Internet of Things Technology in Ecological Security Assessment System of Intelligent Land

ZHAOYA CHEN[1], **LIJUN MA**[2], **FENGYU GUO**[3], **AND YAHENG CHEN**[2]
[1]College of Resources and Environment Science, Hebei Agricultural University, Baoding 071000, China
[2]College of Land and Resources, Hebei Agricultural University, Baoding 071000, China
[3]Urban and Rural Construction Institute, Hebei Agricultural University, Baoding 071000, China

Corresponding authors: Fengyu Guo (xyzz0312@126.com) and Yaheng Chen (chenyaheng@126.com)

ABSTRACT The ecological security of land is related to the sustainable development of human beings. With the continuous progress of science, the evaluation standard of land ecological security is also changing. The traditional safety evaluation method is time-consuming and laborious, with high cost of data acquisition, small amount of data and large error of data itself. The establishment of intelligent land ecological security evaluation system through the Internet of things is conducive to a better understanding of the overall land situation in the region. Taking the land data of Cangzhou City, China as an example, this study uses the Internet of things technology to establish a set of intelligent evaluation and grading model. The random forest algorithm is used to evaluate the land ecological security, and combined with the existing data to predict the land ecological security in 2020-2025. Compared with other security evaluation methods, the intelligent land ecological security evaluation system in this study has the advantages of short operation time, low energy consumption and high accuracy. This method has a strong guiding significance for the future land ecological security management.

INDEX TERMS Internet of things, intelligent land, random forest algorithm, short operation time, Cangzhou, evaluate.

I. INTRODUCTION

As the basic environment for human survival activities, the land provides abundant resources. Among them, the quality and quantity of land resources restrict China's ecological security and economic development [1]. However, with the progress of the society, the population base keeps expanding. At the same time, in order to pursue high economic growth rate, human beings have developed land resources to different degrees. Unreasonable development methods have led to the reduction of green space, desertification and salinization of land. A series of land ecological problems, such as soil erosion, reduce the function of the land ecosystem and the ability to deal with and resist ecological risks, causing losses to the national economy and affecting the sustainable development of the society [2]–[4]. Therefore, it is of great significance to the ecological security evaluation of land.

In this context, the Internet of things technology for land ecological security evaluation came into being. The Internet

The associate editor coordinating the review of this manuscript and approving it for publication was Chun-Wei Tsai.

of things is the global infrastructure of the information society, and capable of connecting physical and virtual objects to each other to provide advanced services based on existing and evolving interoperable information and communication technologies. The Internet of things has identification, data collection, processing, and communication capabilities to make full use of objects to serve a wide range of applications, while ensuring necessary privacy. The goal is to realize all kinds of objects, including physical and virtual objects, to contact, interact, and communicate with each other through various networks [5]–[7]. There are various types of sensors and RFID devices in the Internet of things, and these sensors have different protocols, different structure, and performance, and their data structures are also different [8]. Data formats transmitted between modules of the Internet of things system also include text data, or multimedia data, such as picture format, video format, and so on. There are static data, and streaming data. Data polymorphism and heterogeneity of sensing devices lead to data heterogeneity [9], [10]. A large number of heterogeneous characteristics greatly increase the workload and difficulty of data analysis and processing and

system development. Currently, the popular database management mode and system are difficult to deal with the massive heterogeneous data from multiple sources and face great problems in the process of storage, processing, and analysis [11]. The explosion of information and the emergence of massive data processing force the traditional database to innovate in technology. At the application level, massive data processing has become the basis of information scientific discovery and research. In terms of technology, mass data processing will be a revolution of traditional database technology and an inevitable trend of the development of Internet of things and cloud computing. It is necessary to introduce new computing technology to filter and clean the data and conduct data analysis and mining [12], [13]. The traditional database has many shortcomings in real-time access analysis, recovery and backup of massive data, and effective data mining. Massive data storage also faces constraints such as energy consumption and space limitations [14]. Random forest algorithm is a typical classification algorithm, proposed by professor Breiman, academician of the National Academy of Sciences, in 2001. It has good classification and prediction performance in biological, chemical, agricultural, and other fields [15]. Due to the randomness of random forest, sample number, and attribute number, it is very suitable to select parallelization method to establish multiple decision trees. In the classification process of Internet of things data processing, the parallel random forest algorithm based on Hadoop can be used to mine and classify data [16], [17]. At the same time, there are still few studies on the application of random forest algorithm in the land ecological security evaluation and prediction. The random forest algorithm is used to mine the data effectively, which has high research value for the evaluation of land ecological security.

Under such a background, the Internet of things technology for land ecological security evaluation came into being. Based on this, the Cangzhou city is taken as the research area to build the land ecological security evaluation index system. Random forest algorithm is designed and applied in the evaluation of land ecological security to analyze the current situation of land in Cangzhou city and predict its future trend.

