

THE DESIGN OF JUJUBE IRRIGATION SYSTEM USING LINEAR REGRESSION ANALYSIS, BP NEURAL NETWORK AND RANDOM FOREST

利用线性回归分析、BP神经网络与随机森林的枣树灌溉系统设计

Wenhao DOU, Sanmin SUN ^{*}, Pengxiang XU ¹

College of Water Conservancy and Architecture Engineering, Tarim University, Alar, Xinjiang / China

Tel: +8609974680383; E-mail: ssmaq@126.com

Corresponding author: Sanmin Sun

DOI: <https://doi.org/10.35633/inmateh-70-16>

Keywords: Random Forest, BP neural network, Linear regression analysis, Intelligent irrigation system

ABSTRACT

This paper evaluates linear regression analysis, BP neural network, and a random forest prediction model for the prediction of jujube water demand. The results highlight that the R^2 of the random forest is 0.941 and the residual distribution is the most stable. Hence, the random forest is more suitable for prediction, and therefore, an intelligent irrigation system is established employing random forest, where the cloud server is the upper computer and a Raspberry Pi is the lower computer, and at the same time, a PC and a mobile interface was built to present various information about the developed irrigation system.

摘要

本文搭建线性回归分析、BP神经网络与随机森林预测模型预测枣树需水量，结果表明随机森林的 R^2 为 0.941 且残差分布最稳定，随机森林更适合用于预测。利用随机森林建立一套智能灌溉系统，使用云服务器作为上位机，树莓派作为下位机，同时搭建 PC 端与移动端操作页面。

INTRODUCTION

Water demand data is important for jujube growth, with water demand prediction being a key link to achieving intelligent irrigation, and prediction accuracy directly affecting the water-saving effect. Hence, many experts and scholars apply computer technology to predict water demand to realize intelligent agricultural water-saving (Friha et al., 2021).

With the maturity of artificial intelligence technology, more intelligent algorithms are applied to predict crop water demand (Liu et al., 2021). Given that linear regression analysis, BP neural network, and random forest have extensively been used and proven effective in data analysis (Abdol et al., 2020), these three prediction models on water prediction demand for jujube trees were evaluated.

Considering agricultural intelligent irrigation systems, the one based on ZigBee is unsuitable for building sensors in a wide range of farmland due to its small networking range (Emad Ahmed Mohammed et al., 2023). Additionally, the intelligent irrigation control system based on NB IoT has a wide coverage and much-connected equipment but requires a very high network (Cheng et al., 2021; Aqeel-Ur-Rehman et al., 2011). The LoRa communication system has wide coverage, low energy consumption, and no requirements for the network. Therefore, the developed system uses LoRa communication for data transmission.

Currently, the core controller of most intelligent irrigation systems is a single-chip computer. The development cycle of SCM is quick, but the running speed is slow, the resources are few, and the update code must be rewritten, which is relatively troublesome. Therefore, a new generation of intelligent irrigation systems with Raspberry Pi as the core controller has emerged (Zhao et al., 2022; Tan et al., 2019). Raspberry Pi is a microcomputer that uses an ARM core processor to run the Linux operating system. It can directly program and run programs locally, and after connecting to the Internet, it can directly perform remote operations and complete operations that a microcontroller cannot complete.

Spurred by the above findings, this paper evaluates linear regression analysis, BP neural network, and a random forest prediction model and selects the best model as the irrigation system prediction model. Given the experimental results, an intelligent control system for agricultural irrigation is constructed using Raspberry Pi as the core controller and LoRa communication as the data transmission system.

¹ Dou Wenhao, M.S. Stud. Eng.; Sun Sanmin, Prof. Ph.D. Eng.; Xu Pengxiang, M.S. Stud. Eng.

MATERIALS AND METHODS

System architecture

For this intelligent irrigation system, the current sensor technology, Internet of Things technology, wireless communication technology, and artificial intelligence technology were exploited. The developed system comprises a built data acquisition module, control module, irrigation module, cloud server, and online control platform. Solar modules power the system.

