology were selected for harvesting at intervals of 10-15 days, and a total of seven harvests were conducted. At each harvest, the selected plants were carefully removed from the soil and quickly brought back to the laboratory, where each of the aboveground organs of the tomato was individually sectioned from the base to the apical growth point. Then the length and width of each leaf, the transverse and longitudinal diameter of each fruit were quickly measured, after which each organ was placed in a kraft paper bag into the oven to kill the green at 105° for 30 minutes, and then dried at a constant temperature of 80°C, until the mass was constant. Then, the biomass (dry mass) of each organ was measured (*Cheng et al., 2022*).

Geometric morphometric data on tomato leaves and fruits from each harvest were recorded and they were numbered according to their position on the plant. Fig. 5 and Fig. 6 show the geometrical morphometric data of the leaves at different positions at each harvest. It can be seen that the length and width of the leaves at different positions showed an increasing trend with the growth time. Fig. 7 and Fig. 8 show the geometrical morphometric data of the fruits at different positions at each harvest, and it can be seen that the transverse and longitudinal diameters of the fruits at different positions show the same trend of growth with the change in growth time. Moreover, the average leaves and fruits biomass per plant at each harvest and the average plant height and stem thickness were recorded. As shown in Fig. 9, leaf biomass, fruit biomass, plant height, and stem thickness increased over time, indicating that the plants were maturing as they grew.