II. LITERATURE REVIEW
A. RESEARCH ON LAND ECOLOGICAL SECURITY EVALUATION

As an important part of the whole ecological environment, many researchers have evaluated and analyzed the ecological security of land. Reference [18] took 11 prefecture-level cities in Shanxi Province as evaluation units and built a comprehensive evaluation index system based on PSR model. The entropy weight method and the grey relational evaluation model were integrated to evaluate the comprehensive index, pressure index, state index, and response index of ecological security in the research area from 2005 to 2015, and the spatial and temporal evolution was analyzed. The results showed that: from 2005 to 2015, the comprehensive index of ecological safety in the study area showed an overall upward trend, and the safety level changed from level II and III to level IV and level V. The regional ecological environment has improved significantly, but the trend has slowed down, and the pressure caused by human activities on the ecological environment was generally small and has decreased [18]. Based on the PSFR model, [19] selected 16 sub-indicators and 34 evaluation indicators to establish a 4-level indicator system. Considering the entirety and difference of Luanhe river basin, the Luanhe river basin in Hebei province was divided into six assessment units to carry out ecological safety assessment according to the environmental characteristics and functional zoning. The results showed that the ecological security level of the whole basin was between grade III and grade II [19]. Reference [20], based on the PSR model, constructed an evaluation system of land ecological security in Lianyungang city from three aspects: pressure, state, and response. The evaluation and analysis of land ecological security in Lianyungang city from 2000 to 2016 were conducted with matter-element analysis method. The results showed that the ecological security level of land in Lianyungang city has increased from ''unsafe'' to ''relatively safe'' during the study period. The trend of overall security level was getting better, but the state of ''relatively safe'' was unstable, and the ecological security of land was not optimistic [20].

B. APPLICATION OF RANDOM FOREST ALGORITHM

Reference [21] applied random forest algorithm to seismic reservoir prediction and established a regression model between random forest seismic properties and porosity parameters. The results showed that the random forest model had strong robustness and tolerance to abnormal data different from the sample data. The key problem in the application of random forest algorithm in earthquake reservoir prediction was that no noise was included in the training data [21]. Reference [22] established a random forest model using the air quality monitoring network to predict the maximum 8-hour daily average ozone concentration in China in 2015 ($[O_3]$ MDA8). Compared with the chemical transport model which required a large number of variables and expensive calculation, the random forest model had a comparable or higher prediction performance based on only a few readily available variables, but the calculation cost was much lower. The results showed that the national population-weighted $[O_3]$ MDA8 was estimated to be $84 \pm 23 \mu g/m^3$ per year, with the highest seasonal average in summer ($103 \pm 8 \mu g/m^3$) [22].

To sum up, the PSR model is widely used in the evaluation of land ecological security, and good evaluation effects have been obtained. However, these studies only evaluate the local ecological security in a certain period of time, and do not predict the future land ecology based on the evaluation results. As a kind of data mining algorithm, random forest algorithm can process high-latitude data with more attribute values, has higher prediction accuracy, low cost, and fast training speed compared with other algorithms. However, the application of it in the evaluation of land ecological security is still insuf-

ficient. In order to fill the gap, based on the evaluation and analysis of the current situation of land ecology in Cangzhou city, the random forest algorithm is used to predict its future development trend, so as to provide a theoretical reference for the field of land ecology research.

III. LAND AND ECOLOGICAL SECURITY EVALUATION INDEX AND DATA MINING ALGORITHM DESIGN

A. OVERVIEW OF STUDY AREA

Cangzhou is located in the north China plain, facing Beijing and Tianjin in the north and Bohai sea in the east. Its geographical coordinates are 37°29 'to 38°57' N, 115°42 'to 117°50' E. It covers an area of 13,420,000 km2 and has a total population of 7.8 million. The region belongs to the impact topography of Huang-Huai-Hai sea, with flat terrain and low fluctuation, and the altitude is between 0 and 17 m. Located in the mid-latitude coastal area on the east coast of Eurasia, the temperate monsoon climate is formed by climate circulation and thermal effects. In the same period of hot winter and hot summer and hot rain, the precipitation concentration degree is strong, the annual average temperature is 12.5 °C, and the average annual precipitation is between 500 mm and 600 mm. Cangzhou water system belongs to the outflow area of Bohai sea, the river is short, the ratio is lower, and there is spring flood and summer flood. The city has an average annual land volume of 2.85 billion m3, including 1.462 billion m3 of surface runoff and 1.369 billion m3 of underground land. However, the per capita land volume is only 385 m3, less than 1/6 of the national average level, which is a typical water shortage city. As the coordinated development of the Beijing-Tianjin-Hebei region, ''One Belt And One Road'', cooperation and development in the circum Bohai sea region, such as superposition of important strategic opportunities, Cangzhou coastal, location, transportation, industry and other comprehensive advantages, healthy and rapid economic development, the strategic position of Hebei province significantly increased, and gradually become an important coastal open city of ohai rim region. However, with the promotion of regional urbanization and industrialization in recent years, the population of the city has been increasing year by year, the water wetland has been expropriated, and the large-scale industrial base is backward, which on the one hand causes the ecological environment of the land in the region to be destroyed. Taking Cangzhou as the research area, this research aims to clarify the spatial and temporal distribution characteristics of regional land and comprehensively evaluate the land status, so as to provide basic information for regional land management and economic structure adjustment.

B. EVALUATION INDEX OF LAND ECOLOGICAL SECURITY

The index system can fully reflect the regional ecological security, which must be based on the structural complexity of the ecosystem. In order to better analyze the impact of human activities on the ecology, three models of pressure-state-response (PSR) based on the connotation of ecological
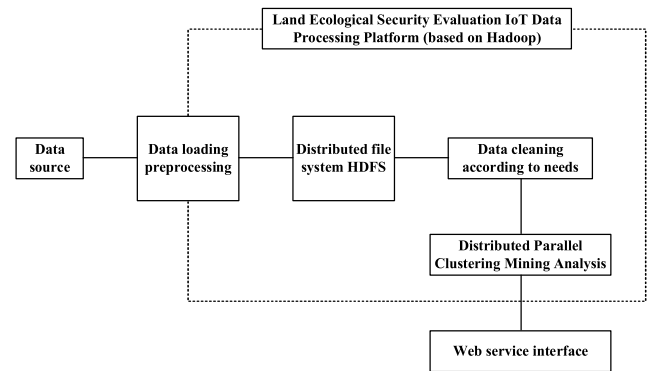

FIGURE 1. Land ecological security evaluation data processing process of Internet of things.

security were established to better and more directly and comprehensively analyze the problems of land ecological security and the relationship between economic and social development and resources, environment and ecological security. The PSR model includes three models: ecological pressure, ecological state and ecological response. The data obtained from the three models determine the evaluation index. According to the PSR conceptual framework, Cangzhou city's land ecological security evaluation index system is constructed from the aspects of population, economy, resources and environment, etc., as shown in table 1.