Specifically, wind speed, solar radiation, soil temperature, humidity, and other sensors are used to collect various environmental information in the jujube garden. The data are then transmitted via the LoRa communication system, and the Raspberry Pi is the control module that transmits data and provides the appropriate instructions. Besides, the irrigation module comprises a water pump, solenoid valve, and flowmeter and employs LoRa communication technology for data transmission. Additionally, the PC terminal and mobile terminal pages are designed for viewing various sensor data from the jujube garden and controlling the solenoid valve for irrigation. The system can conduct real-time environmental monitoring, data collection, transmission and analysis, water demand prediction, and remote irrigation control. The developed irrigation system is illustrated in Fig.1.

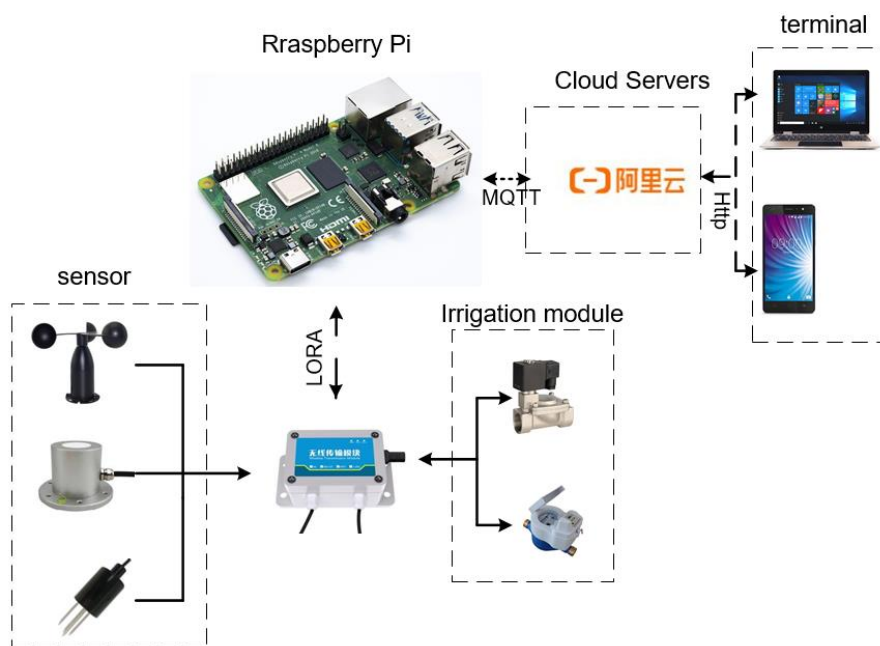


Fig. 1 - System Framework

During irrigation, the system first collects data through the data acquisition module and transmits it to the control module through LoRa communication. Then, the control module uploads the data to the online control page and the cloud server and makes the irrigation decisions, which are passed to the irrigation module for implementation. The actual irrigation system is illustrated in Fig.2.