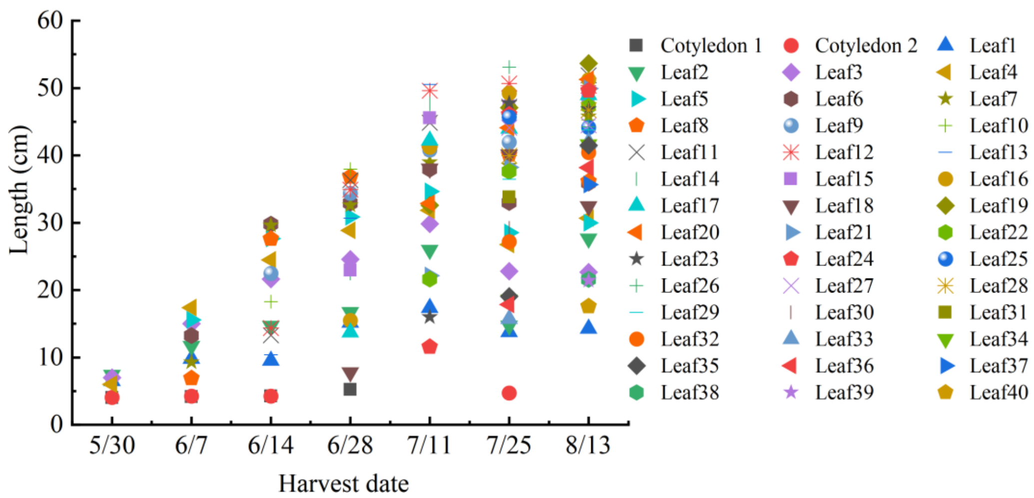

Fig. 5 - The data on leaf length per harvest

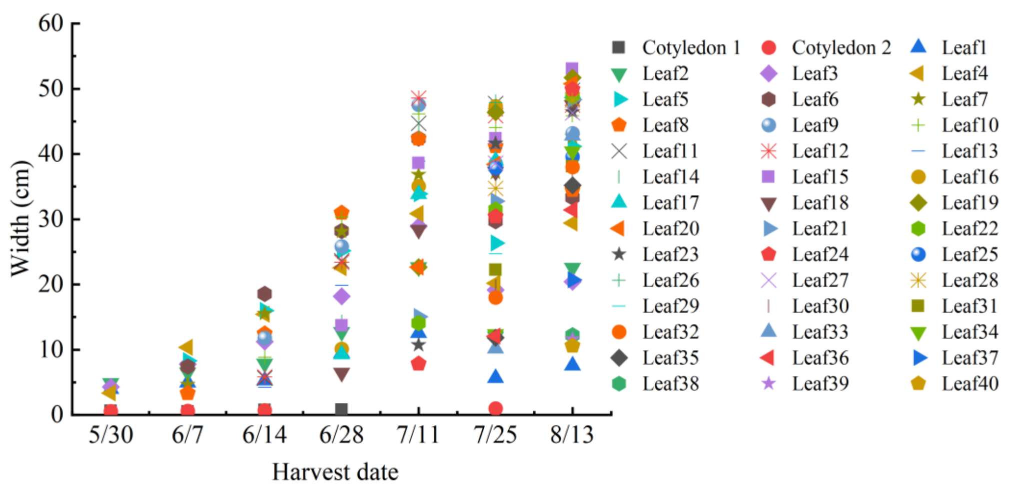

Fig. 6 - The data on leaf width per harvest

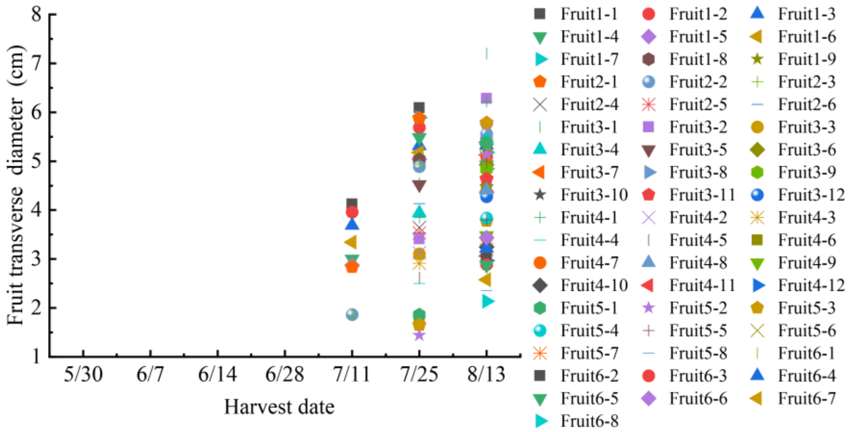

Fig. 7 - The data on fruit transverse diameter per harvest

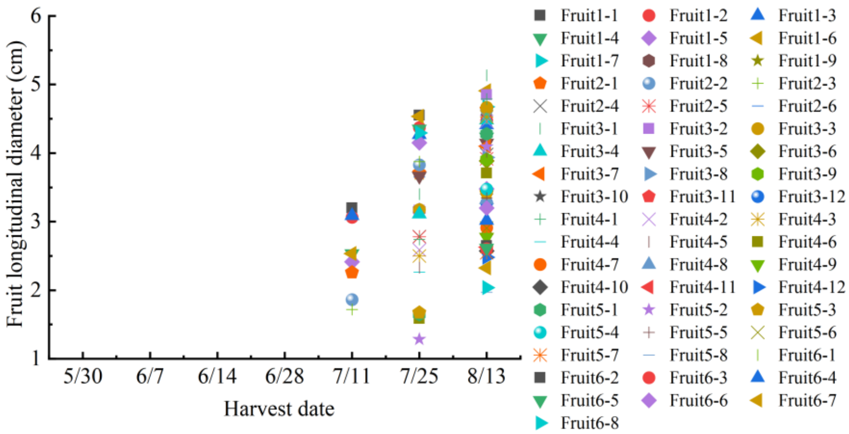

Fig. 8 - The data on fruit longitudinal diameter per harvest

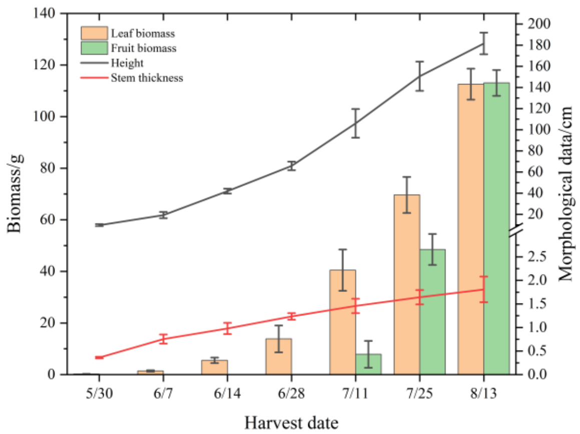

Fig. 9 - Average leaf and fruit biomass per plant and plant height and width per harvest

Model Selection

In this study, classical machine learning algorithms were used to predict the biomass of tomato organs. Among these algorithms, the generalized linear regression supported Ridge Regression (Ridge), Lasso regression algorithm (Least absolute shrinkage and selection operator, Lasso), and ElasticNet (EN) were used, Support Vector Machine algorithm (SVM), Multi-Layer perceptron (MLP), K-NearestNeighbor (KNN), Random Forest (RF) algorithm belonging to bagging in integrated learning, and Gradient Boosting Decision Tree (GBDT) algorithm which belongs to boosting in integrated learning. In addition, the above algorithmic models were also stacked based on the stacking mode of integrated learning to construct a model to realize the prediction of tomato biomass.

In the field of machine learning, the choice of algorithms and the tuning of parameters have always been headache-inducing challenges. Although there are many algorithms available, no algorithm is foolproof. As technology continues to evolve, new techniques have emerged that can provide some help in algorithm selection and parameter tuning, and one of the most popular techniques is Stacking.

Stacking is a technique used to enhance the performance of machine learning models (*Maddaloni et al, 2022; Jahnavi et al., 2023*). The technique generates final predictions by combining predictions from different algorithms. This approach can help to solve many machine learning problems, especially when a single algorithm is not sufficient to solve the problem. Stacking usually consists of two steps: the first step is to use multiple base models to generate predictions, and the second step is to use another model to integrate these predictions and generate the final prediction. The basic process of its implementation is generally divided into two steps. The first step is to generate the prediction results. In the first step, multiple base models were used to generate prediction results. For each base model, the training data are split into two parts: one part is used to train the model and the other part is used to generate the prediction results. Different models such as decision trees, random forests, support vector machines, multilayer perceptual machines, etc. can be used. Each model generates a prediction result; the second step integrates the prediction results. In the second step, another model is used to integrate these predictions and generate the final prediction results. Algorithms such as linear regression, logistic regression, random forests, support vector machines, neural networks, etc. can be used to accomplish this step. One thing that must be noted here is that the model in the second step must use the predictions from the first step as input. This will ensure the consistency of the entire Stacking process.

RESULTS

Result Evaluation

In this study, R^2, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) were used to evaluate the model. The closer the value of R^2 is to 1, the better the interpretability and performance of the model. The smaller the MAE and RMSE values, the higher the consistency between the predicted values and the actual values, thus proving the more accurate prediction results of the model (*Wang et al., 2022*).

Construction and Evaluation of Tomato Leaf Biomass Prediction Model

The leaf is an important organ of the plant, which is the main tissue for photosynthesis, absorbing energy from sunlight through chlorophyll and other photosynthetic pigments and converting it into chemical energy to support plant growth and development. Leaves also absorb water from the plant body from the ground to the atmosphere through transpiration, facilitating water and nutrient uptake and transport, and regulating the temperature of the plant body.

By quickly realizing the prediction of leaf biomass, it is possible to understand the total amount and trends of leaves in different plant communities and ecosystems. This helps to assess the dynamic distribution and spatial pattern of plant biomass and reveals the structure and function of ecosystems, thus providing important information for ecological studies.

After all the data obtained from the seven experiments were divided into a training set and a test set according to the different organs in the ratio of 75% and 25%, then a model was constructed with the geometric morphology data of leaves (length and width of leaves) and its corresponding biomass data and the geometric morphology data were uses as inputs to the model to predict its corresponding biomass. The eight classical machine learning models (Ridge, Lasso, EN, SVM, MLP, KNN, RF, and GBDT), with their corresponding hyperparameters are shown in Table1.

Table 1

Hyperparameters of the 8 base models when the input feature of the model is the geometric shape of the leaves

Model	Para1	Value1	Para2	Value2	Para3	Value3
Ridge	alpha	10				
Lasso	alpha	0.01	max_iter	10		
EN	alpha	0.1	l1_ratio	0.9		
SVM	C	1	kernel	rbf		
MLP	activation	relu	hidden_layer_sizes	(25,25, 25,25)	solver	adam
KNN	algorithm	ball_tree	leaf_size	3	n_neighbors	9
RF	min_samples_leaf	6	min_samples_split	0.1	n_estimators	6
GBDT	learning_rate	0.1	loss	lad	n_estimators	96

The above 8 base models are taken as base-model and linear regression algorithm is used as stacking meta-model. The above 8 classical machine learning models were integrated, using common input data, and 9 different sets of predicted data were obtained, after normalizing the predicted values with the actual values to construct a 1:1 comparison graph, the specific results are shown in Fig. 10 and Fig. 11.