The evaluation standard of land ecological security is the basis of quantitative and graded evaluation. At present, there is no unified understanding at home and abroad. Based on the research experience of relevant scholars [23], the evaluation standard of land ecological security is divided into five levels: safe, relatively safe, critical safe, less safe and unsafe, and the range of the corresponding standard magnitude is given, as shown in table 2.

C. DATA MINING ALGORITHM DESIGN OF LAND ECOLOGICAL SECURITY EVALUATION

The data processing process of Internet of things for land ecological security evaluation is shown in Figure 1. Hadoop distributed framework is adopted for distributed data processing. First, the collected data are loaded and processed to convert it into data in a uniform format. The distributed file system HDFS is used to store a large amount of data. Hive distributed data warehouse and MapReduce framework are adopted to clean and filter the data according to the demand. The redundant information is deleted, and finally the stored data is mined to provide efficient and predictable services for land ecological security evaluation.

In the Internet of things of land ecological security evaluation, the data that needs to be mined is huge and grows exponentially. Therefore, the extended data mining algorithm is usually used to process large-scale data and improve the operation speed of the system. The MapReduce model framework was used to process and analyze data on the basis of Hadoop processing platform. According to the characteristics

TABLE 1. Candzhou land ecological security evaluation index system.

Rule hierarchy		Marker bed	Index calculation	Index properties
Pressure system P	Population pressure	A1 population density (person /km²)	Total population/area	-
		A2 natural population growth rate (‰)	——	-
	Economic and social pressure	A3 agricultural economic structure (%)	Agricultural GDP/ total GDP	-
		urbanization level (%)	——	+
		raw coal consumption level (%)	Total coal use/energy consumption	-
		land pesticide load (kg/ mu)	Pesticide use/arable area	
		A7 energy consumption per unit GDP (ten thousand yuan/ton)	——	-
State system S	State of ecological resources	A8 forest coverage (%)	——	+
		A9 water yield modulus (10,000 m³/km²)	Total water resources/area	
		A10 natural disaster area ratio (%)	Affected area/area coverage	-
	Cultivated land quality state	A11 water and soil coordination (%)	Irrigated area/total cultivated area	
		A12 drought and flood protection yield (%)	Drought-waterlogging conservative area/total cultivated area	+
		A13 per capita arable land area (mu)	The representative quantity of cultivated land resources	+
		A14 output value of cultivated land (kg/hm²)	Grain output/arable land area	+
Response system R	Agricultural science and technology response	A15 power level of agricultural machinery	Total power of farm machinery/arable land area	+
	Environmental governance response	A16 comprehensive utilization rate of industrial solid waste	——	+
		A17 sewage treatment rate (%)	Sewage treatment/sewage discharge	+
		A18 soil erosion control rate (%)	Soil erosion control/total area of soil erosion	+
	Ecological protection response	A19 environmental pollution control investment ratio (%)	——	+
		A20 area ratio of nature reserve (%)	Nature reserve area/area coverage	+
		A21 wetland water area ratio (%)	Wetland water area/area coverage	

TABLE 1. *(Continued.)* Candzhou land ecological security evaluation index system.

Note: the safer the ecology is, the larger the positive index is.

of the existing data mining algorithm, whether it can be MRed so as to carry out distributed programming was analysed to improve the computing efficiency.

Decision tree is an important concept in random forest because unpruned decision tree is the base classifier of random forest and decision tree is the basic unit of random forest. Many decision trees together constitute the random forest, and the collection of different decision trees adopts the random way, so these decision trees become the random decision trees. In a random forest, there's no connection between each random decision tree. When the data to be tested is introduced into the random forest, it is essentially to classify the random decision tree and finally take the most output results obtained from the decision tree as the final results.

Decision tree belongs to the category of single classifier, and its production process can be summarized into three steps: first, it needs to use recursive analysis to analyze the samples and get the tree structure which is like reverse growth. Then, the paths of each part of the decision tree are analyzed to obtain a series of rules. The third step is to make the type distinction and correlation prediction of data and samples according to the rules. The core idea of decision tree classification is to generate classification rules through path analysis, and follow the classification rules to conduct data mining on samples and related data.

Decision tree uses single classifier to classify and make decision. The following problems exist in decision tree classification.

I. Cumbersome classification rule. Due to the use of local greed method, the decision tree only takes one attribute as the basis to establish the decision tree in the process of rule classification. Therefore, the classification rules related to the decision tree may be very complicated, which requires the optimization of the decision tree, and pruning is usually adopted to achieve the simplification of the decision tree.

II. The possible result is a locally optimal solution. Some decision tree algorithms lack the process of multiple selection and comparative analysis in the retrieval process. Therefore, some decision trees may have local optimal solutions instead of one or more overall optimal solutions.

TABLE 2. Land ecological security evaluation grading standards.

