Research Article

Tao Song, Yulong Si, Jie Gao, Wei Wang*, Congwei Nie, and Jiří Jaromír Klemeš

Prediction and monitoring model for farmland environmental system using soil sensor and neural network algorithm

https://doi.org/10.1515/phys-2022-0224
received September 22, 2022; accepted January 02, 2023

Abstract: In this study, data fusion algorithm is used to classify the soil species and calibrate the soil humidity sensor, and by using edge computing and a wireless sensor network, farmland environment monitoring system with a two-stage calibration function of frequency domain reflectometer (FDR) is established. Edge computing is used in system nodes, including the saturation value of the soil humidity sensor, the calculated soil hardness, the calculation process of the neural network, and the model of soil classification. A bagged tree is adopted to avoid over-fitting to reduce the prediction variance of the decision tree. A decision tree model is established on each training set, and the C4.5 algorithm is adopted to construct each decision tree. After primary calibration, the root mean squared error (RMSE) between the measured and standard values is reduced to less than 0.0849%. The mean squared error (MSE) and mean absolute error (MAE) are reduced to less than 0.7208 and 0.6929%. The bagged tree model and backpropagation neural network are used to classify the soil and train the dynamic soil dataset. The output of the trained neural network is closer to the actual soil humidity than that of the FDR soil humidity sensor. The MAE, the MSE, and the RMSE decrease by 1.37%, 3.79, and 1.86%. With accurate measurements of soil humidity, this research shows an important guiding significance for improving the utilization efficiency of agricultural water, saving agricultural water, and formulating the crop irrigation process.

Keywords: humidity sensor, sensor calibration, backpropagation, bagged tree, neural network

1 Introduction

Abuse of earth resources results in unprecedented environmental crises. The development of resource utilization mainly focuses on non-renewable fossil resources and mineral resources currently, such as the replacement of conventional fluid with nanofluids [1], energy storage with nanocomposite phase change materials [2], and increase in heat transfer with magnetic fields [3]. Plate heat exchangers [4] and double tube heat exchangers [5] with novel structures have been introduced in distributed energy systems for energy saving. However, water resources have received less attention than fossil and mineral resources. Water resource is indispensable for social and economic development [6]. About 85% of the global freshwater is used for agricultural irrigation [7], and about 53% of Chinese freshwater resource is used for agriculture applications [8]. Only 40–50% of agricultural water in developing countries has been effectively utilized [9], and the utilization efficiency of agricultural water is 70–80% in developed countries [10]. The accurate measurement of soil humidity is of great significance for improving the utilization efficiency of agricultural water, saving agricultural water, and formulating the crop irrigation process [11].

The frequency domain reflectometer (FDR) method is commonly used to measure soil humidity [12]. The FDR method is used to obtain soil humidity by measuring the dielectric constant of the soil [13]. As the FDR soil humidity sensor is buried in soil for a long time [14], it is aged or corroded due to soil texture and other factors. The data

* **Corresponding author: Wei Wang,** Innovation and Entrepreneurial Center, Hebei University of Technology, Tianjin 300401, China, e-mail: 2013084@hebut.edu.cn
Tao Song, Yulong Si, Jie Gao: School of Electronic Information Engineering, Hebei University of Technology, National Demonstration Center for Experimental (Electronic and Communication Engineering) Education, Hebei University of Technology, Tianjin 300401, China
Congwei Nie: Information Technology Department, Hebei GongNuo Testing Technology Co., Ltd, Shijiazhuang 050200, China
Jiří Jaromír Klemeš: Sustainable Process Integration Laboratory – SPIL, NETME Centre, Faculty of Mechanical Engineering, Brno University of Technology – VUT Brno, Technická 2896/2, 616 69 Brno, Czech Republic

drift occurs due to the influence of soil temperature and soil hardness [15]. In addition, the physical properties of soil are affected by nanoparticles due to the application of nano-agriculture [16].

First, linear data drift over the full humidity range is caused by aged and corroded sensors. According to this feature, the data characteristics measured by standard sensors are recorded first [17]. Then, the offset value is calculated according to the data characteristics after aging or corrosion. Second, nonlinear drift in the output of sensors is caused because of the difference in soil temperature, hardness, and soil type [18]. Based on the characteristics of nonlinear data drift of sensors [19], the FDR soil humidity sensor shows the soil humidity output value [20], to establish the neural network relation among soil temperature, soil hardness, and the actual soil humidity value [21].

A novel soil classification method based on data fusion of soil environmental parameters is proposed. A two-stage calibration method is designed to improve the measurement accuracy of the FDR soil humidity sensor. A farmland environment monitoring system with a two-stage calibration function of the FDR sensor is established using edge computing and a wireless sensor network.

2 Experimental investigations

In order to verify a linear rule of data drift due to aging and corrosion problems of FDR soil humidity sensors, the soil samples were collected on the Beichen Campus of Hebei University of Technology in Tianjin, China.

2.1 Primary calibration of FDR soil humidity sensor

The standard FDR soil humidity sensor marked A was accurate, and the other sensors marked B and C were used for 1 year and 2 years, respectively. In order to verify the linear drift characteristics of the sensors, three sensors were used for the experimental analysis of the soil at the same humidity [22].

Three sensors were buried in an even array in a plastic vessel, which was filled with dry soil samples. Add water evenly, so that the soil humidity inside the vessel gradually increases to the maximum, and a full humidity range was measured simultaneously with these three sensors. Test results are presented in Figure 1. It is found that the growth rate of the sensor output is slowing down with the increase in water injection. When its output reaches the saturation of soil humidity sensors, the output values of the sensors no longer increase with the amount of water injected. The output saturation values of sensors A, B, and C are 46.63, 55.79, and 58.93%. Sensors B and C, which are buried in soil for a few years, provide a large data offset than the standard sensor A. This part of the data offset is linear, and the curves show a larger overall value within the whole range of humidity.

According to the characteristics of sensors, the sensor with numerical fault can be calibrated by data compensation. As long as the output values of the standard sensor and the error sensor under a different soil humidity are measured, the data compensation can be carried out according to the difference.

In the actual greenhouse or farmland, it is difficult to use standard sensors and non-standard sensors to

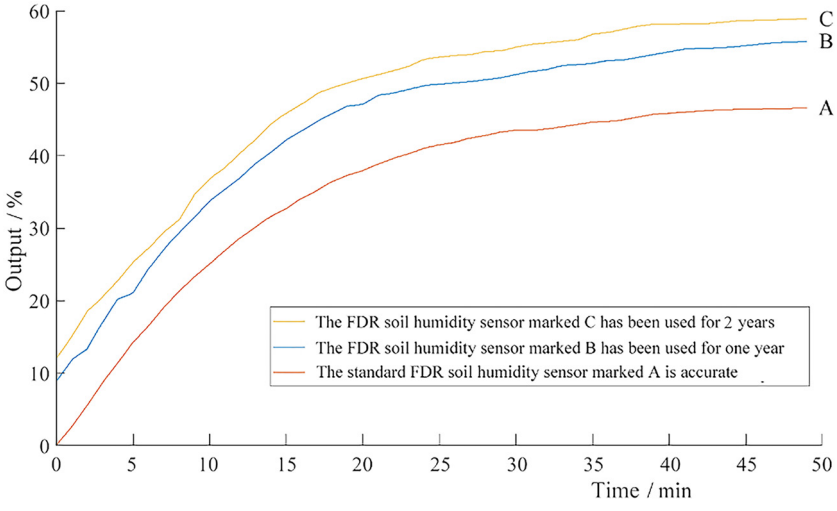

Figure 1: Relation curves between sensor outputs and water injection time.

measure the output data under the same soil humidity conditions without manual operation [23]. The data compensation is carried out by measuring the saturation process curves of the non-standard soil sensors and the standard soil sensor. The soil humidity sensor will reach its output saturation value during irrigation time or rainy season, which corresponds to the highest points of sensors A and B as shown in Figure 1. The compensation value for the linear data is shown in Eq. (1) as follows:

$$V = VS_i - VS_n, \tag{1}$$

where V represents the compensation value for linear data, VS_i represents the saturation value from a non-standard humidity sensor, and VS_n represents the saturation value from the standard soil humidity sensor.

Aging and corrosion on humidity sensors are slow processes. The compensation value V is calculated according to Eq. (1) after the output saturation value of the non-standard sensor is monitored by the agricultural personnel for irrigation or in the rainy season. The sensor is calibrated according to the following equation:

$$E = M - V, \tag{2}$$

where E represents the calibrated soil humidity, M represents measured values from the non-standard soil humidity sensor, and V is the linear data compensation.

In Eq. (1), the output saturation value of the standard soil moisture sensor is unique and measured by the standard soil moisture sensor, while the output saturation value of each non-standard soil moisture sensor is different. The compensation value for linear data of each non-standard soil moisture sensor is obtained by subtracting the two. For a specific non-standard sensor, Eq. (2) is used to subtract the measured value from the compensation value of linear data to carry out data compensation for the non-standard soil moisture sensor. The data of the non-standard soil moisture sensor that is closer to the actual soil moisture after linear data compensation are obtained, namely, the E value.

2.2 Secondary calibration of FDR soil humidity sensor

Because only the linear compensation of the sensor cannot meet the requirements of agricultural utilizations, it is necessary to carry out the secondary calibration. Soil distributed in China is roughly divided into loam, brown soil, and sandy soil. Three soil samples were used in this experiment. For the same soil humidity in actual farmland, the FDR humidity sensor is affected by soil temperature, soil hardness, and soil type [24]. In the secondary calibration, the data set of corresponding relationships among soil temperature, soil hardness, soil type, output values of the humidity sensor, and actual soil humidity was obtained through many experiments. First, the soil was classified by the relationship among soil type, soil temperature, soil hardness, and output humidity of the FDR soil sensor after the primary calibration [25]. The backpropagation (BP) neural network algorithm was used to train the data set to get the characteristics of the neural network [26]. According to the network characteristics, the actual soil humidity value was obtained by inputting soil temperature, soil hardness, and output humidity of the FDR soil sensor.