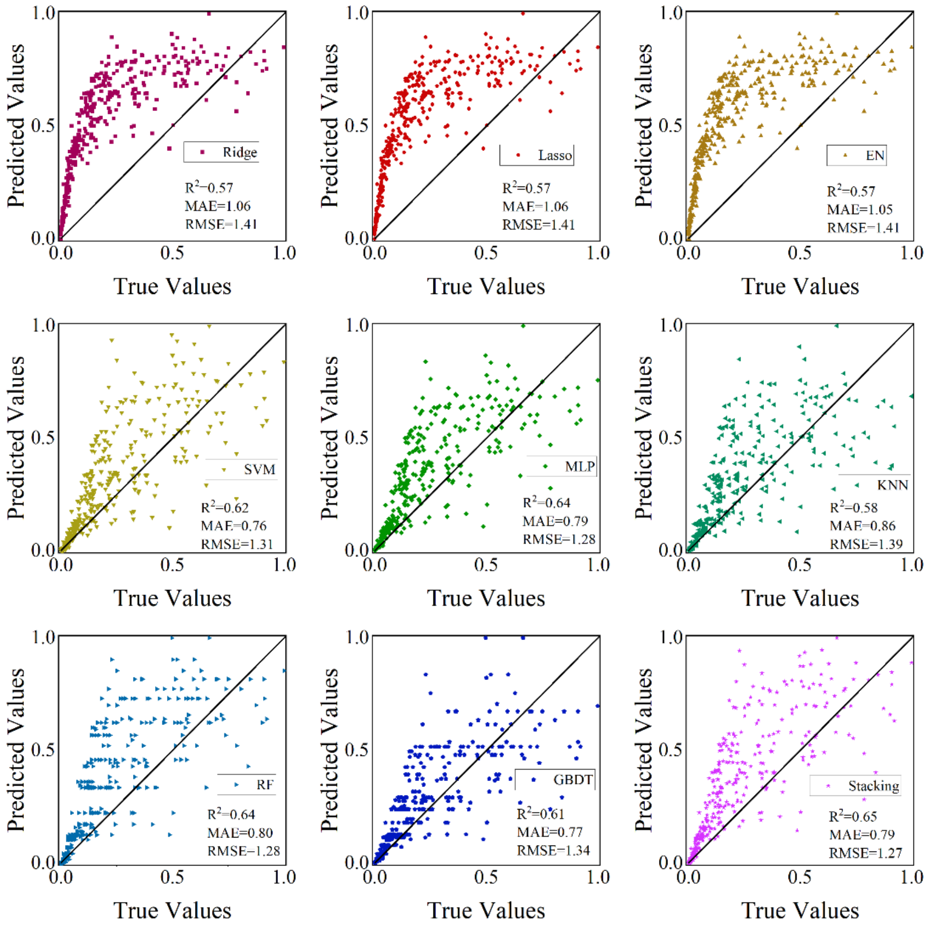

Fig. 10 - The 1:1 plot of predicted versus actual values for the nine models when the input feature of the model is the geometry of the blade

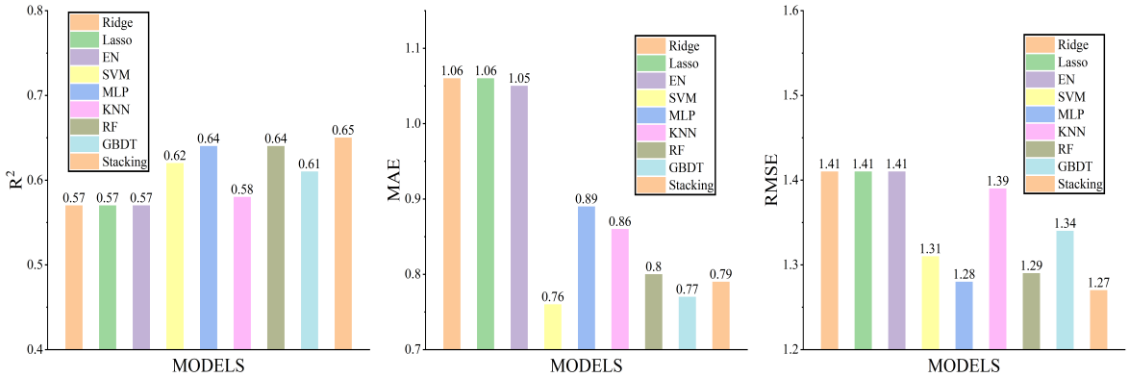

Fig. 11 - Comparison of the evaluation metrics of the nine models when the input feature of the model is the geometric form of the blade

The prediction results of MLP and RF in a single model are optimal. The model constructed based on the stacking approach is superior to the base learner whose prediction results are best. However, the prediction results of the nine models were not very good, so plant height and stem thickness of the current plant individuals of the leaf geometry data were added to the input features of the models. The results obtained were superior to the prediction results without adding the individuals' plant height and stem thickness.

The hyperparameters of the 8 machine-learning models with the addition of individual plant height and stem thickness are shown in Table 2.

Table 2

Hyperparameters of the 8 base models when the input features of the models are the geometric shape of the leaves and the height and stem thickness of the corresponding individual plants

Model	Para1	Value1	Para2	Value2	Para3	Value3
Ridge	alpha	1				
Lasso	alpha	0.001	max_iter	20		
EN	alpha	0.1	l1_ratio	0.1		
SVM	C	5	kernel	rbf		
MLP	activation	relu	hidden_layer_sizes	(25,25, 25,25)	solver	adam
KNN	algorithm	kd_tree	leaf_size	1	n_neighbors	5
RF	min_samples_leaf	3	min_samples_split	0.1	n_estimators	51
GBDT	learning_rate	0.1	loss	ls	n_estimators	96

Similarly, the above 8 base models were used as the base-model and the linear regression algorithm was used as the meta-model for stacking. The obtained 9 sets of predicted data were normalized and a 1:1 comparison graph between them and the actual values was constructed, and the specific results are shown in Fig. 12 and Fig. 13.