Grading	A1	A2	A3	A4	A5	A6	A7
Safe	<30	<5	<5	>70	<10	<3	<0.2
Relatively safe	[30,70)	[5,10)	[5,10)	[50,70)	[10,20)	[3,8)	[0.2,0.5)
Critically safe	[70,150)	[10,20)	[10,15)	[30,50)	[20,40)	[8,12)	[0.5,0.7)
Relatively not safe	[150,300)	[20,50)	[15,25)	[10,30)	[40,60)	[12,20)	[0.7,1)
Not safe	≥300	≥50	≥25	≤10	≥60	≥20	≥1

Grading	A8	A9	A10	A11	A12	A13	A14
Safe	>55	>100	<5	>80	>80	>3	>5000
Relatively safe	[35,55)	[70,100)	[5,10)	[60,80)	[60,80)	[2,3)	[3500,5000)
Critically safe	[20,35)	[40,70)	[10,20)	[40,60)	[40,60)	[0.8,2)	[2000,3500)
Relatively not safe	[10,20)	[10,40)	[20,30)	[20,40)	[20,40)	[0.4,0.8)	[1000,2000)
Not safe	≤10	≤10	≥30	≤20	≤20	≤0.4	≤1000

Grading	A15	A16	A17	A18	A19	A20	A21
Safe	>45	>90	>90	>85	>5	>25	>15
Relatively safe	[30,45)	[70,90)	[70,90)	[65,85)	[4,5)	[15,25)	[10,15)
Critically safe	[10,30)	[50,70)	[50,70)	[40,65)	[2,4)	[10,15)	[5,10)
Relatively not safe	[5,10)	[20,50)	[20,50)	[20,40)	[1,2)	[5,10)	[3,5)
Nor safe	≤5	≤20	≤20	≤20	≤1	≤5	≤5

III. The existence of overfitting problem. In the classification and decision-making process of decision tree, due to the variety of classifiers and rules, overfitting may occur due to the excessive influence of noise in the learning process of decision tree.

In order to make up for the shortcomings of decision tree, a stochastic forest algorithm is proposed. Random Forest (RF) establishes a forest in a random way. The Forest here is composed of a large number of decision trees, and there is no correlation between each decision tree. After the establishment of the forest, the new input sample data was used for prediction, judgment and classification. Based on the decision tree established by the random forest in the previous stage, the classification of the sample data is determined. The number of choices for each category is calculated. The category with the largest number of choices is the predicted category.

Random forest is a fusion algorithm integrating the formation of N $\{p(x, \Theta_n), k = 1, 2, \ldots\ldots, N\}$ classification regression trees. Based on random subspace theory and self - aggregation method, random vector $(X, Y)\{\Theta_n, k = 1, 2, \ldots\ldots, N\}$ is selected randomly. Let X and Y be input and output vectors in the random subset of independent random vectors (X, Y), respectively. For the output p(h) of the prediction sample, there is a generalization error:

$$E_{X,Y}[Y - p(x)]^2 \tag{1}$$

The random forest output is based on the mean value of N regression trees $\{p(\Theta, X_n), k = 1, 2, \ldots\ldots, N\}$. When $k \to \infty$, then there is:

$$E_{X,Y}\left[Y - \bar{p_k}(X, \Theta_k)\right]^2 \to E_{X,Y}[Y - E_\Theta(X, \Theta_k)]^2 \tag{2}$$

Its generalization error (RE) is:

$$RE = E_\Theta E_{X,Y}[Y - p(X, \Theta)]^2 \tag{3}$$

For all single trees, the average generalization error is:

$$RE \leq \bar{r}\,RF \tag{4}$$

In the equation, $\bar{r}$ is residual, and the individual trees Θ are relatively independent.

The value of the attribute variable of the node after it forms the decision tree is generated from several randomly selected attribute subsets. For the samples to be tested, the random forest allows each tree to vote through self-help aggregation, and the highest category of votes is the output result, namely:

$$P(x) = \arg\max \sum^{k} I[p_i(x) = Y] \tag{5}$$

In the equation, P(x) is the result of the random forest combination model, pi is the single tree classification model, and I is the indicator function.

In the classification process of Internet of things data processing for land ecological security assessment, the Hadoop-based random forest algorithm can be used to classify and process the data, and its algorithm process is shown in Figure 2. In this algorithm, the input data is preprocessed, the decision tree is established and stored in HDFS. The data is aggregated and the prediction results are output. Compared with other algorithms, the Hadoop-based random forest algorithm can process data with more attribute values, so that relatively important attributes can be obtained after data training instead of pruning each tree.

1) DATA PRE-PROCESSING

For the input Data sample generate Data and Dataset, where Data records every Data point, and Dataset records Data format. The data structure is composed of five data types: ignored represents the attribute value ignored
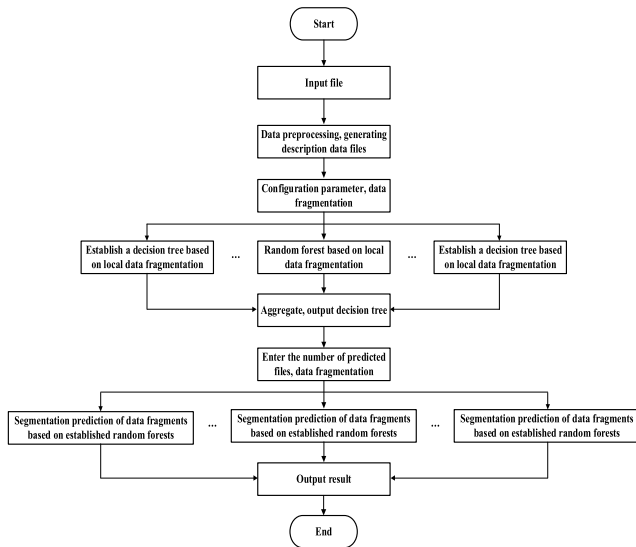
FIGURE 2. Hadoop-based stochastic forest algorithm flow chart.
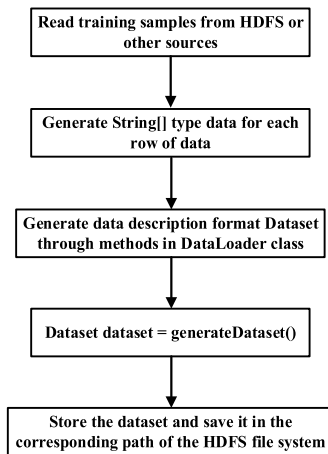

FIGURE 3. Data preprocessing.

by the classification, values represents the specific value of the classification type of the sample, nbinstance represents the data amount of the input sample, attribute represents the saved sample data type, and labelld represents the saved specific value of the specific data type in the sample. The process of data preprocessing is shown in Figure 3.