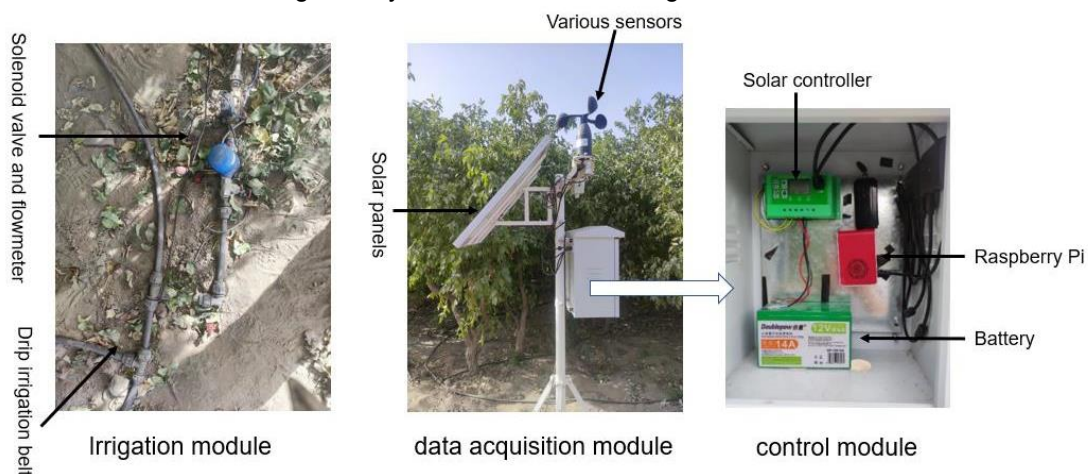


Fig. 2 - Physical system

Irrigation prediction model

The data are independently predicted according to linear regression analysis, BP neural network, and random forest, and the best model is selected for the intelligent irrigation system. The sample data were obtained from various sensors in the Zaoyuan Irrigation Experimental Base of the School of Water Conservancy and Building Engineering of Tarim University in Alar from 2001 to 2021. From the sample data, 80% is allocated to the training set and 20% to the verification set. Examples of raw data are listed in Table 1.

Table 1

Raw data							
date	Average temperature X1 (°C)	Average wind speed X2 (m/s)	relative humidity X3(%)	Sunshine hours X4(h)	Pressure X5(Pa)	solar radiation X6(W/m2)	water demand Y(mm)
20010401	17.38	3.62	18.66	6.62	895.2	177.09	4.13
20010402	16.67	4.69	20.53	5.43	893.7	144.21	4.01
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
20210929	20.31	1.64	19.45	5.21	895.5	143.82	2.98
20210930	22.11	2.28	16.49	5.11	895.7	140.90	3.49

Linear regression analysis

The traditional linear model has a simple structure and a small computational burden. When the time series are in a stable state, the linear model produces good predictions and captures the linear relationship of the time series (Huang et al., 2022).

The regression model considers the water demand of jujube trees as the dependent variable Y in the regression analysis, and the other meteorological factors are the independent variable Xi.

The equation and analysis results are as follows:

$$Y=0.316X_1+0.768X_2+0.023X_3+3.147X_4-0.037X_5-0.093X_6+24.618 \tag{1}$$

Table 2

Results of linear regression analysis								
variable	Denormalization coefficient		Standardization coefficient		T	P	R ²	F
	B	Standard error	Beta					
constant	24.618	4.136	—		5.95	0		
X ₁	0.316	0.006	0.65		52.99	0		
X ₂	0.768	0.019	0.34		39.55	0		
X ₃	0.023	0.002	0.13		11.93	0	0.857	F=1987.2
X ₄	3.147	0.182	1.819		17.28	0		
X ₅	-0.037	0.005	-0.08		-8.08	0		
X ₆	-0.093	0.007	-1.49		-14.14	0		

Table 2 reveals that the F value of the regression equation is 1987.2 and the P value of the model is less than 0.01, indicating that the multiple linear regression equation is very significant. However, the fitting effect of R² is 0.857, which is not adequate. To further improve the prediction effect, BP neural network and random forest will be used.

BP neural network

A BP neural network is a multi-layer feedforward network with a hidden layer, which is the core of the feedforward network and reflects the essence of an artificial neural network. BP neural network systems solve the learning problem of hidden unit connection weight in multi-layer networks. A typical BP neural network model is depicted in Fig.3, where the input layer is X1~X4, and Y is the output vector.