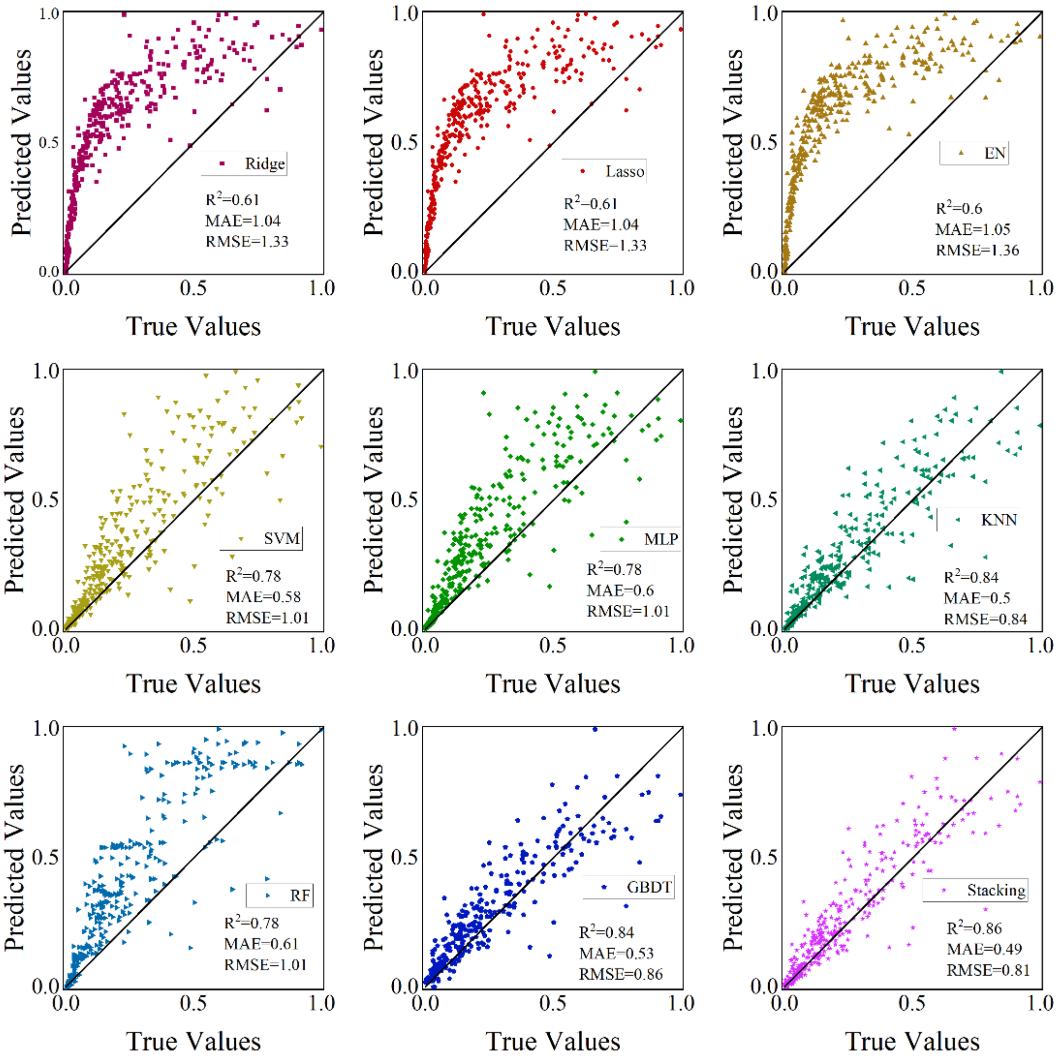

Fig. 12 - The 1:1 plot of predicted versus actual values for the nine models when the input features of the models are the geometric shape of the leaves versus the height and stem thickness of the corresponding individual plants

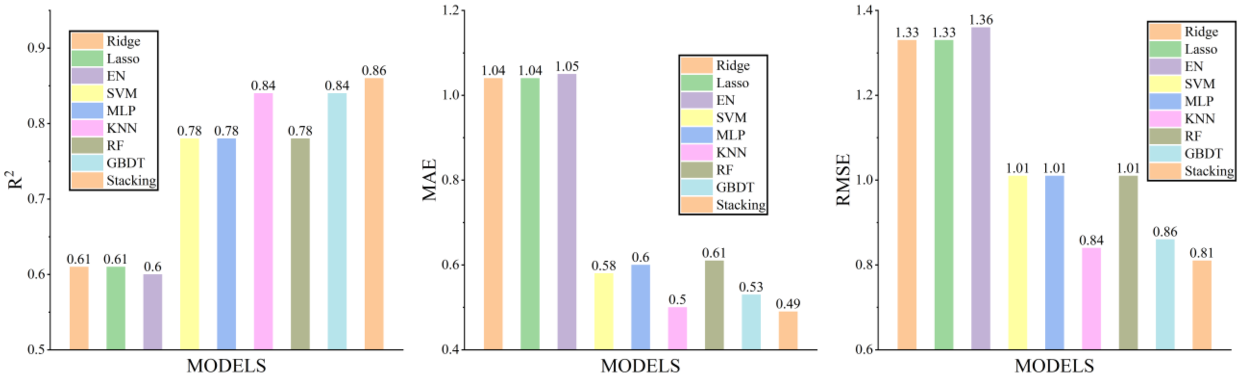

Fig. 13 - Comparison of the evaluation metrics of the nine models when the input features of the models are the geometric shape of the leaves with the height and stem thickness of the corresponding individual plants

The prediction accuracy of the model was significantly improved with the addition of plant height and stem thickness corresponding to individual plants, and it is desirable to have a prediction model with higher accuracy. In agriculture and horticulture, the prediction of plant leaf biomass enables the assessment of plant growth and yield potential. This helps to adjust fertilization and irrigation strategies, optimize the growing environment of crops, and improve agricultural productivity and resource efficiency. Leaf biomass prediction is important for ecological research and precision agriculture research, promoting sustainable development of agriculture and environmental protection.

Construction and Evaluation of Tomato Fruit Biomass Prediction Models

When a model is constructed with the geometric morphology data of the fruit (transverse and longitudinal diameters of the fruit) and its corresponding biomass data and the geometric morphology data are use as inputs to the model to predict its corresponding biomass, the 8 classical machine learning models (Ridge, Lasso, EN, SVM, MLP, KNN, RF, and GBDT), with their corresponding hyperparameters are shown in Table 3.

Table 3

Hyperparameters of the 8 base models when the input feature of the model is the geometric form of the fruit

Model	Para1	Value1	Para2	Value2	Para3	Value3
Ridge	alpha	50				
Lasso	alpha	0.1	max_iter	10		
EN	alpha	0.1	l1_ratio	0.1		
SVM	C	1	kernel	rbf		
MLP	activation	relu	hidden_layer_sizes	(30, 30, 30)	solver	adam
KNN	algorithm	ball_tree	leaf_size	3	n_neighbors	9
RF	min_samples_leaf	3	min_samples_split	0.1	n_estimators	21
GBDT	learning_rate	0.1	loss	huber	n_estimators	21

Similarly, the stacking model was constructed based on the above-mentioned base model. A 1:1 comparison plot was constructed with the actual values after normalizing the obtained 9 sets of predicted data, as shown in Fig. 14 and Fig. 15.