2) BUILDING RANDOM FOREST

Step 1: The Decision TreeBuilder type object is created and the corresponding parameters are set such as the input path. The number of attributes m and the parallel configuration parameter conf are selected, and the description file generated in the previous stage through the distributed cache file is read to parse the input sample file.

Step 2: Creation of a random forest.

This step is mainly to use the MR programming model to transform the data of each shard in the map function and build a tree in the clean function. The relevant data structure and
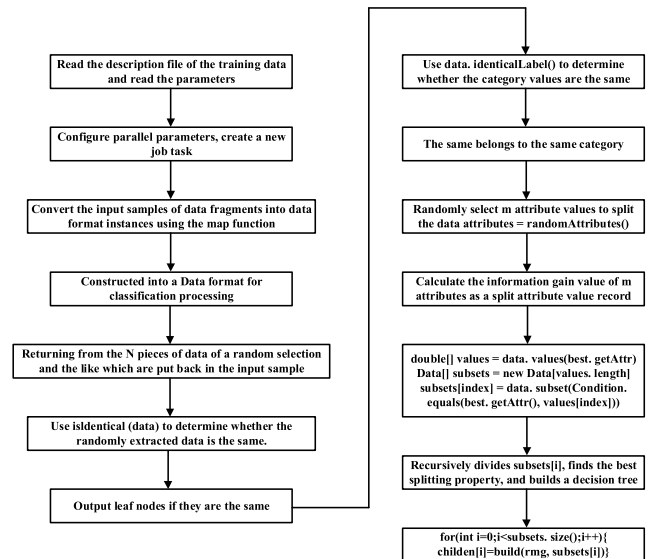

FIGURE 4. Establishment of the data structure and function process of decision tree.

function process of distributed decision tree establishment are shown in Figure 4:

A: Reading the description file of training data and parameters, such as the number of trees built, randomly selected attribute trees, configuring parallel parameters, and creating new job tasks.

B: The map function transforms the input sample of this data sharding into instances and constructs the Data format for classification processing.

C: N pieces of data of equal size are randomly selected from the input samples to be put back.

D: isIdentical(data) is used to determine whether these randomly selected data have the same attribute values. If they are the same, then the leaf node is output.

E: Judginf the category. If the category value of all data under the branch is the same, then it is judged as leaf node. The data.identicallabel () method is used for judgment. if eligible, the data belong to the same category.

F: m attribute values are randomly selected and the data are split attributes = randomAttributes The information gain values of these m attributes are calculated respectively and recorded as split attribute values.

G: The value range of the property selected in step f is obtained and the data are divided into several parts according to the property.

double[] values = data.values(best.getAttr);

Data[] subsets = new Data[values.length];

subsets[index] = data.subset(Condition.equals(best.getAttr, values[index]));

h: The subsets[I] are divided recursively to find the best split properties to build the decision tree.

for(int i = 0; i < subsets.siza(); i++){

childen[i]= build(rng,subsets[i])

}

Step 3: The random forest created by several random trees created following the steps above are written to HDFS to be saved.

3) TEST PREDICTION CLASSIFICATION

Step 1: The parameters should be configured, and the input file and algorithm path are defined.

Step 2: MapReduce framework is adopted for prediction. First, the test data is transformed, and all trees are traversed, and the corresponding leaf nodes are traversed in each decision tree to get the classification results.

Step 3: The corresponding results are output and the accuracy is calculated.

D. EXPERIMENTAL ENVIRONMENT

In this study, the Hadoop distributed cluster composed of four machines is used as the data processing platform for testing the Internet of things. A node is used as the Master Master node of HDFS and MapReduce, namely NameNode and JobTracker node, which is mainly responsible for metadata management and task scheduling of distributed data processing platform. The other nodes are Slave working nodes, namely DataNode and TaskTracker, which are mainly responsible for data storage and specific distributed computing processes.

IV. RESULTS AND CONCLUSION

A. DATA RESOURCES

The research data are mainly from China statistical yearbook (1997-2017). Due to the limitation of data availability, relevant data are also drawn from China urban statistical yearbook (1997-2017), China rural statistical yearbook (1997-2017) and Cangzhou statistical yearbook (1997-2017).

B. TEAT RESULTS OF RANDOM FOREST ALGORITHM

In order to study the performance of the random forest algorithm, accuracy, average recall, iteration times and running time were used to analyze and compare the random forest algorithm in this study. Where, accuracy represents the ratio of the number of correct clustering data to all data, and average recall represents the average of the ratio of the number of correctly divided into various categories to the data objects. The formula of accuracy and average recall is as follows.

$$\text{Precision} = \frac{N_r}{N_a} \times 100\% \qquad (6)$$

$$\text{Recall} = \frac{1}{k} \sum_{a=1}^{k} \frac{N_{a \to a}}{N_{a \to a} + N_{a \to b}} \times 100\% \qquad (7)$$

The sample data is divided into training samples and detection samples to test the algorithm. The size of the two sample data is 1 GB, and the sample size is 10 million. Under the same training sample and detection sample, the random forest algorithm and decision tree algorithm are compared with SVR [24] and generalized neural network algorithm [25].