In the experiment, the temperature range was from 10 to 50°C and the temperature gradient was 1°C. The soil hardness means the amount of soil pressure per cm^2. The pressure value was measured by pressure sensors, and the surface area of the sensor-sensitive element was measured. The soil hardness is calculated by Eq. (3) as follows:

$$H = F/S, \tag{3}$$

where H represents the soil hardness, F represents the pressure value measured by the pressure sensor, and S represents the surface area of the sensor-sensitive element. The suitable range of soil hardness for crop growth in this experiment was 20–80 N/cm^2.

A number of soil samples were thoroughly dried and divided into 24 parts, and each piece weighed 2,400 g marked m_0. Each sample of soil was placed in a plastic cup. The 24 cups were labeled as $Ai \sim Fi$ (i is 1, 2, 3, and 4). An FDR soil humidity sensor, a soil temperature sensor, and pressure sensors were buried in each cup. Table 1 shows the weight of soil in each plastic cup, the weight of water injected, and the soil humidity. The plastic cup was sealed with plastic wrap to prevent water volatilization in the soil, and water was distributed evenly in the soil for 24 h in the plastic containers. Every accurate soil humidity in 24 plastic containers was calculated by Eq. (4) as follows:

$$Hu = \frac{m_i}{m_0 + m_i}, \tag{4}$$

where Hu represents the actual soil humidity, m_i represents the weight of water added into the cup, and m_0 represents the weight of soil.

The weight of water injected into the soil was measured three times, and the average value was taken to calculate the soil moisture value after different weights of water were injected into the dry soil through Eq. (4).

Table 1: Labels and soil humidity

Vessel number (i = 1, 2, 3, 4)		*Ai*	*Bi*	*Ci*	*Di*	*Ei*	*Fi*
Weight of water (unit: g)	The first measurement	125.3	262.8	429.1	608.3	800.7	1028.0
	The second measurement	126.2	264.8	423.0	600.3	796.3	1027.6
	The third measurement	126.5	273.4	419.9	591.4	803.0	1032.6
Average weight of water (unit: g)		126.0	267.0	424.0	600.0	800.0	1029.0
Soil humidity (unit: %)		5	10	15	20	25	30

After calculation, the soil moisture of six groups of containers was 5, 10, 15, 20, 25, and 30%, respectively. The output of the soil humidity sensor shows deviations from the actual soil humidity due to effects of soil temperature, hardness, and type. The secondary calibration of the FDR soil humidity sensor is used to eliminate the measurement deviations caused by soil temperature, soil hardness, and soil type. The measured results are improved using the secondary calibration, closer to the actual measurement accuracy.

The testing system was established to measure and store experimental data as shown in Figure 2.

The small experimental measurement system consisted of an FDR soil humidity sensor, soil temperature sensor, pressure sensor, data transmission unit, and cloud monitoring platform. The sensor whose output was deviated from the actual soil humidity was the sensor to be calibrated at the second stage. The data transmission unit USR-G781 is produced by Jinan YouRen Technology Company, China. The soil in each vessel was extruded to improve the soil hardness in each vessel as shown in Table 2.

The output of the FDR soil humidity sensor in 24 vessels was measured with a temperature gradient of 1°C from 0 to 50°C. In Table 1, the accurate soil humidity calculated in the 24 vessels is 5, 10, 15, 20, 25, and 30%. The output of the FDR soil humidity sensor is affected by soil temperature, soil hardness, and soil types [27]. The output measured is shown in Figures 3–5.

The experimental process was repeated using samples of brown soil and sandy soil. The output of FDR soil humidity sensors is shown in Figures 4 and 5. From Figures 3–5, it is found that the output of the FDR soil humidity sensor is significantly affected by soil temperature, soil hardness, and soil type. In Figure 3, the error reaches its maximum value of 5.28% at 50°C. The soil hardness is 80 N/cm^2, and the soil humidity is 20%.

In Figure 4, the error reaches its maximum value of 7.79% at 50°C. The soil hardness is 80 N/cm^2, and the soil

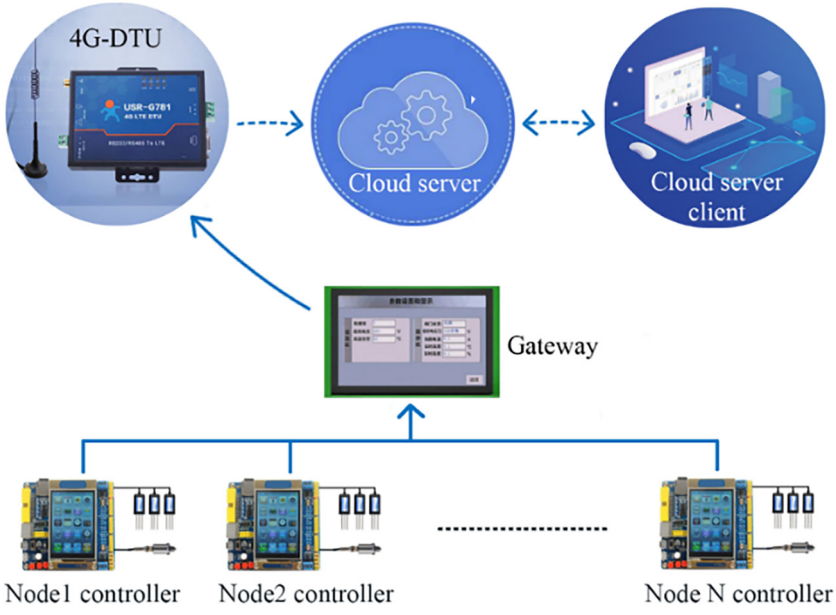

Figure 2: Experimental testing system.

Table 2: Correspondence between labels and soil hardness

Vessel number (X = A, B, C, D, E, F)	X1	X2	X3	X4
Soil hardness (unit: N/cm^2)	20	40	60	80

humidity is 20%. The error reaches its maximum value of 5.25% at 50°C. The soil hardness is 80 N/cm^2, and the soil humidity is 20%. In soil humidity measurement, it is necessary to eliminate the influence of soil temperature, soil hardness, and soil type on soil humidity sensor.

The output of the FDR soil humidity sensors buried in three soils is used to build 984 groups of data set. Each group of data set is composed of soil temperature, soil hardness, FDR soil humidity sensor output, actual soil humidity, and the type of soil.

In order to eliminate the influence of soil type on the output of the FDR soil humidity sensor, a decision tree is a popular model for the classification of soil [28]. Bagged tree and random forests are the most famous and competitive extensions of decision trees. Li *et al.* [29] pointed out that the bagged tree combined with C4.5 is an excellent performance integration approach for classification problems [30]. A bagged tree that reduces the prediction variance of the decision tree was adopted to avoid overfitting.

The bagged tree involves three steps. The original data are separated into subset data in step 1. Step 2 is to build a classifier on each data subset. Step 3 is the majority voting, which selects the best classifier from all the classifiers [31].

In the model of bagged tree, the C4.5 algorithm was used to construct trees based on a subset. The data set included soil temperature, soil hardness, and the output of the FDR soil humidity sensor. There were three labels: loam, brown soil, and sandy soil. In the C4.5 algorithm, the gain rate of the attribute needs to be calculated, and Eqs (5)–(8) show the calculation method. The basic information entropy represents the degree of chaos in the data set as Eq. (5). Information entropy is a basic concept of information theory, which describes the uncertainty of the occurrence of each possible event of an information source.

$$\text{entropy}(D) = -\sum_{j=1}^{|C|} \Pr(c_j)\log_2 \Pr(c_j), \tag{5}$$

where D represents a data set, $\Pr(c_j)$ represents the probability of category c_j in data set D, C represents the

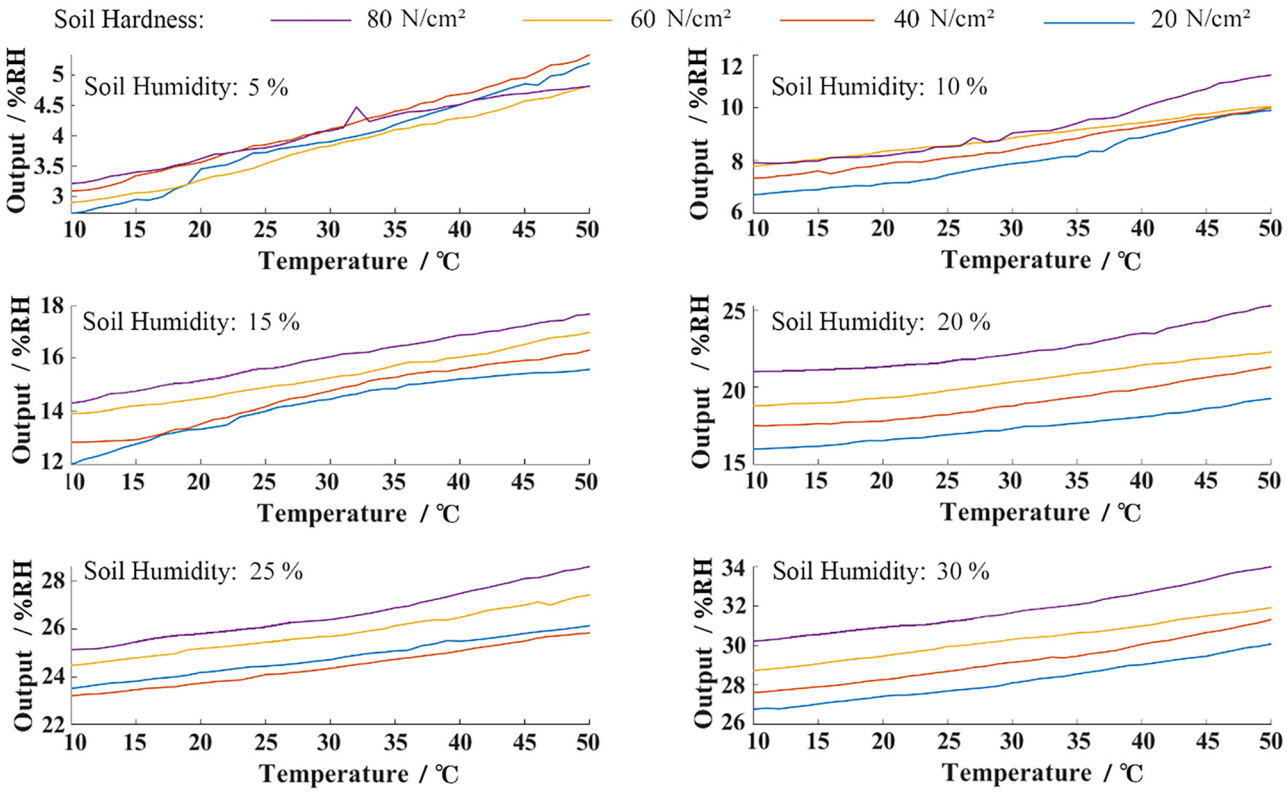

Figure 3: Output of FDR soil humidity sensor at various soil temperatures and hardness in loam.