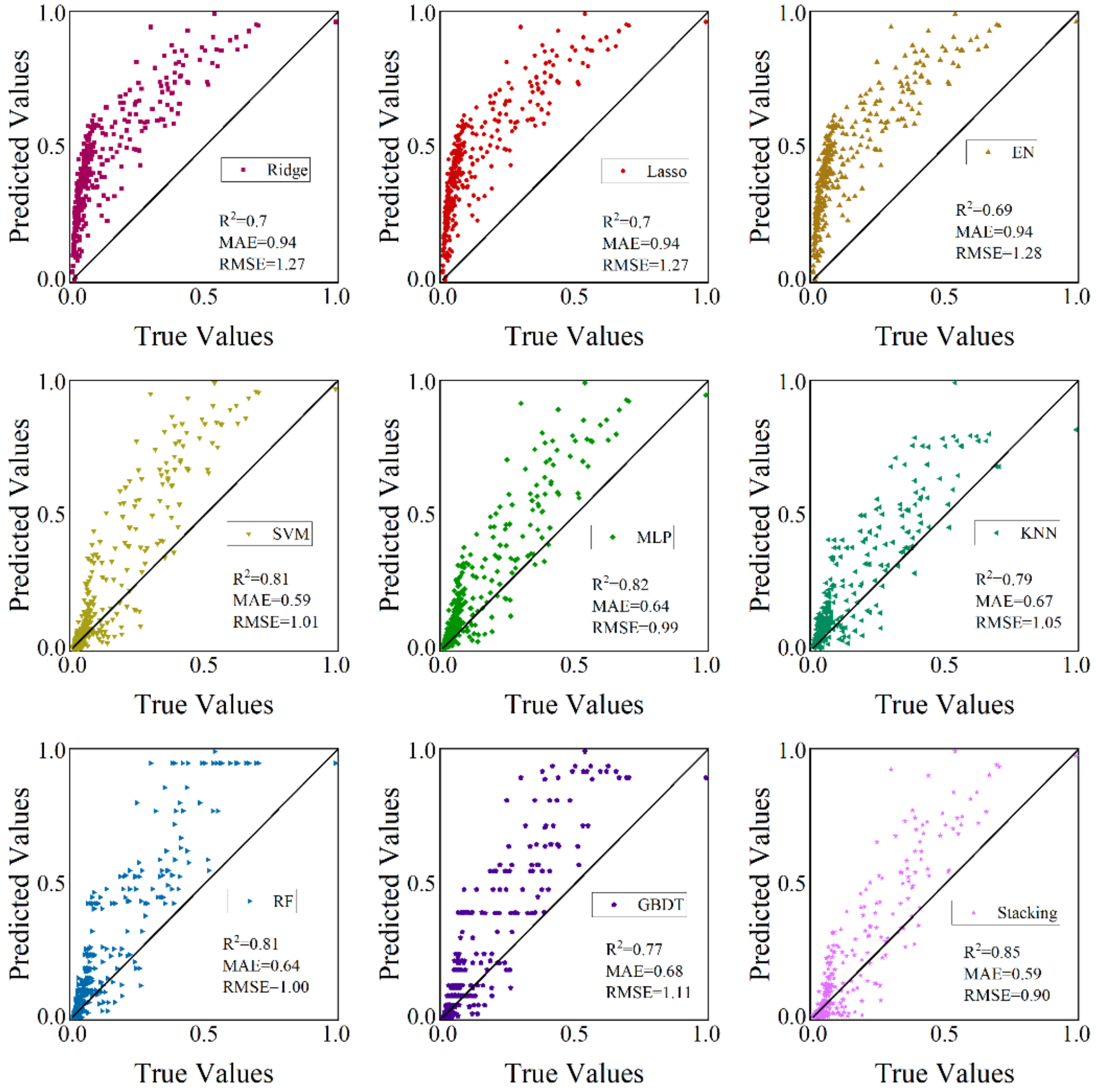

Fig. 14 - The 1:1 plot of predicted versus actual values for the nine models when the input features of the models are the geometric shape of the leaves versus the height and stem thickness of the corresponding individual plants

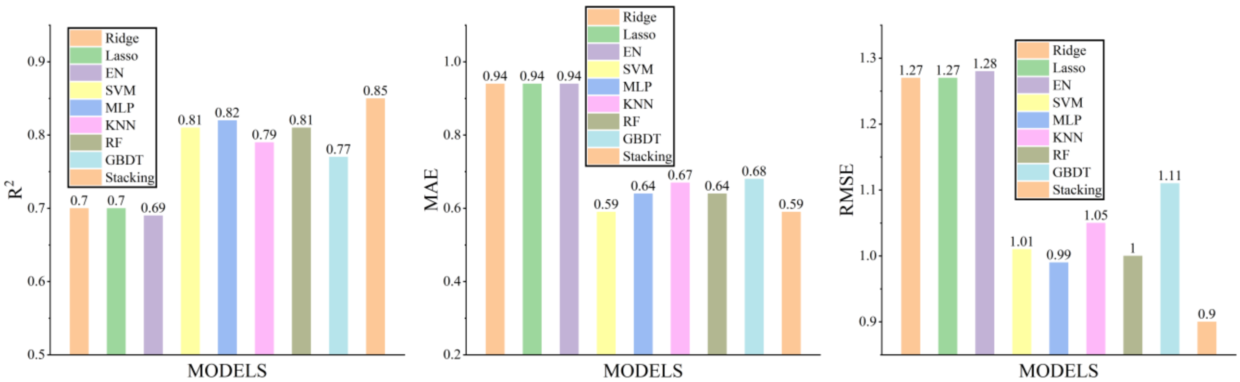

Fig. 15 - Comparison of the evaluation metrics of the nine models when the input feature of the model is the geometric form of the fruit

As in the case of the leaves, the plant height and stem thickness of the corresponding individual plant were added to the input features. The results obtained were again superior to the predictive models without the added plant height and stem thickness. The hyperparameters of the 8 machine learning models after adding individual plant height and stem thickness are shown in Table 4.

Table 4

Hyperparameters of the 8 base models when the input feature of the model is the geometric form of the fruit

Model	Para1	Value1	Para2	Value2	Para3	Value3
Ridge	alpha	10				
Lasso	alpha	0.1	max_iter	10		
EN	alpha	0.1	l1_ratio	0.9		
SVM	C	5	kernel	rbf		
MLP	activation	relu	hidden_layer_sizes	(25, 25, 25, 25)	solver	lbfgs
KNN	algorithm	ball_tree	leaf_size	3	n_neighbors	9
RF	min_samples_leaf	3	min_samples_split	0.1	n_estimators	21
GBDT	learning_rate	0.1	loss	huber	n_estimators	21

The above 8 base models were used as the base-model and the linear regression algorithm was used as the meta-model for stacking. The obtained 9 sets of predicted data were normalized and a 1:1 comparison graph between them and the actual values was constructed, as shown in Fig. 16 and Fig. 17.