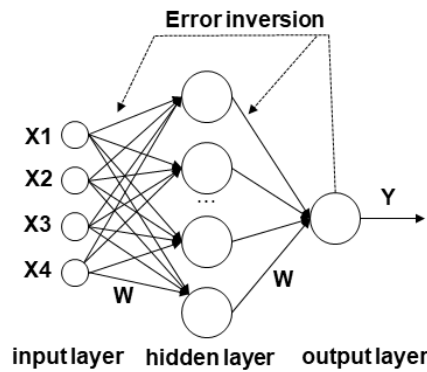


Fig. 3- BP Neural Network Topology

This paper's BP neural network prediction model employs the LM training algorithm. The sample data in this paper comprise 3843 groups of medium-scale data, and the LM algorithm has the fastest training speed. Moreover, 'tansig' is selected as the transfer function, 'trainlm' as the training function, and 'learnqdm' as the learning function. The number of iterations is set to 1000, the learning rate is 0.1, and the error precision is 0.001. The hidden layer is determined as 11, according to the determination method of 2N+1 (Lv et al., 2019).

The results highlight that the deterministic parameter R^2 of the BP neural network prediction model is 0.849, indicating that the model's prediction accuracy is insufficient. Table 3 shows that the model's MSE value has a large error, with the training effect not meeting the actual demand.

Table 3

BP Neural network analysis results			
MSE	RMSE	MAE	R^2
1.281	1.132	0.926	0.849

Stochastic forest model

The random forest exploits a decision tree as the basic learner through Bootstrap technology. It randomly extracts k samples from the sample set N, with the samples replaced in N. The sample size in the random extraction is the same as the original data and establishes k decision tree models for k samples. Each time a training set is used to obtain a model, the final decision tree model will provide k decision results, and the final results will be voted according to each record. Fig.4 illustrates the principle of a random forest, where D is the sample set, and D1, D2 ... Dk are the decision trees formed by random sampling.

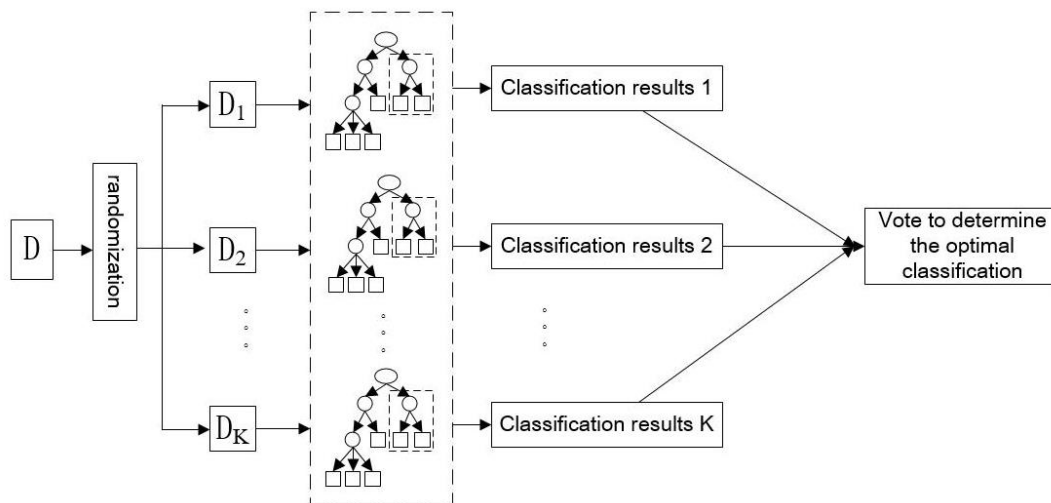


Fig. 4 - Principle of random forest

According to the data sample size, the number of decision trees in the random forest model is set to 100, the maximum depth of the decision tree is 10, and the maximum number of leaf nodes is 50. In this work, sampling with placement is performed. After importing the training data into the model, the corresponding results are reported in Table 1, which reveals that the deterministic parameter R^2 is 0.928, i.e., the model training effect is appealing and meets the requirements of daily use.

Table 4

Random Forest Analysis Results			
MSE	RMSE	MAE	R ²
0.359	0.6	0.45	0.928

Comparison and analysis of prediction models

This paper considers the jujube garden as an example to analyze and compare the application of linear regression analysis, BP neural network, and random forest prediction model in water demand. After training the model, 615 groups of sample data are randomly selected from the 3843 groups of experimental data to predict the water demand of jujube trees using the three competitor models, with Table 5 presenting the actual and predicted water demand per model.