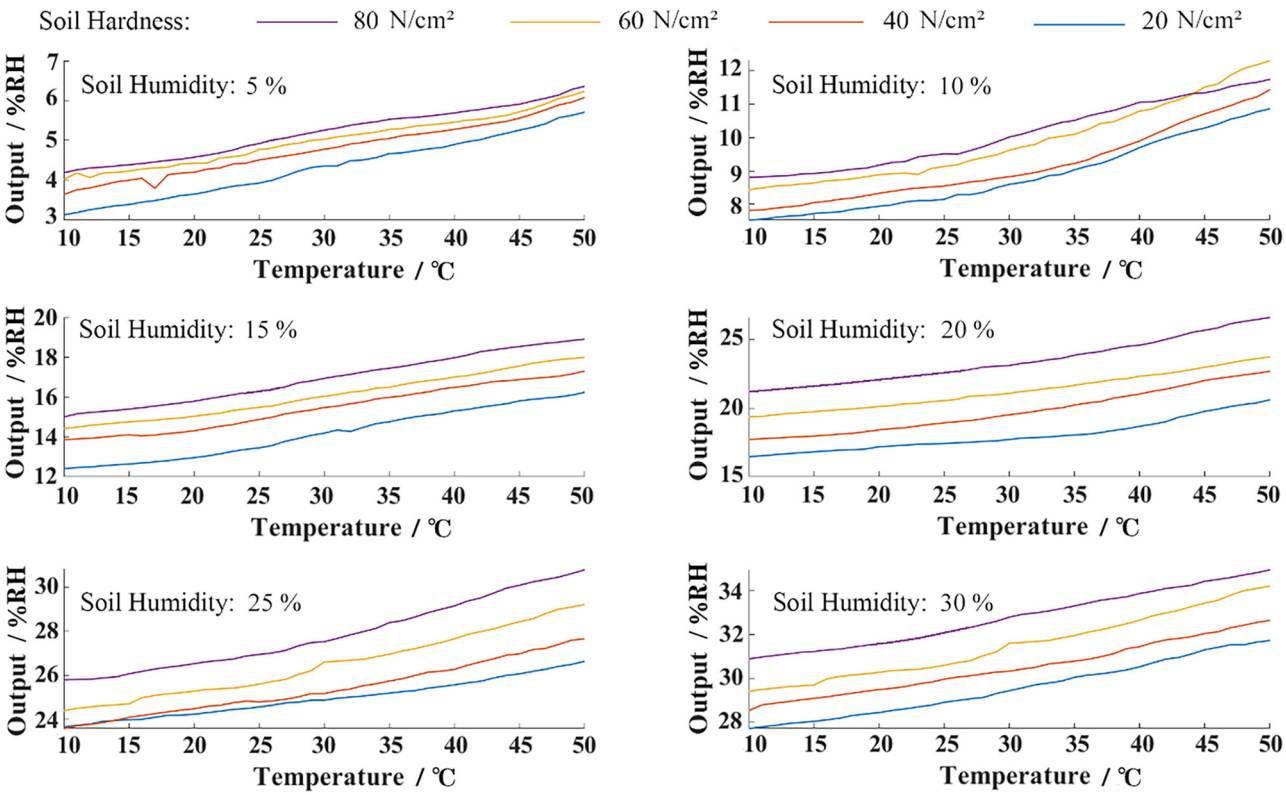

Figure 4: Output of FDR soil humidity sensor at various soil temperatures and hardness in brown soil.

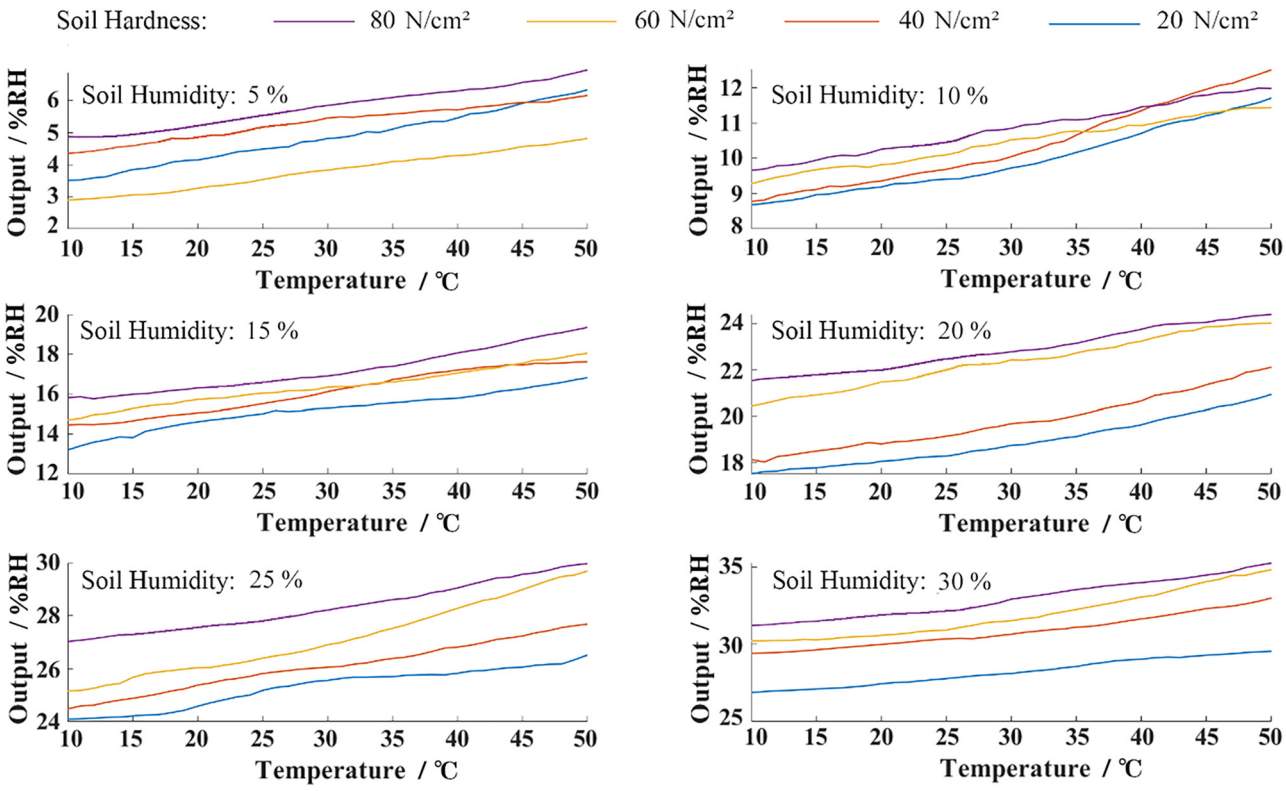

Figure 5: Output of FDR soil humidity sensor at various soil temperatures and hardness in sandy soil.

number of categories in data set D, and entropy(D) represents the entropy of the data set D. After the data set is divided into v subsets according to attributes, the entropy of the data set is calculated as the following equation:

$$\text{entropy}(A_i, D) = \sum_{j=1}^{v} \left| \frac{D_j}{D} \right| * \text{entropy}(D_j), \tag{6}$$

where A_i is an attribute of data set D, D_j is a subset of data set D divided by attributes, v is the number of attributes in the data set, and entropy(D_j) is the entropy of the subset D_j and entropy(D_j). The information gain of attribute A_i is calculated by the following equation:

$$\text{gain}(D, A_i) = \text{entropy}(D) - \text{entropy}(A_i, D), \tag{7}$$

where gain(D, A_i) represents the information gain of attribute A_i. The information gained is biased to select attributes with more values. In order to correct this bias, the entropy of the data set relative to the distribution of attribute values is used to normalize the information gain, which called the information gain rate and calculated as follows:

$$\text{entropy}(A_i, D) = \sum_{j=1}^{v} \left| \frac{D_j}{D} \right| * \text{entropy}(D_j), \tag{8}$$

where gainRatio(D, A_i) represents the information gain rate of attribute A_i.

Figure 6 shows the flowchart of the C4.5 algorithm. The process of building a decision tree using the C4.5 algorithm is divided into four steps. In step 1, the total data sample *MP* is input. In step 2, the entropy of *MP* and the entropy of each attribute are calculated. In step 3, the information gain rate of each attribute is calculated, and the data set is divided into the categories with the maximum information gain rate to generate nodes. In step 4, the decision tree and mapping rules are obtained by repeating step 1–step 3.

Figure 7 shows the three steps of the construction of the bagged tree. In Figure 7, *MP* dataset is the entire dataset, and n samples are selected from *MP* dataset as the training set by placing back samples. A decision tree model was established on each training set, and the C4.5 algorithm was adopted to construct each decision tree. The models were used to vote on the results of the test samples, so as to obtain the final prediction results.

The bagged tree algorithm is not dependent on one decision tree, but by many decision trees to jointly determine the prediction results, which can improve the accuracy and stability of the results, but also avoid the problem of over-fitting.

In this study, BP neural network was used for data fusion. The mathematical equation of the mapping relationship between input and output in advance is not necessary to determine in the artificial neural network [32]. Learning specific rules through training obtained the result closest to the expected output value [33]. The BP neural network is a multi-layer feed-forward network trained according to error BP. With the gradient descent and searching gradient, this BP algorithm [34] minimized the mean squared error (MSE) of the actual output value and the expected output value of the network [35].