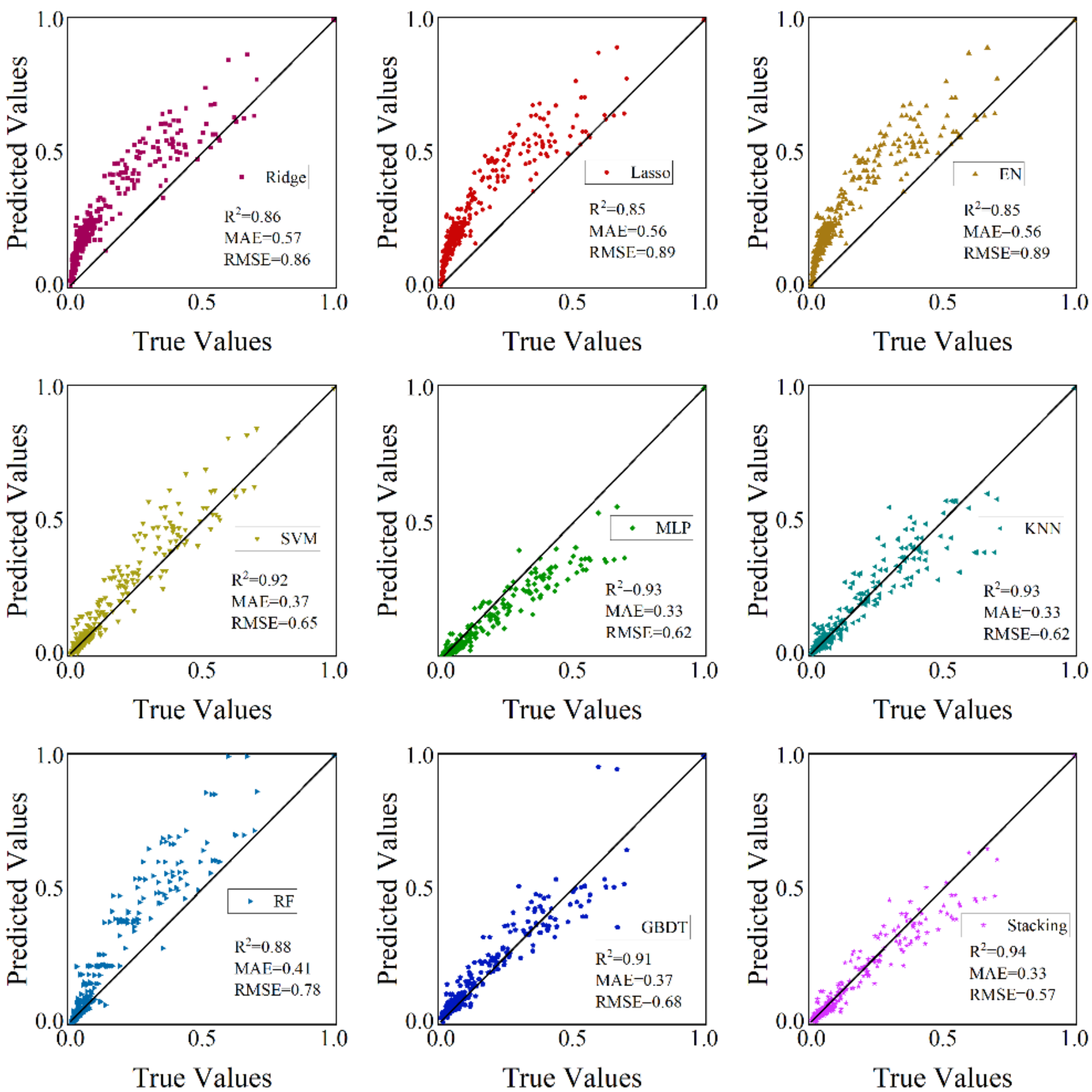

Fig. 16 - The 1:1 plot of predicted versus actual values for the nine models when the input features of the models are the geometric shape of the leaves versus the height and stem thickness of the corresponding individual plants

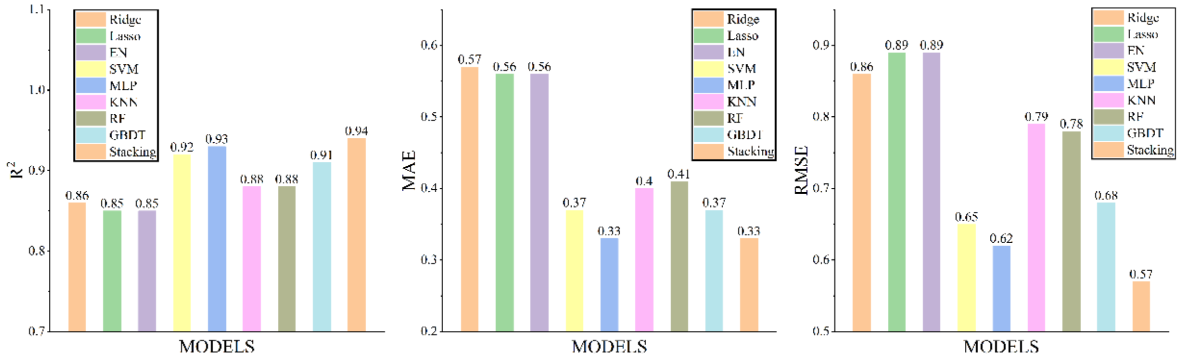

Fig. 17 - Comparison of the evaluation metrics of the nine models when the input features of the models are the geometric form of the fruit with the height and stem thickness of the corresponding individual plants

The leaf biomass prediction model constructed based on stacking had R2=0.65, MAE=0.79 and RMSE=1.27, whereas the evaluation indexes of the leaf biomass prediction model constructed based on stacking improved to a certain extent after adding the height and stem thickness of the current individual tomato plants to the input features. Compared with the model without the addition of height and stem thickness, the R2 increased by 0.21, MAE decreased by 0.3 and RMSE decreased by 0.46. The specific values were R2=0.86, MAE=0.49, and RMSE=0.81. Similarly, the fruit biomass prediction model was constructed based on stacking, with R2=0.85, MAE=0.59, and RMSE=0.9. After adding the plant height and stem thickness of the current tomato plant individuals to the input features, the evaluation indexes of the fruit biomass prediction model constructed based on stacking were improved to a certain extent, with an increase of 0.09 in the R2, a decrease of 0.26 in the MAE, and a decrease of 0.33 in the RMSE compared with that of the model with no addition of the plant height and stem thickness, with the specific values of R2=0.94, MAE=0.33, and RMSE=0.57. It is evident that the model's prediction accuracy can be effectively improved after adding the corresponding plant height and stem thickness of individual plants.

Model validation

The predicted and measured values of different parts of tomato leaves and fruits at various periods were compared to assess the model's generalisation ability.

Fig. 18 shows the predicted values of leaves biomass at each growth cycle of the four positions compared with the measured values, and the results of section 9 were better, with RMSE=0.1473 g and MAE=0.1072 g between its predicted and actual values. The RMSE between the predicted and measured values of biomass of leaves during each growth cycle ranged from 0.1473 g to 0.5229 g; MAE ranged from 0.1072 g to 0.4190 g. By comparing the predicted and measured values of leaf biomass, it was found that the predicted values might be larger or smaller than the measured values. The predicted results of leaf biomass fluctuated wildly, and the reason for the significant fluctuation of leaf error might be the large error mixed in the data collection process of the experiment.

Fig. 19 shows the predicted values of fruits biomass at each growth cycle of the four positions compared with the measured values, and the results of section fruit 3-2 were better, with RMSE=0.1206 g and MAE=0.0987 g between its predicted and actual values. The RMSE between the predicted and measured values of biomass of fruits during each growth cycle ranged from 0.1206 g to 0.4113 g; MAE ranged from 0.0987 g to 0.2965 g. By comparing the predicted and measured values of fruit biomass, it was found that the predicted values might be larger or smaller than the actual values. However, the predicted results of fruit biomass fluctuated less compared with the results of leaf biomass, which might be due to the smaller samples of fruits.