Table 5

Serial No	Actual water demand / (mm·d ⁻¹)	Predicted water demand / (mm·d ⁻¹)		
		Linear regression analysis	BP neural network	Random forest
1	4.1340	4.8142	4.7200	4.7825
2	4.0755	4.8334	4.7014	4.3487
⋮	⋮	⋮	⋮	⋮
614	6.9904	5.9874	5.8915	6.0657
615	6.5048	5.3346	5.1512	5.6568

The prediction results are displayed, and the model's predicted values are linearly fitted with the real values. Fig.5 compares the residual error of the fitting curve between some predicted and real water demand values based on linear regression analysis, BP neural network, and random forest prediction model.

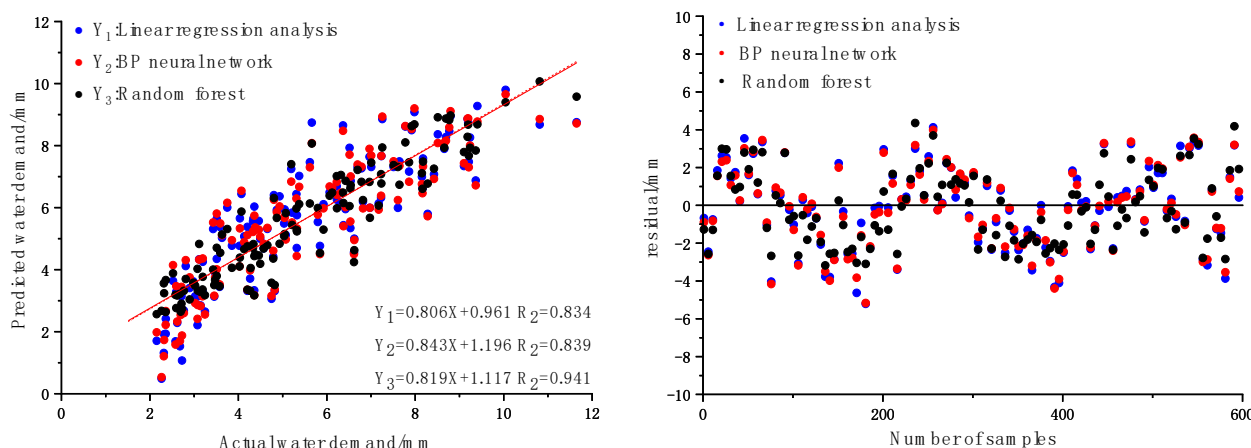


Fig. 5 - Fitting curve and residual

Fig. 5 highlights that the predicted value using the linear regression equation and the actual value have a fit of 0.839. For the BP neural network, the fitting value is 0.834, and for the random forest, the fitting value is 0.941. Considering the residuals, the residuals of the linear regression equation and the BP neural network are relatively large and volatile, while the residuals of the random forest are small and stable. Compared with linear regression analysis and BP neural network, the random forest regression has a better prediction effect on the water demand of jujube trees. Thus, the random forest has higher accuracy and is more suitable for the water storage prediction of jujube trees.

Irrigation control platform

The Alibaba Cloud is utilized for PC and mobile control platforms to facilitate remote control and real-time irrigation system monitoring. Our platform interface displays environmental parameters, sensor status, irrigation control, and equipment positioning. Specifically, the environmental monitoring part displays the real-time environmental data in the orchard, which users can view in real-time. The sensor status part displays the operational status of the sensor to find damaged offline equipment quickly.

The irrigation control allows users to query the crop water demand predicted by the model, realize the remote switch of the user control field solenoid valve, and display the irrigated water volume and the instantaneous flow of the drip irrigation belt through the flowmeter. Besides, the device location displays the device location to prevent the loss of the device. At the same time, to overcome the problem that operators cannot use the PC interface to view farmland environmental information and control the solenoid valve irrigation, a mobile terminal operation interface was also built. The PC end control interface is shown in Figure 6.