The basic BP algorithm includes two processes: forward propagation of signal and backward propagation of error [36]. The error output, the adjustment weight, and threshold are calculated from input to output. In forward propagation, the input signal acts on the output node through the hidden layer, which generates the output signal through nonlinear transformation. If the actual output is inconsistent with the expected output, the process of backward propagation is to be carried out.

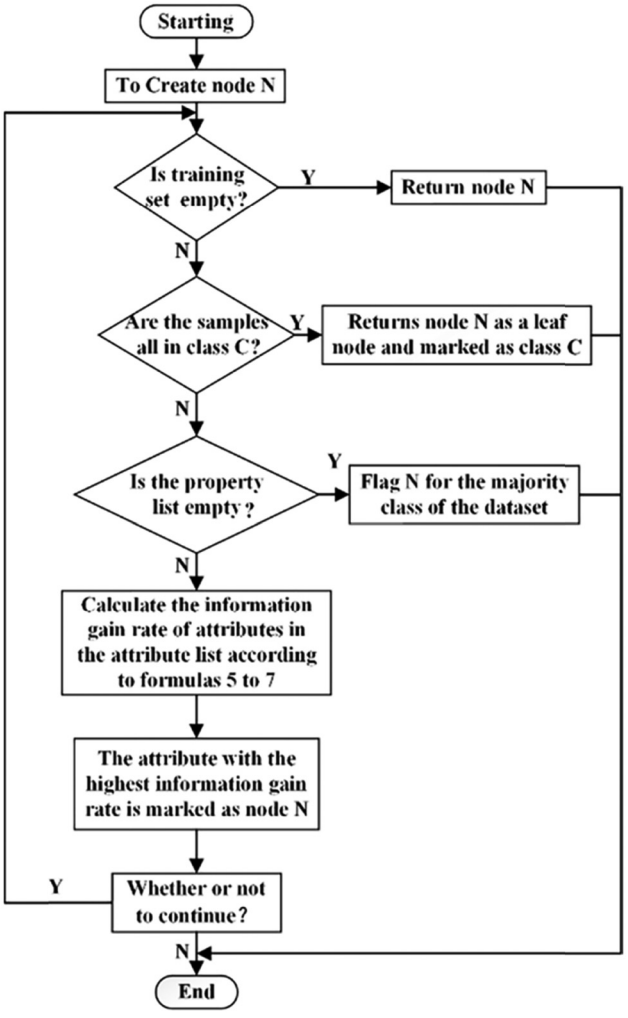

Figure 6: Flowchart of the C4.5 algorithm.

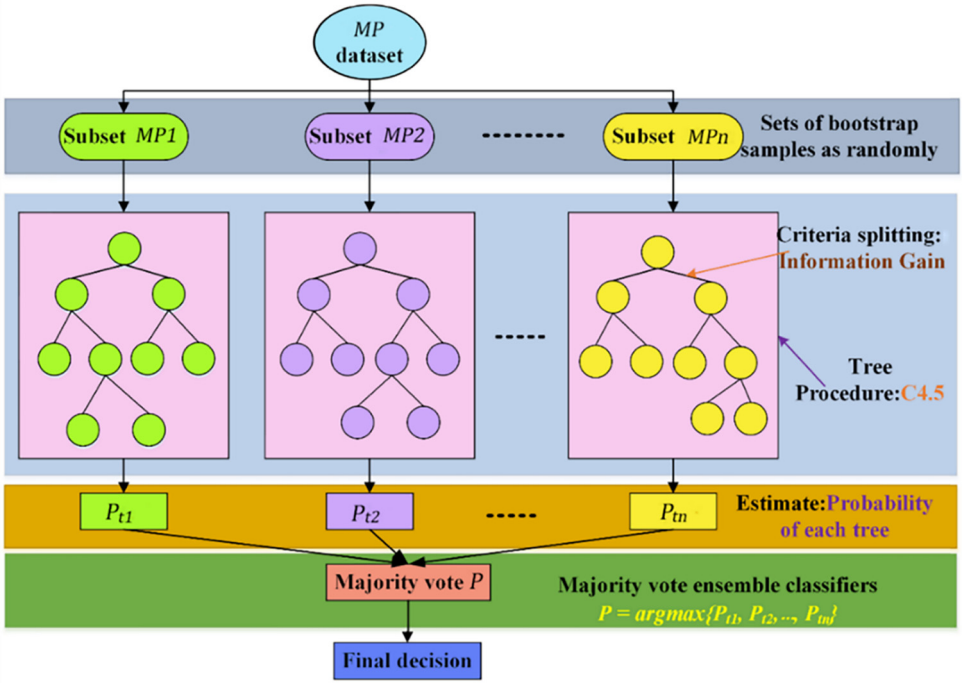

Figure 7: Model of the bagged tree.

Backward propagation of error is apportioned to all units of each layer, which means that the output error is transmitted back to the input through the hidden layer. The error is lowered along the gradient direction by adjusting the connection strengths between the input node and the hidden layer node and between the hidden layer node and the output node [37]. After repeated learning and training, the network parameters (weights and thresholds) corresponding to the minimum error are determined [38]. In this case, the trained neural network processes the input information of similar samples with the least output error after nonlinear transformation.

The topology of the three-layer neural network is shown in Figure 8.

There are n nodes in the input layer of BP network, q nodes in the hidden layer, and m nodes in the output layer. The weight between the input layer and the hidden layer is v_{ki}, and the weight between the hidden layer and the output layer is w_{jk}. The transfer function of the hidden layer is $f_1(\cdot)$, and the transfer function of the output layer is $f_2(\cdot)$; then, the output of the node of the hidden layer is shown as follows:

$$z_k = f_1\left(\sum_{i=0}^{n} v_{ki}x_i\right), \tag{9}$$

where the range of k is 1, 2,$\cdots$,q. The output of the node in the output layer is

$$y_j = f_2\left(\sum_{k=0}^{q} w_{jk}z_k\right). \tag{10}$$

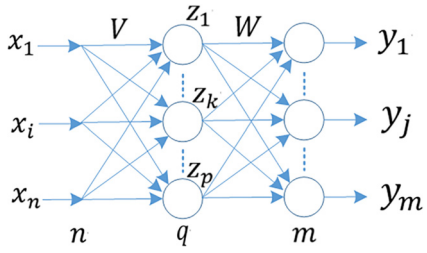

Figure 8: Topology of the three-layer neural network.

At this point, BP network has completed the mapping from N-dimensional space vector to M-dimensional space. p learning samples are input into the BP network, and they are denoted by $x_1, x_2, x_3, \cdots, x_p$. The pth sample is input to the network, and the output $y_j^p (j = 1, 2, \ldots, m)$ is obtained. The squared error function is used, so the error E_p of the pth is obtained.

$$E_p = \frac{1}{2}\sum_{j=1}^{m}(t_j^p - y_j^p)^2, \tag{11}$$

where t_j^p is the expected output. For p samples, the global error is shown as follows:

$$E = \frac{1}{2}\sum_{p=1}^{P}\sum_{j=1}^{m}(t_j^p - y_j^p) = \sum_{p=1}^{P}E_p. \tag{12}$$

The cumulative error BP algorithm is used to adjust w_{jk} to make the global error E smaller.

$$\Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}} = \sum_{p=1}^{P}\left(-\eta \frac{\partial E_p}{\partial w_{jk}}\right), \tag{13}$$

where η is the learning rate. The error signal is δ_{yj}.

$$\delta_{yj} = -\frac{\partial E_p}{\partial S_j} = -\frac{\partial E_p}{\partial y_j}\frac{\partial y_j}{\partial S_j}, \tag{14}$$

where the first term is shown in Eq. (15) and the second term is the partial of the transfer function of the output layer, which is shown in the following equation:

$$\frac{\partial E_p}{\partial y_j} = \frac{\partial}{\partial y_j}\left[\frac{1}{2}\sum_{j=1}^{m}(t_j^p - y_j^p)\right] = -\sum_{j=1}^{m}(t_j^p - y_j^p), \tag{15}$$

$$\frac{\partial y_j}{\partial S_j} = f_2'(S_j). \tag{16}$$

So, δ_{yj} can be expressed as follows:

$$\delta_{yj} = \sum_{j=1}^{m}(t_j^p - y_j^p) f_2'(S_j). \tag{17}$$

According to the chain rule,

$$\frac{\partial E_p}{\partial w_{jk}} = \frac{\partial E_p}{\partial S_j}\frac{\partial S_j}{\partial w_{jk}} = -\delta_{yj} z_k = -\sum_{j=1}^{m}(t_j^p - y_j^p) f_2'(S_j)\cdot z_k. \tag{18}$$

Therefore, the weight adjustment formula of each neuron in the output layer is as follows:

$$\Delta w_{jk} = \sum_{p=1}^{P}\sum_{j=1}^{m}\eta(t_j^p - y_j^p) f_2'(S_j)\cdot z_k. \tag{19}$$

The change of hidden layer weight is given as follows:

$$\Delta v_{ki} = \sum_{p=1}^{P}\left(-\eta \frac{\partial E_p}{v_{ki}}\right). \tag{20}$$

The error signal is given as follows:

$$\delta_{zk} = -\frac{\partial E_P}{\partial S_k} = -\frac{\partial E_p}{\partial z_k}\frac{\partial z_k}{\partial S_k}, \tag{21}$$

where the first term is shown in the following equation:

$$\frac{\partial E_p}{\partial z_k} = \frac{\partial}{\partial z_k}\left[\frac{1}{2}\sum_{j=1}^{m}(t_j^p - y_j^p)\right] = -\sum_{j=1}^{m}(t_j^p - y_j^p)\frac{\partial y_j}{\partial z_k}. \tag{22}$$

Eq. (23) is obtained by the chain theorem.

$$\frac{\partial y_i}{\partial z_k} = \frac{\partial y_j}{\partial S_j}\frac{\partial S_j}{\partial z_k} = f_2'(S_j) w_{jk}. \tag{23}$$

The second term in Eq. (21) is the partial derivative of the transfer function of the hidden layer, which is shown in the following equation:

$$\frac{\partial z_k}{\partial S_k} = f_1'(S_k). \tag{24}$$

So, Eq. (25) is given as follows:

$$\delta_{zk} = \sum_{j=1}^{m}(t_j^p - y_j^p) f_2'(S_j) w_{jk} f_1'(S_k). \tag{25}$$

According to the chain theorem,

$$\begin{aligned}\frac{\partial E_p}{\partial v_{ki}} &= \frac{\partial E_p}{\partial S_k}\frac{\partial S_k}{\partial v_{ki}} \\ &= -\delta_{zk} x_i \\ &= -\sum_{j=1}^{m}(t_j^p - y_j^p) f_2'(S_j) w_{jk} f_1'(S_k) x_i.\end{aligned} \tag{26}$$

Thus, the weight adjustment formula of each neuron in the hidden layer is as follows:

$$\Delta v_{ki} = \sum_{p=1}^{P}\sum_{j=1}^{m}\eta(t_j^p - y_j^p) f_2'(S_j) w_{jk} f_1'(S_k) x_i. \tag{27}$$

In this study, the BP neural network was adopted for data fusion, and the data from Figures 3–5 were used as the data set of the BP neural network [39]. The structured setting of the neural network is shown in Figure 9.