Machine learning algorithms are highly flexible and computationally efficient and have been widely used for modeling and prediction of agricultural scenarios (*Ribeiro et al., 2022; Saleem et al.,2021*). The input factors selected in this study were all phenotypic indicators of individual tomato plants and did not include environmental factors. If the variables including environmental factors, location of planting area, and agronomic and management practices are coupled into the simulation of the biomass prediction model, the performance, accuracy, and generalization ability of the model will be greatly improved (*Geng et al., 2021*).

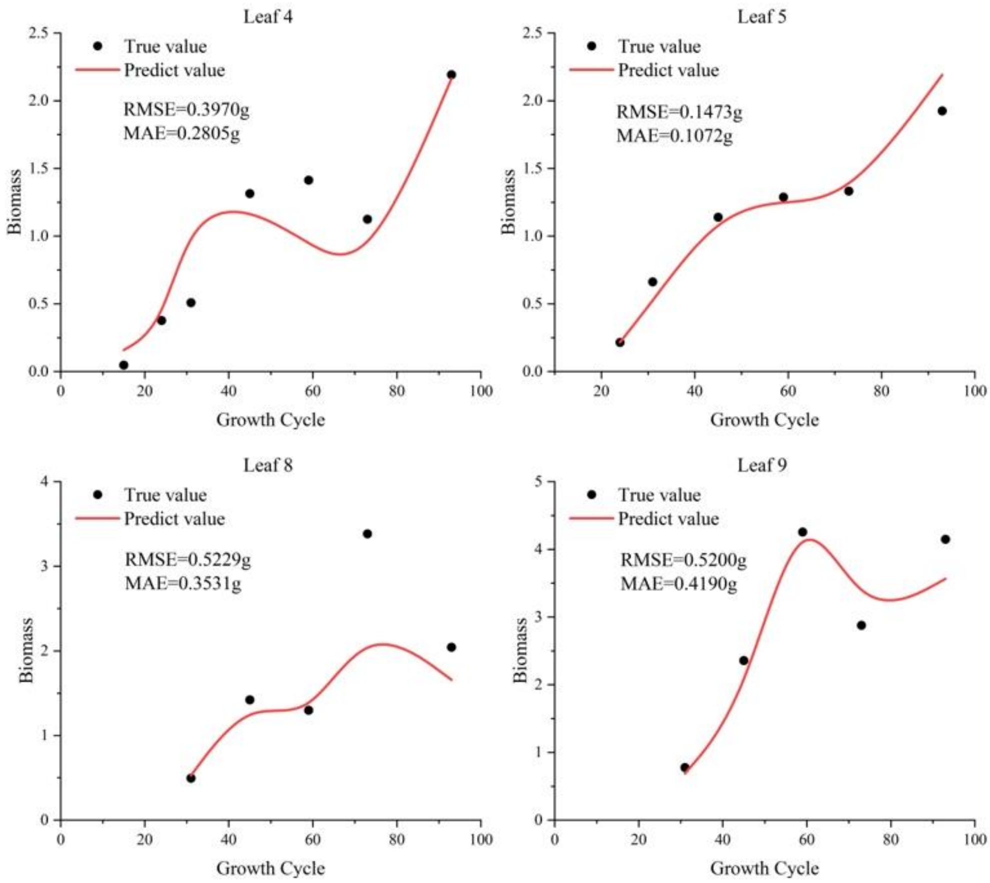

Fig. 18 - Comparison of predict and measured values of leaf biomass at different locations

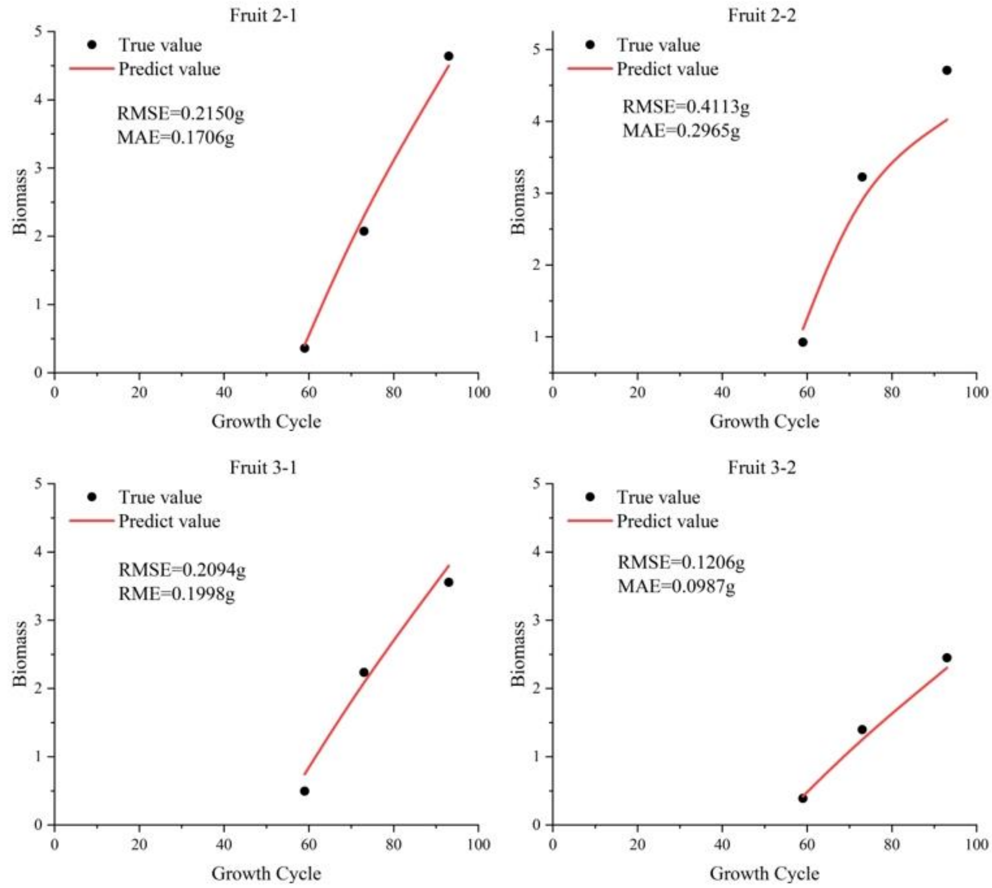

Fig. 19 - Comparison of predict and measured values of fruit biomass at different locations

CONCLUSIONS

In this study, tomato is taken as the research object, tomato organ morphology data were used as variables, morphological factors such as plant height and stem thickness were superimposed, and the biomass prediction model of tomato under organ scale is established based on machine learning and integrated learning, which specifies the optimal model construction method under organ scale and accurately realizes the biomass prediction of leaves and fruits of tomato.