TABLE 3. Data classification results comparison.

Sorting algorithm	Decision tree	Random forest	SVR	functional networks
Accuracy rate (%)	89.72	96.53	95.78	94.62
Average recall (%)	90.18	97.02	96.11	96.06
Training time (s)	468	352	379	394
Test time (s)	230	174	182	202

TABLE 4. Data classification results comparison.

Sorting algorithm	Random forest	RNN	LSTM
Accuracy rate (%)	96.53	54.38	99.32
Average recall (%)	97.02	98.76	98.63
Time consuming (s)	3.36	289.61	735.91

SVR maps data to a higher dimensional space through nonlinear mapping, where linear regression can be carried out. The algorithm of finding the optimal linear regression hyperplane is summarized to solve a convex programming problem with a convex constraint, and the global optimal solution is obtained. However, it has shortcomings such as slow training speed, large training error, and poor generalization ability [26].

It can be concluded from the data in table 3 that the random forest algorithm has the shortest running time, the highest accuracy, the highest average recall rate, and the best performance.

Since the data used in this study belongs to the time series, on this basis, the performance of the random forest algorithm is compared with the cyclic neural network model (RNN) and the long- and short-term memory network model (LSTM). The results are shown in table 4. The accuracy and average recall rate of LSTM are higher than that of random forest algorithm, but the difference between the two is relatively small, while the accuracy of RNN is low and the overall performance is relatively poor. At the same time, although the accuracy and average recall rate of random forest algorithm are slightly lower, its time consumption is the shortest, which greatly saves the running time of the algorithm [27].

C. EVALUATION RESULTS OF LAND SECURITY

The land ecological security level is jointly affected by the index factors in all dimensions of PSR. The key to applying the random forest algorithm to evaluate ecological security is to construct the membership rules between the ecological security level and a single index in the indicator system according to the evaluation classification standard. The specific evaluation process is as follows: step 1: Each index was randomly interpolated into 200 groups of sample data according to the classification standard interval, and the total
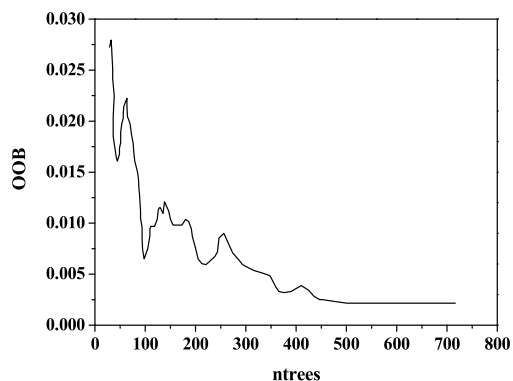
FIGURE 5. The accuracy of stochastic forest algorithm varies with the number of trees.

sample size of 5 evaluation grades was 1000 groups. Step 2: sample settings. 100 groups of data were randomly selected from 5 graded sample groups, and 500 groups of data were set as training samples and the remaining 500 groups as test samples. The index factors in the sample are input vectors, and the evaluation results I, II, III, IV and V corresponding to the classification standard are represented by the Numbers 1, 2, 3, 4 and 5 successively, and taken as output vectors, and then random forest network training is carried out. Step 3: parameter optimization and model optimization.

In the process of model training, sensitive parameters mtry and ntree can be selected according to the size of OOB error outside the bag. In the random forest algorithm, mtry is a category of split attributes, usually set as the square root value of the variable. In this study, the total number of variables is 20, and its rounding is 4 or 5. It has been verified by multiple experiments that the evaluation result is better when mtry is 5. Ntree is the number of trees in the random forest algorithm, which affects the calculation rate and accuracy of the model. As shown in Figure 5, when its value is greater than 500, OOB is small and stable. Step 4: indicator measurement. The random forest algorithm can determine the importance of an index by the change of OOB after adding or subtracting an index factor. Samples that are early enough are generated randomly within the classification interval without loss of generality and can avoid noise pollution, so the measurement of the importance of ecological security evaluation index is more objective [28]. Step 5: model evaluation. The performance of random forest was evaluated according to the output value and actual value generated after the model training, as shown in table 5. It can be concluded that the Re of the training sample and the detection sample are close to 1, and MSE and RMSE are close to 0. The model has high fitting accuracy and good generalization ability, which can be applied to the ecological security evaluation of the target region. Step 6: threshold setting and model application. According to the critical value of each index classification in table 2, random forest algorithm is used for simulation calculation, and its simulation value is taken as the standard threshold of different classification. The thresholds are shown in table 6.
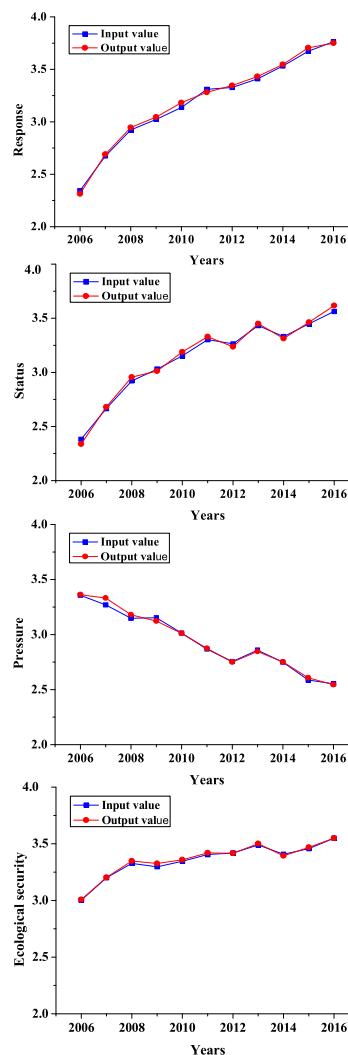

FIGURE 6. Study of ecological security prediction model in Cangzhou city.