In Figure 9, neurons $i1$, $i2$, $i3$, and $i4$ in input layer $L1$ of the neural network are soil temperature, soil hardness, FDR soil humidity sensor, and soil type. There are eight neurons in hidden layer $L2$, and neuron Q in output layer $L3$ is the output of the network, which is the soil humidity value after two-stage calibration.

The neural network was trained by the established data set to determine the weights from the input layer to

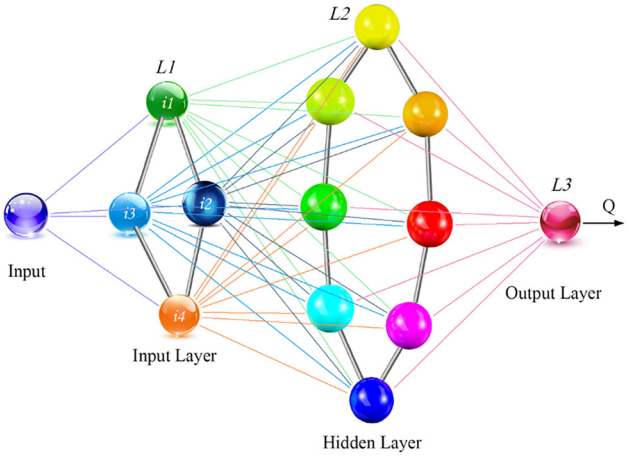

Figure 9: Structure of neural network.

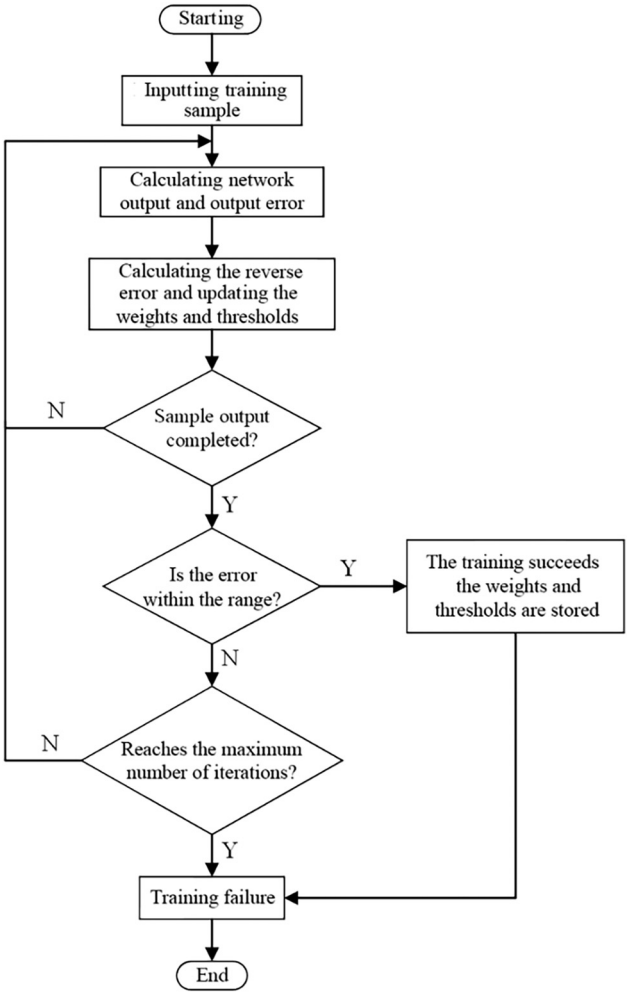

Figure 10: Flowchart of BP neural network.

the hidden layer and from the hidden layer to the output layer and the thresholds of each neuron in the hidden layer and output layer [40]. Figure 10 shows the flowchart of the BP neural network based on the algorithm of error reverse propagation.

After the BP neural network was trained, the weights of the input layer to the hidden layer, the hidden layer to the output layer, and the thresholds of neurons of the hidden layer and neurons of the output layer were obtained [41]. Table 3 shows the weights from input to hidden layers and the thresholds of neurons of the hidden layer.

Table 4 shows the weights from hidden to output layers and the thresholds of neurons of the output layer.

Tables 3 and 4 are deployed to the following system for secondary calibration of the FDR soil humidity sensor.

3 System structure

3.1 System overview

Based on the experimental analysis, a farmland environment monitoring system with two-stage calibration function was constructed.

The primary calibration is performed in this system by measuring historical soil humidity. In the secondary calibration, accurate soil humidity value is obtained through the acquisition of the current soil temperature, soil hardness, humidity output, and network calculations. The soil humidity value calibrated is closer to the actual value. It has a real guiding significance for crop growth monitor and water-saving. The system can also be used to monitor other environmental parameters of crop growth, including ambient temperature, air humidity, carbon dioxide concentration, and light intensity. These environmental parameters also showed vital guiding significance for crop growth and other agricultural activities [42].

3.2 Edge computing

As an extension and supplement of cloud computing, edge computing is considered a new computing paradigm that performs computing tasks at the edge of the network [43]. The core concept of edge computing is to make computing closer to the edge of the network and data sources to provide edge services nearby to meet the

Table 3: Weights and thresholds from input layer to hidden layer

Neuron	*h*1	*h*2	*h*3	*h*4	*h*5	*h*6	*h*7	*h*8
Threshold	−2.3545	1.6818	−1.0091	0.3364	−0.3364	−1.0091	0.5343	−1.1011
*i*1	1.4268	−1.3495	1.6309	−1.4247	−0.5631	−1.5113	0.9078	1.5212
*i*2	−1.4192	1.2833	−0.5860	0.8284	−0.5172	0.4666	−0.9872	1.0212
*i*3	0.9376	1.4137	1.2027	−0.6019	0.5223	0.2766	0.5612	−0.8967
*i*4	−0.7841	−0.2781	−1.0459	1.5702	2.1648	−1.7220	−1.0539	0.3436

Table 4: Weights and thresholds from hidden layer to output layer

Neuron		h1	h2	h3	h4	h5	h6	h7	h8
Q	Threshold 0.7707	0.6042	0.9783	−0.866	0.8788	−0.964	0.3677	0.5675	0.0683

needs of users in agile connection [44], real-time data transmission, data optimization, application intelligence, security, and privacy protection [45]. The edge computing consortium defines four domains for edge computing, that is, device domain, network domain, data domain, and application domain, as shown in Figure 11. The application domain, data domain, network domain, and equipment domain are computing objects of edge computing. In the device domain, edge computing can directly process the perceived information. In the network domain, the automatic conversion of every network protocol is implemented, and the data format is standardized. Computing at the edge of the data domain makes data management more intelligent and flexible. Edge computing can provide localized controlling logic and application intelligence [46].

In a traditional cloud computing environment, only the cloud has a computing function [47]. However, due to the large amount of data collected by the agricultural monitoring system, network delays and congestion will occur if all collected data are uploaded to the cloud [48]. The system constructed in this study adopts the method of lightweight edge computing. The edge computing in this system is in the data domain layer in Figure 11. The humidity sensor saturation value, the calculated soil hardness, the calculation process of the neural network, and the model of soil classification were deployed to the edge of the system to finish the process of two-stage calibration of the FDR soil humidity sensor.

The 32-bit ST microelectronics (STM32) was adopted as the microcontroller unit of edge calculation, and Figure 12 shows the connection between STM32 and the sensors.

The program flowchart of STM32 is shown in Figure 13. The algorithm of model of bagged tree and the algorithm of BP neural network were implanted into STM32.

The STM32 collects data from the FDR soil humidity sensor, soil temperature sensor, and pressure sensor, which converts pressure data into soil hardness data by Eq. (3), and stores soil humidity data. Eq. (2) is used to obtain the soil humidity value after primary calibration. The model of a bagged tree is used to classify the soil. The soil temperature, soil hardness, soil humidity after the primary calibration, and kind of soil are used as a group of input data for the neural network. The neural network calculates the accurate soil humidity after the second calibration. When the STM32 receives the command for data collecting, it sends the soil humidity data to the long-range radio (LoRa) sub-nodes in the agriculture Internet of things (IoT).

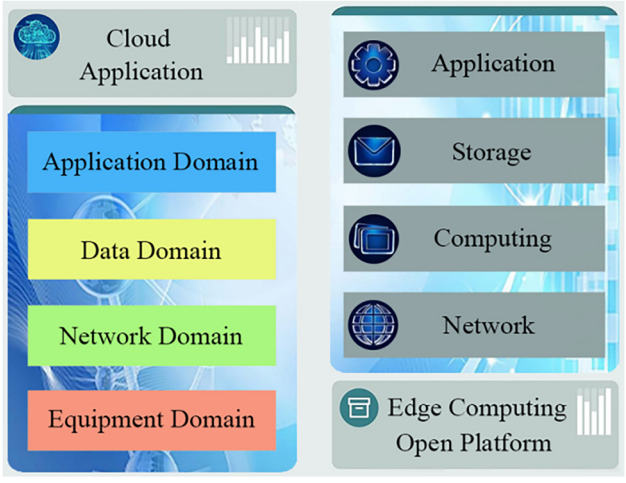

Figure 11: Field of edge computing.