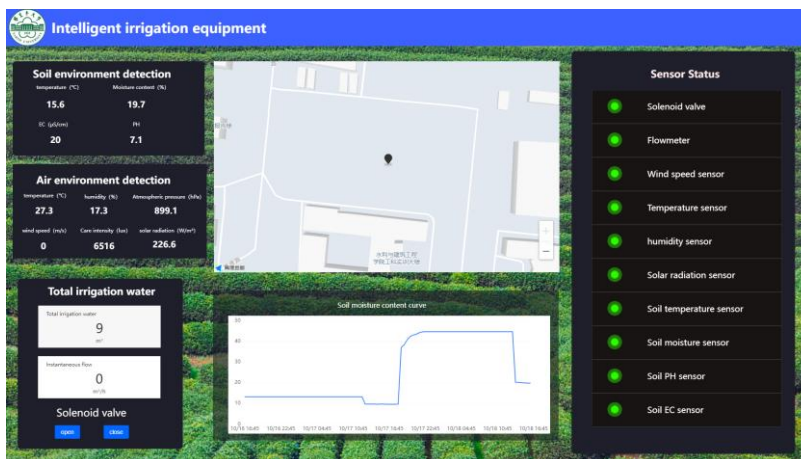


Fig. 6 - PC end control interface

The mobile end control interface is illustrated in Figure 7.



Fig. 7 - Mobile terminal control interface

RESULTS

Irrigation experiment

After the system is built, its parts are tested to confirm normal operation. Precisely, the data obtained by the sensors are compared with the data obtained by the weather station in the jujube garden. Moreover, it is tested that the operation interface can display data in real time and that the data transmission is normal. Additionally, it is confirmed that the PC and mobile interfaces can remotely control the solenoid valve, and the remote control switch of the solenoid valve operates normally. The abovementioned tests aim to confirm that the system achieves the expected effect.

The irrigation test site is located in the jujube garden of the irrigation test base of the School of Water Conservancy and Building Engineering of Tarim University in Alar City. The jujube trees are 10 years old, and drip irrigation has been adopted. The irrigation test environment is shown in Fig. 8.



Fig. 8 - The irrigation test environment

During irrigation, first, the Stacking integrated learning prediction model predicts the water demand according to the meteorological data collected by the sensors in the jujube garden. The predicted value is compared with the soil moisture storage calculated by the soil moisture sensor. When the predicted water demand exceeds the soil moisture storage, the system judges whether to conduct irrigation. The irrigation stops when the soil moisture sensor detects that the soil moisture storage exceeds the predicted water demand. Otherwise, irrigation will not be carried out.

The calculation formula of soil water storage is:

$$S = \rho b \times h \times w \tag{2}$$

where:

- S — Soil water storage, [mm];
- ρb — Soil bulk density, [g/cm³];
- h — Soil thickness, [mm];
- w — Soil moisture, [%];

In this experiment, the soil moisture sensor is buried at 20 cm, so $h=200$ mm. According to the previous experimental data, the soil bulk density of the test site is 1.43 g/cm³ (Zhou et al., 2021).

Considering our test, the soil water storage is calculated based on the data obtained from the soil humidity sensor, and the system irrigation results are reported in Table 6, revealing that the system operates normally and starts irrigation when the soil water storage is less than the predicted value. No irrigation will be carried out when the soil water storage exceeds the predicted value.