TABLE 5. Accuracy of stochastic forest algorithm.

Dimension	Samples	R^2	MSE	RMSE
Pressure	Training	0.9990	0.0672	0.0132
	Detection	0.9983	0.0912	0.0241
State	Training	0.9972	0.1023	0.0635
	Detection	0.9981	0.0841	0.0349
Response	Training	0.9988	0.0693	0.0578
	Detection	0.9993	0.0551	0.0205
Ecological safety	Training	0.9977	0.1001	0.0733
	Detection	0.9985	0.0712	0.0542

D. LAND SAFETY FORCAST RESULTS

Ecological security is restricted by many different factors. The reason why it appears randomness is that it has changed
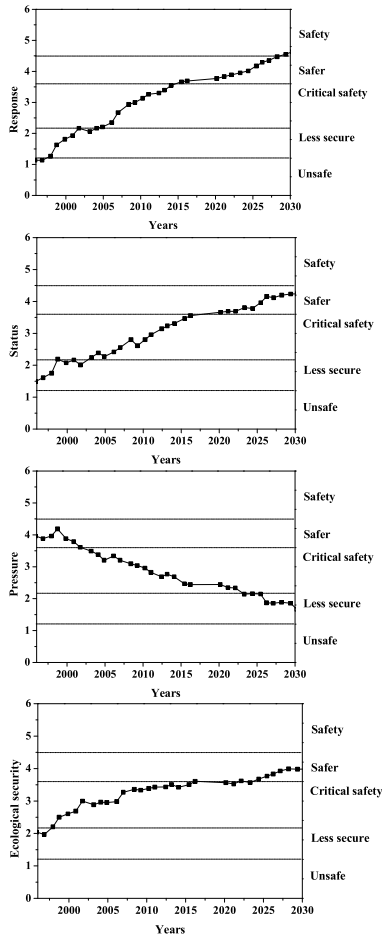
FIGURE 7. Evaluation and prediction results of land ecological security in Cangzhou city.

TABLE 6. Ecological security index classification.

Gradin g	Pressure	State	Response	Evaluation of ecological safety
Safe	(4.5301,5.2 681]	(4.2152,5.3 765]	(4.3624,5.4 012]	(4.5214,5.2 625]
Relativ ely safe	(3.1628,4.5 301]	(3.6707,4.2 152]	(3.5155,4.3 624]	(3.2071,4.5 214]
Critical ly safe	(2.2415,3.1 628]	(2.3721,3.6 707]	(2.2081,3.5 155]	(2.1620,3.2 071]
Relativ ely not safe	(1.3746,2.2 415]	(1.4686,2.3 721]	(1.5324,2.2 081]	(1.4011,2.1 620]
Not safe	(0,1.3746]	(0,1.4686]	(0,1.5324]	(0,1.4011]

in a certain space and time range. Therefore, to predict ecological security, its rules must be obtained first. According to the random forest regression principle, the land ecological pressure, state, response index and ecological security index of Cangzhou city from 1996 to 2016 were set as training samples to predict the security index of land

ecological pressure, state and response of Cangzhou city from 2020 to 2030. As shown in Figure 6, the single prediction step size was set as 1. The design program of Matlab2016b is applied to implement simulation and prediction. After a lot of experiments, and the embedded dimension was determined to be 10 (that is, the index of the first 10 years was used to predict the index of the next year), and the optimal parameters in each prediction model were selected. According to the calculation, the precision of the model is all within the acceptable range, so the land ecological security in Cangzhou city from 2020 to 2025 is predicted.

Through the application of the previous evaluation method for land ecological security, various aspects of land ecological security data were obtained. The main research years were land pressure in Cangzhou city from 1996 to 2016 and from 2020 to 2030, as shown in Figure 7. It can be concluded that the data of response, state and pressure in this decade ranged from 1 to 4, with the characteristics of instability, and evolved from a safer level to a less safe level. This is due to the rapid development of Cangzhou city since the new century, a large number of land expropriation, land ecological environment deterioration, forest coverage, agricultural output per unit area are low, so there is instability.

According to the prediction results, the pressure index tends to decrease in the next few years, and the resulting land ecological security will also face great pressure. The area of returning farmland to forest will be increased, and the area of saline and alkaline land will be reduced.

V. CONCLUSION

In this research, Cangzhou city is taken as the research area, land ecological security evaluation index system is constructed, and random forest algorithm is designed. It is compared with the decision tree algorithm, and the running time, accuracy and average recall of the algorithm are studied and applied to land ecological security evaluation. The present situation of land in Cangzhou city is analyzed and its future trend is predicted. Compared with decision tree algorithm, SVM regression prediction algorithm, and generalized neural network algorithm, stochastic forest algorithm has short running time, high accuracy and average recall. Moreover, the algorithm has high fitting accuracy and good generalization ability, and can be applied to ecological security evaluation of the target region. The land ecology of Cangzhou city is unstable from 1996 to 2016. The land ecology prediction of Cangzhou city using this algorithm in the next 10 years shows that the pressure index tends to decrease, the area of returning farmland to forest will be improved, and the area of saline and alkaline land will decrease. The algorithm has simple model, fast learning rate, high tolerance for dimensionality and data noise, and can exclude the interaction between internal dimensions. The effects of the out-of-pocket error object model are evaluated. Its shortcoming is that it is not sensitive to attribute response with more values, which can be considered in future research.