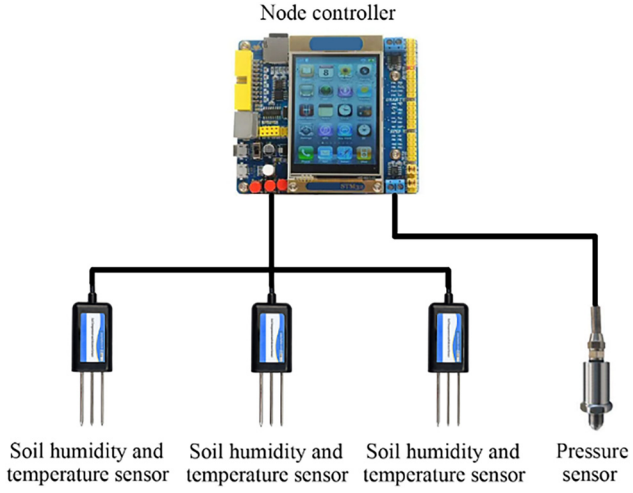

Figure 12: Connection between STM32 and sensors.

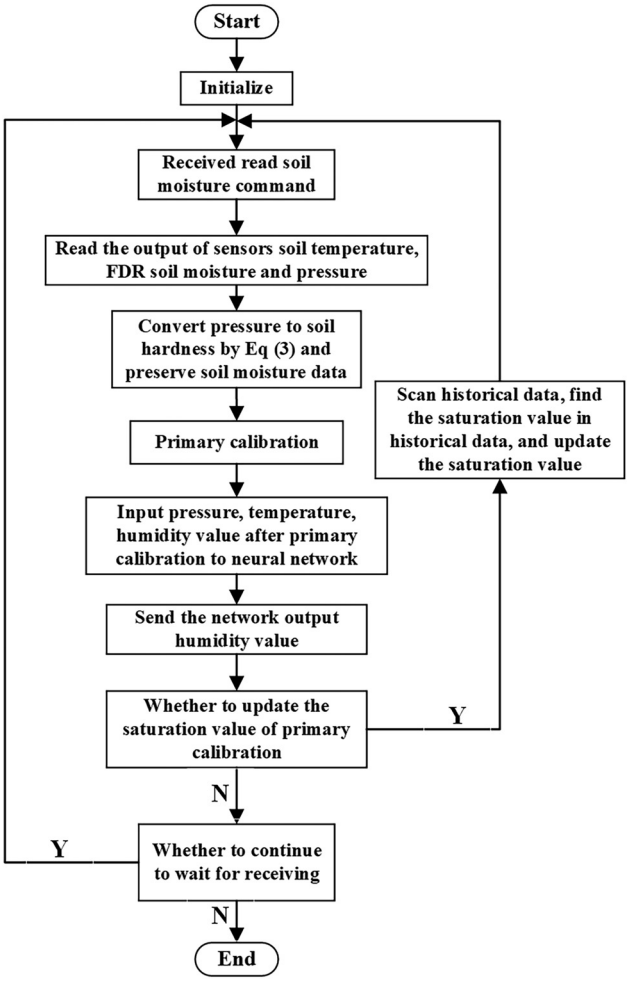

Figure 13: Controller program flowchart.

3.3 System composition

The system consists of the perception, network, and application layers [49]. The overall architecture of the system is shown in Figure 14. The perception layer consists of five sensors and a microprocessor STM32, in which three sensors are connected to STM32. Modbus remote terminal unit (RTU) protocol is adopted for communication between LoRa sub-nodes and the perception layer, and Modbus RTU protocol is also adopted for light, temperature, humidity, and CO_2 concentration sensors. When STM32 receives the data transmission command from LoRa sub-nodes, the soil humidity data calibrated are packaged and sent to LoRa sub-nodes by Modbus RTU protocol [50].

LoRa technology is used for data transmission at the network layer. LoRa spreads farther than another wireless mode under the same power consumption, realizing the unification of low power consumption and long distance [51]. Compared with traditional wireless communication, the distance of LoRa is 3–5 times larger than that of conventional radio frequency communication, up to 2–5 km [52].

In addition, LoRa technology supports the topology of the star network [53]. LoRa technology is very suitable for agricultural IOT [54]. The network layer consists of LoRa sub-nodes and LoRa gateway. LoRa sub-nodes send the perception layer data to the LoRa gateway, and then, the data are transferred to the cloud monitoring platform by the 4G network, which receives the data from the sub-nodes and sends the data to the cloud monitoring platform through the SIM card.

The cloud monitoring platform adopts the cloud platform of the Youren Cloud, Ltd., Shandong Province, China. The data transmission mode between the cloud monitoring platform and LoRa gateway is network pass-through. On the cloud monitoring platform, users can check the growing environment of crops in the farmland at any time to take related agricultural activities.

Figure 15 shows the data transmission modes among the cloud monitoring platform, LoRa gateway, LoRa sub-nodes, and sensors. After completing the network construction, users log in to the cloud monitoring platform to view the environmental conditions of the farmland. At this time, the soil humidity displayed on the platform is the soil humidity value calibrated by the two-stage calibration method in this study. Farmers send a command to update the calibration information in the cloud every time when they irrigate their fields or during the rainy season. When the STM32 receives the command, it will update primary calibration data according to the flowchart in Figure 13. The secondary calibration is performed automatically in the STM32 without the farmer's involvement. In addition, light intensity, carbon dioxide concentration, ambient temperature, and humidity in the farm are also displayed to the user in the cloud monitoring platform. According to the environmental information, users can take related agricultural activities to improve the growth condition and yield of crops.

4 Results and discussion

This chapter shows discussions about the results of the primary calibration of soil humidity sensor, the soil classification model, and the performance of the BP neural network.

4.1 Primary calibration

If there are m groups of data, the calculation method of MSE is shown in the following equation:

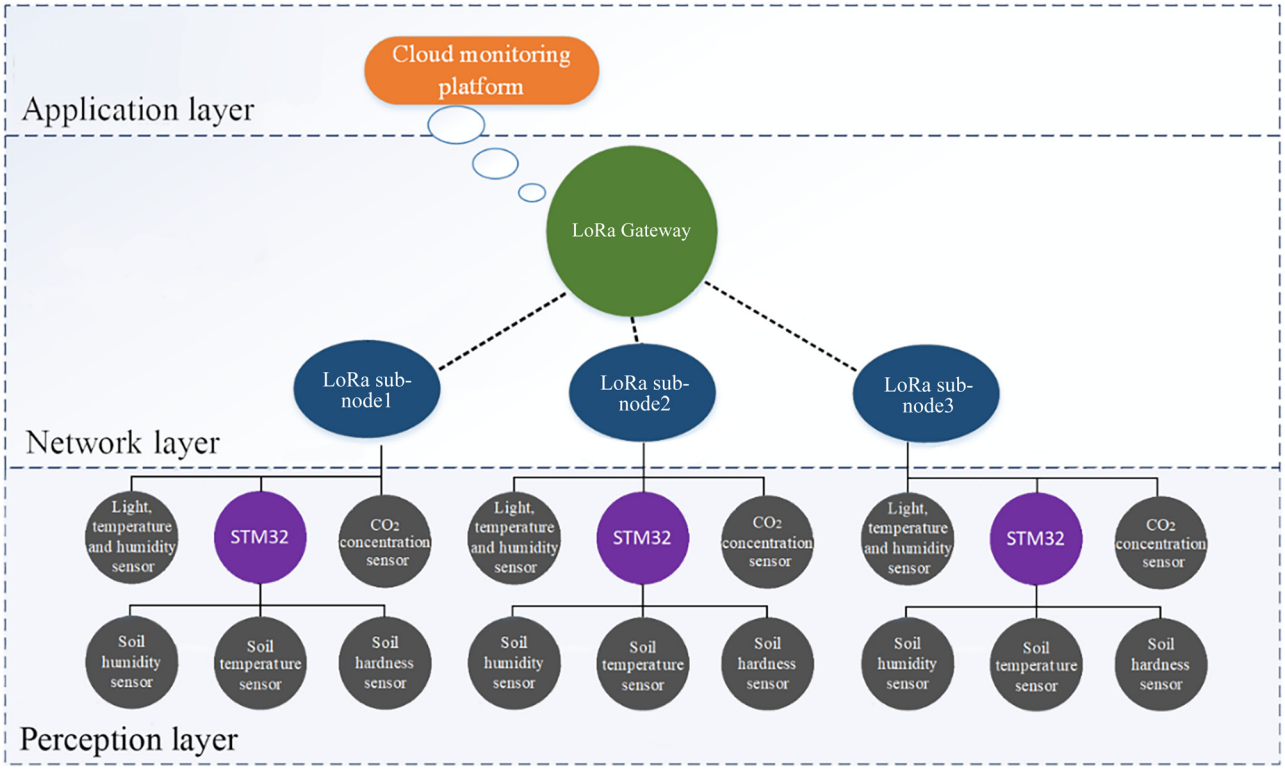

Figure 14: Overall architecture of the system.

$$\mathrm{MAE} = \frac{1}{m}\sum_{i=1}^{m}|f_i - y_i|, \tag{28}$$

where m is the number of groups of data, f_i is the predicted value, and y_i is the real value. The calculation method of RMSE is shown in the following equation:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(f_i - y_i)^2}{N}}. \tag{29}$$

The calculation method of MSE is shown in the following equation:

$$\mathrm{MSE} = \frac{1}{m}\sum_{i=1}^{m}(f_i - y_i)^2. \tag{30}$$

Table 5 shows the MSE, the root mean squared error (RMSE), and the mean absolute error (MAE) and those of sensor B and sensor A and the error of sensor C and sensor A before and after calibration in Eqs (1) and (2).