The model is user-friendly and cost-effective. Farmers can anticipate fruit and leaf biomass by simply collecting geometrical-morphological data on tomato fruits and leaves using a measuring tool. The estimated cost of our test mainly includes measuring tools, ovens, computers, and other equipment, with a total of approximately 15,000 RMB. So that farmers can understand the growth status of the crop in real time, adjust cultivation and management measures and plant protection methods in a timely manner, and manage the tomato crop in a targeted manner to maximize yield and product quality.

Of course, this study still has the following shortcomings: first, the duration of the experiment needed to be shorter, and the subsequent tomato experiment needed to be conducted over a more extended period to improve the model's accuracy and generalization ability with more data. Second, the experiments were conducted only on tomatoes grown in solar greenhouses in the alpine region of the North China Plateau and did not consider other crops (e.g., cucumbers and eggplants, etc.), other growing regions (e.g., lamps in the North China Plateau), and other growing facilities (e.g., glass greenhouses), or field crops, and will be followed up with a more detailed experimental planning and experimental design to increase the abundance of the model's adaptability and the breadth of its use. Third, the acquisition of experimental data is manually obtained by hand, which has a certain degree of subjectivity and instability. This study did not consider the effects of pests and diseases, root systems, and soil on the whole process of tomato growth and development. In the following study, the effects of different environmental factors, agronomic measures and management conditions should be explored as well as the effects of pests and diseases on the biomass prediction model. Therefore, related research should be strengthened and it should be strived to construct more accurate and reliable biomass prediction models and to apply them to actual agricultural production to promote the progress of agricultural science and provide scientific basis and technical support for optimizing agricultural production.

ACKNOWLEDGEMENT

This research, titled ‘Construction and validation of a predictive model for tomato organ biomass at organ scale based on stacking learning’, was funded by the Shanxi Agricultural University Special Merit Program (XDHZFQY2022-02) and the Basic Research Program of Shanxi Provincial Science and Technology Department (202103021224123).

REFERENCES

[1] Colomina I., Molina P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: a review. *ISPRS Journal of Photogrammetry and Remote Sensing*. *92*, 79-97.

[2] Damtew A. (2017). Review on Genetics and Breeding of Tomato (*Lycopersicon esculentum Mill*). *Advances in Crop Science & Technology*, *05*(05), 306.

[3] Yuanyuan Fu, Guijun Yang, Xiaoyu Song, Zhenhong Li, Xingang Xu, Haikuan Feng, Chunjiang Zhao. (2021). Improved Estimation of Winter Wheat Aboveground Biomass Using Multiscale Textures Extracted from UAV-Based Digital Images and Hyperspectral Feature Analysis. *Remote Sensing*. 13(4), 581.

[4] Liying Geng, Tao Che, MingGuo Ma., Junlei Tan, HaiBo Wang. (2021). Corn Biomass Estimation by Integrating Remote Sensing and Long-Term Observation Data Based on Machine Learning Techniques. *Remote Sensing*. 13(12), 2352.

[5] Jahnavi Y., Elango P., Raja S.P., Nagendra Kumar P. (2023). A Novel Ensemble Stacking Classification of Genetic Variations Using Machine Learning Algorithms. *International Journal of Image and Graphics*. (02), 2350015.

[6] Ji Wen Cheng, Shusen Wang, Yuyang Luo, Cen Zhang. (2022). Prediction model of above-ground biomass of Salix based on BP neural network (基于 BP 神经网络的沙柳地上生物量预测模型), *Forestry Science Research*. (003), 035.

[7] Qi Liu, Yaxin Wang, Jie Yang, Wuping Zhang, Huanchen Wang, Fuzhong Li, Guofang Wang, Yuansen Huo, Jiwan Han. (2023). Smartagb: Aboveground Biomass Estimation of Sorghum Based on Spatial

Resolution, Machine Learning, and Vegetation Index. *EAI Endorsed Transactions on Internet of Things*. *9(1)*, e1.

[8] Lingzhi Li, (2013), Quantitative study of tomato growth with different nitrogen supply levels based on functional structure feedback mechanism (基于功能结构反馈机制下番茄生长对不同供氮水平定量化研究), [Doctoral dissertation, Shanxi Agricultural University]

[9] Maddaloni P., Continanza D.N., Del Monaco A., Figoli D., Di Lucido M., Quarta F., Turturiello G. (2022). Stacking machine-learning models for anomaly detection: comparing AnaCredit to other banking datasets. *Bank of Italy Occasional Paper*. (689).

[10] Qiaoxue Dong, Yiming Wang, Jialin Hou. (2007), Tomato structural-functional model: Organ-based functional model and validation (番茄的结构-功能模型Ⅱ:基于器官水平的功能模型与验证研究), *Chinese Journal of Ecological Agriculture*. (01),122-126.

[11] Junior F.M.R., Bianchi R.A., Prati R.C., Kolehmainen K., Soininen J.P., Kamienski C.A. (2022). Data reduction based on machine learning algorithms for fog computing in IoT smart agriculture. *Biosystems Engineering*. 223, 142-158.

[12] Saleem M.H., Potgieter J., Arif K.M. (2021). Automation in Agriculture by Machine and Deep Learning Techniques: A Review of Recent Developments. *Precision Agriculture*. 22(6), 2053-2091.

[13] Weifeng Wang, Yuancai Lei, Xuefeng Wang, Haoyan Zhao. (2008). Review of forest biomass models (森林生物量模型综述), Journal of Northwest Forestry College. 23(2), 58-63.

[14] Xu Wang, Yushuai Wang, Xuemeng Lian, Yuguang Wang, Lihua Yu, Gui Geng, (2022), Inversion of chlorophyll content in sugar beet canopy based on UAV multispectral technique (基于无人机多光谱技术的甜菜冠层叶绿素含量反演). *Sugar Crops of China*. 44(4), 36-42.

[15] Zhong Xin Chen, Jianqiang Chen, Huajun Tang, Yun Shi, Peileng, Jia Liu, Limin Wang, Wenbin Wang, Yanmin Yao, (2016), Progress and perspectives on agricultural remote sensing research and applications in China (农业遥感研究应用进展与展望), *Journal of Remote Sensing*. 20(5), 748-767.

[16] Yang Liu, Jue Huang, Qian Sun, Haikuan Feng, Guijun Yang, Fuqin Yang. (2021). Estimation of aboveground biomass of potato based on UAV digital image (利用无人机数码影像估算马铃薯地上生物量), *National Remote Sensing Bulletin*. 25(9), 2004-2014.

[17] Zhaoying Mu, Zipeng Zhang, Hao Zhang, Lichun Jiang. (2024). Applying Machine Learning Algorithm Models to Predict Aboveground Biomass of Larix gmelinii in Xing'an (应用机器学习算法模型预测兴安落叶松地上生物量), *Journal of Northeast Forestry University*. 52(03), 41-47.