Table 6

Systematic irrigation test				
Number of tests	Soil water storage /mm	Model prediction value /mm	Solenoid valve status	
1	5.3	6.7	open	
2	5.9	6.6	open	
3	5.7	8.3	open	
4	7.1	6.5	close	
5	6.1	5.2	close	

CONCLUSIONS

This paper evaluates linear regression analysis, BP neural network, and random forest to predict the water demand of jujube trees. Considering the meteorological factors as independent variables and the jujube water demand as the dependent variable, the results reveal that the fitting coefficient R^2 of the random forest model is 0.972, which is better than the linear regression analysis and BP neural network. Additionally, the residual error of the random forest is stable. Overall, the random forest prediction model is highly accurate and can be used in actual production. At the same time, hardware such as the data acquisition module and the irrigation module and apply LoRa communication technology were developed to solve the problem of RS485 communication requiring a wired connection. Overall, the absolute error between the model-predicted water demand and the actual crop water demand is small, demonstrating that all modules of the irrigation system operate normally in daily production, and the whole system operates stably.

ACKNOWLEDGEMENT

The authors were funded for this project by the science and technology project of Xinjiang Production and Construction Corps (No.2021CB021), and science and technology plan project of the first Division of the City of Alar (No.2022XX01).

REFERENCES

- [1] Abdol R. Z., Mohammad R. M., & Ali S. (2021). Investigating of the climatic parameters effectiveness rate on barley water requirement using the random forest algorithm, Bayesian multiple linear regression and cross-correlation function. *Paddy and Water Environment*, Vol. 19, pp. 137-148, Germany.
- [2] Aqeel-Ur-Rehman, Abbasi A.Z., Islam N., Shaikh Z.A., (2014). A review of wireless sensors and networks' applications in agriculture, *Comput. Stand. Interfac*, Vol.36, pp. 263–270, Pakistan.
- [3] Cheng Z., Xin Y., Yao Y., Hou C., (2021). Design of intelligent irrigation system based on NB IoT (基于 NB-IoT 的智能灌溉系统设计). *Integrated Circuit Applications*, Vol. 6, pp. 44-45, Shanghai/ China.
- [4] Emad Ahmed Mohammed., Hakam Marwan Zaidan, Zaid Ghanim Mohammed, (2023). Implementation of WSN based Smart Irrigation System. *Przeglad Elektrotechniczny*, Vol. 6, pp. 27-31, Mosul.
- [5] Friha O., Ferrag M.A., Shu L., Maglaras L.A., Wang X. (2021). Internet of things for the future of smart agriculture: a comprehensive survey of emerging technologies, *IEEE CAA J. Autom*, Vol. 8, pp. 718–752, Algeria.
- [6] Huang H., Zhang Z., Lin Z., Liu S., (2022) Hourly water demand forecasting using a hybrid model based on mind evolutionary algorithm. *Water Supply*, Vol. 1, pp. 917-927, United Kingdom.
- [7] Liu J., Liu X., Wu H., Deng H., Li Z. (2021). Application of support vector machine model based on GA optimization in forecasting water demand of green pepper crops (基于 GA 优化的支持向量机模型在青椒作物需水量预测中的应用). *Water Saving Irrigation*, Vol. 1, pp. 70-76, Hubei/China.
- [8] Lv H., Feng Qian., (2019). A review of stochastic forest algorithm research (随机森林算法研究综述). *Journal of Hebei Academy of Sciences*, Vol. 3, pp. 37-41, Hebei/China.
- [9] Tan Y., Qin F., (2019). Design and research of indoor intelligent irrigation system based on Raspberry Pi (基于 Raspberry Pi 的室内智能灌溉系统设计与研究). *Water saving irrigation*, Vol. 7, pp.105-108, Hubei/China.
- [10] Zhao S., Chen X., (2022). Intelligent irrigation system based on Raspberry Pi and cloud platform (基于树莓派和云平台的智能灌溉系统). *Computer System Application*, Vol. 4, pp.123-129, Beijing/China.
- [11] Zhou S., Sun S., Yao B., Li F., He T., Wang W., (2021). Effects of mist irrigation on canopy environment, photosynthetic characteristics, yield and quality of jujube orchards (弥雾灌对枣园冠层环境和光合特性及产量品质的影响). *Journal of Agricultural Machinery*, Vol. 5, pp. 249-257, Beijing/China.