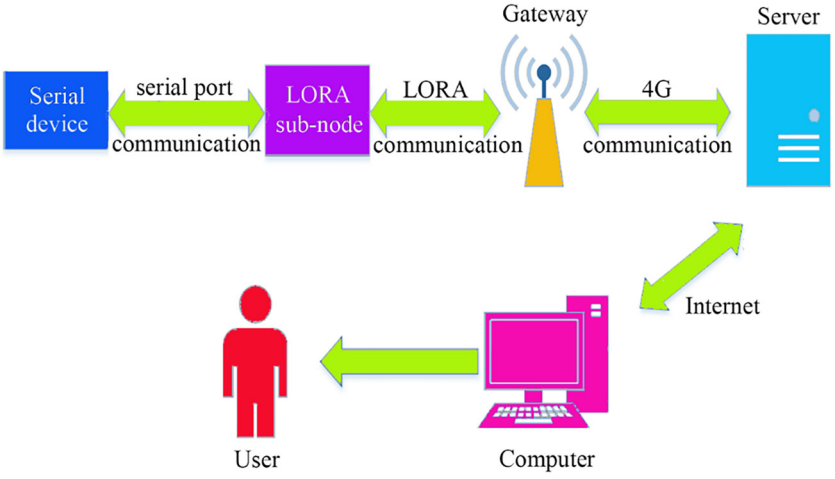

Figure 15: Way in which parts of a system communication.

Table 5: Error comparison of sensors before and after first-stage calibration

Error types		MSE	RMSE	MAE
Errors between sensor B and sensor A	Before calibration	73.4681	0.8571%	8.5509%
	After calibration	0.7208	0.0849%	0.6929%
Errors between sensor C and sensor A	Before calibration	145.4828	1.2062%	12.0460%
	After calibration	0.5139	0.0717%	0.5102%

Compared with sensor A, the RMSE of sensor B and sensor C is reduced to a value below 0.0849% after first-stage calibration. For the MSE and the MAE, sensors B and C show the highest values of 0.7208 and 0.6929%. After first calibration, the output value of soil sensor is closer to the standard soil sensor output value.

4.2 Soil classification model

The accuracy of data classification reaches 90.0% using the bagged tree model with the training set. Regarding the data characteristics in the dataset, the bagged tree classification model is much more accurate than other methods.

There are some accuracies of some other methods. The linear discriminant, the support vector machine, the decision tree, the k nearest neighbors, and the bagged tree algorithms show classification accuracies of 34.4, 33.1, 42.6, and 52.8%. Figure 16 shows the confusion matrix and receiver operating characteristic (ROC) curves using the bagged tree training dataset.

A total of 2,952 groups of test samples were tested for the bagged tree classification model. In confusion matrix of bagged tree classification model, a total of 2,657 groups of test samples were accurately classified and a total of

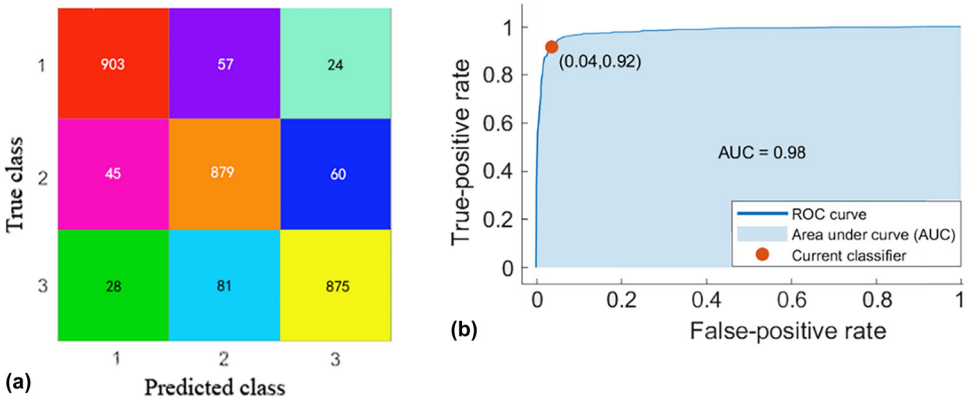

Figure 16: Confusion matrix and ROC curves for the bagged tree model. (a) The confusion matrix of bagged tree and (b) the ROC curves of the bagged tree.

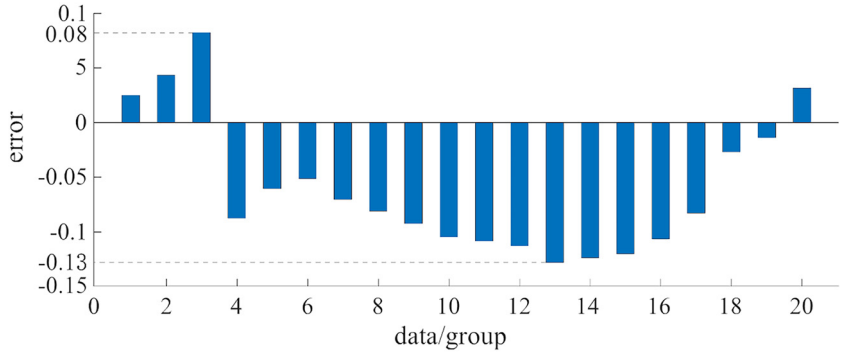

Figure 17: Output error between the neural network and actual soil humidity.

295 groups of test samples were inaccurately classified. The classification accuracy of bagged tree classification model is 90%. In ROC curves, false-positive rate is the probability that a wrong sample is predicted to be positive and true-positive rate is the probability that a correct sample is predicted to be positive. Area under curve (AUC) is the area under the ROC curve. AUC is an evaluation index of classification model, and the closer it is to the 1, the better the effect of classification model. Among the above classification algorithms, the bag tree algorithm has the largest AUC value, which is as high as 0.98.

The bagged tree is the most suitable algorithm for this data set. Using the bagged tree algorithm, soil classification prepares for the BP neural network and secondary calibration of the FDR soil humidity sensor.

4.3 Secondary calibration

After the neural network training was completed, 20 sets of data were verified. Figure 17 shows the output error between the network and the actual soil humidity.

The absolute error between the output of neural network and actual soil humidity is controlled within 0.13% in Figure 17. This absolute error meets the needs of agricultural soil humidity measurement.

Table 6 shows the error between output of the network and the actual soil humidity and the error between input of the network and the actual soil humidity.

MAE, MSE, and RMSE between the output of the network and actual soil humidity are 1.37%, 3.79, and 1.86%. less than those between the input of network and actual soil humidity. The output of the neural network almost eliminates the effects of soil temperature, soil hardness, and soil type on the output of the FDR soil humidity sensor. The accuracy of soil humidity measurement is improved by using the FDR soil humidity sensor in agriculture. This result improves the utilization rate of agricultural water and saves agricultural water resources, which accurately determines the process of crop irrigation.

Various algorithms have been tried for the soil classification models, but only the bagged tree algorithm accurately classifies soil based on the mechanism of voting. For the BP neural network model, the value of soil humidity is closer to the real soil humidity according to the input, and its training results meet the expected expectations. The effects of soil temperature, soil hardness, and soil type on the FDR soil humidity sensor are removed by the calibration method in this study, and the accurate soil humidity value is obtained.

Table 6: Error among network input, output, and real soil humidity

Error classification	MAE	MSE	RMSE
Error between output and actual value	0.08%	0.01	0.09%
Error between input and actual value	1.45%	3.80	1.95%
difference value	1.37%	3.79	1.86%

5 Conclusions

A two-stage calibration method for the measurement of FDR soil humidity sensor is proposed in this study. Linear and nonlinear data drifts of the FDR soil humidity sensor are analyzed.

The RMSE between measured values of sensor B and sensor C after first-stage calibration and that of standard sensor A is reduced to less than 0.0849%. Both the MSE and the MAE are reduced to less than 0.7208 and 0.6929%. In the secondary calibration, the data sets of soil temperature, soil hardness, soil types, soil humidity sensor output, and actual soil humidity are established experimentally. The data set trains the bagged tree for soil classification model and the BP neural network. The accuracy of the soil classification model reaches 90%. The output of the trained neural network is closer to the actual soil humidity than that of the FDR soil humidity sensor. The MAE, MSE, and RMSE decrease by 1.37%, 3.79, and 1.86%. The calibrated soil humidity value is indispensable for agricultural activities and saving agricultural water.

The dynamic monitoring system of the crop growth environment is built based on the two-stage calculation. The process of sensor secondary calibration is deployed to the edge of the network so that the calibrated soil humidity is collected in real time to guide the user to carry out corresponding agricultural irrigation activities in time, which reduces unnecessary water consumption in agriculture. The system adds the function of collecting the growth environment parameters of crops, which is convenient for users to carry out other agricultural activities according to environmental changes.

A novel idea for water-saving irrigation and water-resource saving globally is provided in this study, but some shortcomings exist in this study. The hardness measurement of soil in this experiment was insufficient. The soil was only roughly divided into three categories, and the measurement range of soil hardness was limited. In the subsequent work, soil classifications in both China

and around the world will be shown in detail, and the measurement range of soil hardness will be further expanded.

Funding information: This work was supported by the Natural Science Foundation of Hebei Province (Grant No. 19227212D), the Natural Science Foundation of Hebei Province of China (Grant No. E2021202163), and the Special Project of Science and Technology Winter Olympics in the Hebei Technology Innovation Guidance Plan (Grant No. 21474501D). This research is also supported by a project "Sustainable Process Integration Laboratory – SPIL," project No. CZ.02.1.01/0.0/0.0/15_003/0000456 funded by EU as "CZ Operational Programme Research, Development and Education," Priority 1: Strengthening capacity for quality research under the collaboration agreement with Hebei University of Technology, Tianjin, China.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: The authors state no conflict of interest.

References

[1] Zhao M, Zhao L, Xiong X, He Y, Huang W, Liu Z, et al. Performance analysis of a plate heat exchanger using various nanofluids. Int J Heat Mass Transf. 2020;158:119993–2281.
[2] Wang J, Li Y, Zheng D, Mikulčić H, Vujanović M, Sundén B. Preparation and thermophysical property analysis of nano-composite phase change materials for energy storage. Renew Sustain Energy Rev. 2021;151:111541.
[3] Wang J, Li G, Zhu H, Luo J, Sundén B. Experimental investigation on convective heat transfer of ferrofluids inside a pipe under various magnet orientations. Int J Heat Mass Transf. 2019;132:407–19.
[4] Zheng D, Yang J, Wang J, Kabelac S, Sundén B. Analyses of thermal performance and pressure drop in a plate heat exchanger filled with ferrofluids under a magnetic field. Fuel. 2021;293:120432.
[5] Zheng D, Du J, Wang W, Klemeš JJ, Wang J, Sundén B. Analysis of thermal efficiency of a corrugated double-tube heat exchanger with nanofluids. Energy. 2022;256:124522.
[6] Béné C, Prager SD, Achicanoy HAE, Toro PA, Lamotte L, Cedrez CB, et al. Understanding food systems drivers: A critical review of the literature. Glob Food Security. 2019;23:149–59.
[7] Zhang B, AghaKouchak A, Yang Y, Wei J, Wang G. A water-energy balance approach for multi-category drought assessment across globally diverse hydrological basins. Agric For Meteorol. 2019;264:247–65.
[8] Bai Z, Caspari T, Gonzalez MR, Batjes NH, Mäder P, Bünemann EK, et al. Effects of agricultural management practices on soil quality: A review of long-term experiments for Europe and China. Agric Ecosyst Environ. 2018;265:1–7.
[9] Rigby D, Cáceres D. Organic farming and the sustainability of agricultural systems. Agric Syst. 2001;68(1):21–40.
[10] Nimmagadda SM, Harish KS. Review paper on technology adoption and sustainability in India towards smart cities. Multimed Tools Appl. 2022;81:27217–45.
[11] Fritz S, See L, Bayas JCL, Waldner F, Jacques D, Becker-Reshef I, et al. A comparison of global agricultural monitoring systems and current gaps. Agric Syst. 2019;168:258–72.
[12] Rajendran S, Priya AK, Senthil Kumar P, Hoang T, Sekar K, Chong KY, et al. A critical and recent developments on adsorption technique for removal of heavy metals from wastewater-A review. Chemosphere. 2022;303:135146.
[13] Jiao W, Tian C, Chang Q, Novick KA, Wang L. A new multi-sensor integrated index for drought monitoring. Agric For Meteorol. 2019;268:74–85.
[14] Wild J, Kopecký M, Macek M, Šanda M, Jankovec J, Haase T. Climate at ecologically relevant scales: A new temperature and soil humidity logger for long-term microclimate measurement. Agric For Meteorol. 2019;268:40–7.
[15] Huuskonen J, Oksanen T. Soil sampling with drones and augmented reality in precision agriculture. Comput Electron Agric. 2018;154:25–35.
[16] Wang J, Li G, Li T, Zeng M, Sundén B. Effect of various surfactants on stability and thermophysical properties of nanofluids. J Therm Anal Calorim. 2021;143:4057–70.
[17] Kamilaris A, Prenafeta-Boldú FX. Deep learning in agriculture: A survey. Comput Electron Agric. 2018;147:70–90.
[18] Parsons DJ, Rey D, Tanguy M, Holman IP. Regional variations in the link between drought indices and reported agricultural impacts of drought. Agric Syst. 2019;173:119–29.
[19] Iizumi T, Ramankutty N. How do weather and climate influence cropping area and intensity? Glob Food Secur. 2015;4:46–50.
[20] Yu H, Zou W, Chen J, Chen H, Yu Z, Huang J, et al. Biochar amendment improves crop production in problem soils: A review. J Environ Manag. 2019;232:8–21.
[21] Turner C, Aggarwal A, Walls H, Herforth A, Drewnowski A, Coates J, et al. Concepts and critical perspectives for food environment research: A global framework with implications for action in low-and middle-income countries. Glob Food Secur. 2018;18:93–101.
[22] Ramadan KM, Oates MJ, Molina-Martinez JM, Ruiz-Canales A. Design and implementation of a low cost photovoltaic soil humidity monitoring station for irrigation scheduling with different frequency domain analysis probe structures. Comput Electron Agric. 2018;148:148–59.
[23] Javed M, Sajid M, Yousaf HMZ, Hassan G, Mahmood H. Facile and low cost temperature compensated humidity sensor and signal conditioning system. IEEE Sens J. 2021;21(13):14906–14.
[24] Matthees HL, He Y, Owen RK, Hopkins D, Deutsch B, Lee J, et al. Predicting soil electrical conductivity of the saturation extract from a 1:1 soil to water ratio. Commun Soil Sci Plant Anal. 2017;48(18):2148–54.
[25] Louki II, Al-Omran AM, Aly AA, Al-Harbi AR. Sensor effectiveness for soil water content measurements under normal and extreme conditions. Irrig Drain. 2019;68(5):979–92.
[26] Mosavi MR, Khishe M, Akbarisani M. Neural network trained by biogeography-based optimizer with chaos for sonar data set classification. Wirel Personal Commun. 2017;95:4623–42.

[27] Lu S, Lu Y, Peng W, Ju Z, Ren T. A generalized relationship between thermal conductivity and matric suction of soils. Geoderma. 2019;337:491–7.
[28] Ge Y, Shen L, Sun M. Temperature compensation for optical fiber graphene micro-pressure sensor using genetic wavelet neural networks. IEEE Sens J. 2021;21:24195–201.
[29] Li Y, Chiou CT, Li H, Schnoor JL. Improved prediction of the bioconcentration factors of organic contaminants from soils into plant/crop roots by related physicochemical parameters. Environ Int. 2019;126:46–53.
[30] Tao K, Zhao S, Gao P, Wang L, Jia H. Impacts of Pantoea agglomerans strain and cation-modified clay minerals on the adsorption and biodegradation of phenanthrene. Ecotoxicol Environ Saf. 2018;161:237–44.
[31] Liu Q, Liu Y, Zhang C, Ruan Z, Meng W, Cai Y, et al. sEMG-based dynamic muscle fatigue classification using SVM with improved whale optimization algorithm. IEEE Internet Things J. 2021;8(23):16835–44.
[32] Le TTH, Kang H, Kim H. Household appliance classification using lower odd-numbered harmonics and the bagging decision tree. IEEE Access. 2020;8:55937–52.
[33] Iseki K, Konta T, Asahi K, Yamagata K, Fujimoto S, Tsuruya K, et al. Association of dipstick hematuria with all-cause mortality in the general population: results from the specific health check and guidance program in Japan. Nephrol Dialysis Transplant. 2018;33(5):825–32.
[34] Yang A, Zhuansun Y, Liu C, Li J, Zhang C. Design of intrusion detection system for internet of things based on improved BP neural network. IEEE Access. 2019;7:106043–52.
[35] Guo C, Cui Y, Dong B, Luo Y, Liu F, Zhao S, et al. Test study of the optimal design for hydraulic performance and treatment performance of free water surface flow constructed wetland. Bioresour Technol. 2017;238:461–71.
[36] Ghosh S, Bequette BW. Spectral graph theoretic analysis of process systems: An application to distillation columns. Comput Chem Eng. 2022;161:107748.
[37] Jiang Q, Huang R, Huang Y, Chen S, He Y, Lan L, et al. Application of BP neural network based on genetic algorithm optimization in evaluation of power grid investment risk. IEEE Access. 2019;7:154827–35.
[38] Mosbah H, El-Hawary ME. Optimization of neural network parameters by Stochastic Fractal Search for dynamic state estimation under communication failure. Electr Power Syst Res. 2017;147:288–301.
[39] Rush B, Celi LA, Stone DJ. Applying machine learning to continuously monitored physiological data. J Clin Monit Comput. 2019;33:887–93.
[40] Yang Y, Kim KR, Kou R, Li Y, Fu J, Zhao L, et al. Prediction of effluent quality in a wastewater treatment plant by dynamic neural network modeling. Process Saf Environ Prot. 2022;158:515–24.
[41] Song S, Xiong X, Wu X, Xue Z. Modeling the SOFC by BP neural network algorithm. Int J Hydrog Energy. 2021;46(38):20065–77.
[42] Mahanty T, Bhattacharjee S, Goswami M, Bhattacharyya P, Das B, Ghosh A, et al. Biofertilizers: a potential approach for sustainable agriculture development. Environ Sci Pollut Res. 2017;24:3315–35.
[43] Köksal Ö, Tekinerdogan B. Architecture design approach for IoT-based farm management information systems. Precis Agric. 2019;20:926–58.
[44] Tian Y, Rao H, Zhang K, Tao S, Xue WT. Data security and privacy-preserving in edge computing paradigm: Survey and open issues. IEEE Access. 2018;6:18209–37.
[45] Khan WZ, Ahmed E, Hakak S, Yaqoob I, Ahmed A. Edge computing: A survey. Future Gener Comput Syst. 2019;97:219–35.
[46] Chen B, Wan J, Celesti A, Li D, Abbas H, Zhang Q. Edge computing in IoT-based manufacturing. IEEE Commun Mag. 2018;56(9):103–9.
[47] Gochhayat SP, Kaliyar P, Conti M, Tiwari P, Prasath VB, Gupta D, et al. LISA: Lightweight context-aware IoT service architecture. J Clean Prod. 2019;212:1345–56.
[48] Wang T, Cao Z, Wang S, Wang J, Qi L, Liu A, et al. Privacy-enhanced data collection based on deep learning for internet of vehicles. IEEE Trans Ind Inform. 2020;16(10):6663–72.
[49] Franco P, Martinez JM, Kim YC, Ahmed MA. IoT based approach for load monitoring and activity recognition in smart homes. IEEE Access. 2021;9:45325–39.
[50] Park DS. Future computing with IoT and cloud computing. J Supercomput. 2018;74:6401–7.
[51] Cardenas AM, Nakamura Pinto MK, Pietrosemoli E, Zennaro M, Rainone M, Manzoni P. A low-cost and low-power messaging system based on the LoRa wireless technology. Mob Netw Appl. 2020;25:961–8.
[52] Liao CH, Zhu G, Kuwabara D, Suzuki M, Morikawa H. Multi-hop LoRa networks enabled by concurrent transmission. IEEE Access. 2017;5:21430–46.
[53] Zhang X, Cao Z, Dong W. Overview of edge computing in the agricultural internet of things: Key technologies, applications, challenges. IEEE Access. 2020;8:141748–61.
[54] Sodhro AH, Pirbhulal S, Luo Z, de Albuquerque VHC Towards an optimal resource management for IoT based green and sustainable smart cities. J Clean Prod. 2019;220:1167–79.