REFERENCES

[1] J. Lin, G. Song, and Y. Zhang, "Regionalization planning for prevention and control of saline-alkali land based on a landscape ecology risk pattern theory: A case study in Lindian county, Heilongjiang province," *Acta Ecologica Sinica*, vol. 38, no. 15, pp. 5509–5518, 2018.

[2] C. Zhang, M. Qiao, and W. Yun, "Trinity comprehensive regulatory system about quantity," *Qual. Ecol. Cultivated Land.*, vol. 48, no. 1, pp. 1–6, 2017.

[3] J. D. Spencer, A. Brunelle, and T. Hepola, "Using paleoecology to inform land management as climates change: An example from an oak savanna ecosystem," *Environ. Manage.*, vol. 60, no. 6, pp. 1090–1100, Dec. 2017.

[4] C. Radel, B. Schmook, L. Carte, and S. Mardero, "Toward a political ecology of migration: Land, labor migration, and climate change in northwestern nicaragua," *World Develop.*, vol. 108, pp. 263–273, Aug. 2018.

[5] A. Osseiran, O. Elloumi, J. Song, and J. F. Monserrat, "Internet of Things," *IEEE Commun. Standards Mag.*, vol. 1, no. 2, p. 84, Jul. 2017.

[6] P. Ramezani and A. Jamalipour, "Toward the evolution of wireless powered communication networks for the future Internet of Things," *IEEE Netw.*, vol. 31, no. 6, pp. 62–69, Nov. 2017.

[7] J. Iannacci, "Internet of Things (IoT); Internet of everything (IoE); tactile Internet; 5G—A (not so evanescent) unifying vision empowered by EH-MEMS (energy harvesting MEMS) and RF-MEMS (radio frequency MEMS)," *Sens. Actuators A, Phys.*, vol. 272, pp. 187–198, Apr. 2018.

[8] H. Hellaoui, M. Koudil, and A. Bouabdallah, "Energy-efficient mechanisms in security of the Internet of Things: A survey," *Comput. Netw.*, vol. 127, pp. 173–189, Nov. 2017.

[9] S. Shadroo and A. M. Rahmani, "Systematic survey of big data and data mining in Internet of Things," *Comput. Netw.*, vol. 139, pp. 19–47, Jul. 2018.

[10] V. Losing, B. Hammer, and H. Wersing, "Tackling heterogeneous concept drift with the self-adjusting memory (SAM)," *Knowl. Inf. Syst.*, vol. 54, no. 1, pp. 171–201, Jan. 2018.

[11] C. Fan, F. Xiao, and C. Yan, "Research and applications of data mining techniques for improving building operational performance," *Current Sustain./Renew. Energy Rep.*, vol. 5, no. 2, pp. 181–188, Jun. 2018.

[12] P. Rezaei-Hachesu, A. Oliyaee, N. Safaie, and R. Ferdousi, "Comparison of coronary artery disease guidelines with extracted knowledge from data mining," *J. Cardiovascular Thoracic Res.*, vol. 9, no. 2, pp. 95–101, 2017.

[13] C. Yin, L. Xia, S. Zhang, R. Sun, and J. Wang, "Improved clustering algorithm based on high-speed network data stream," *Soft Comput.*, vol. 22, no. 13, pp. 4185–4195, Jul. 2018.

[14] C. Wu, L. Liu, M. Wei, B. Xi, and M. Yu, "Statistics-based optimization of the polarimetric radar hydrometeor classification algorithm and its application for a squall line in south China," *Adv. Atmos. Sci.*, vol. 35, no. 3, pp. 296–316, Mar. 2018.

[15] K. Sheng, Z. Liu, and D.-C. Zhou, "A multi-class semi-supervised classification algorithm based on evidence theory," *Tien Tzu Hsueh Pao/Acta Electronica Sinica.*, vol. 46, no. 11, pp. 2642–2649, 2018.

[16] Q. Zhang, X. Sun, K. Feng, S. Wang, Y.-H. Zhang, S. Wang, L. Lu, and Y.-D. Cai, "Predicting citrullination sites in protein sequences using mRMR method and random forest algorithm," *Combinat. Chem. High Throughput Screening*, vol. 20, no. 2, pp. 164–173, Jun. 2017.

[17] J. Chen, K. Li, Z. Tang, K. Bilal, S. Yu, C. Weng, and K. Li, "A parallel random forest algorithm for big data in a spark cloud computing environment," *IEEE Trans. Parallel Distrib. Syst.*, vol. 28, no. 4, pp. 919–933, Apr. 2017.

[18] H. L. Liu, Y. L. Xie, and Y. W. Jia, "Ecological security assessment and spatial-temporal evolution of Shanxi province," *Econ. Geography*, vol. 38, no. 5, pp. 161–169, 2018.

[19] M. Wolsink, "Distributed energy systems as common goods: Sociolitical acceptance of renewables in intelligent microgrids," *Renew. Sustain. Energy Rev.*, vol. 127, Jul. 2020, Art. no. 109841.

[20] C. Wang, S. Li, T. Cheng, and B. Li, "A construction of smart city evaluation system based on cloud computing platform," *Evol. Intell.*, vol. 13, no. 1, pp. 119–129, Jun. 2020.

[21] J. Song, L. Yang, Q. Gao, and J. Liu, "Strong tolerance random forest algorithm in seismic reservoir prediction," *Oil Geophys. Prospecting*, vol. 53, no. 5, pp. 954–960, 2018.

[22] Y. Zhan, Y. Luo, X. Deng, M. L. Grieneisen, M. Zhang, and B. Di, "Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment," *Environ. Pollut.*, vol. 233, pp. 464–473, Feb. 2018.

[23] J. Peng, J. L. Jia, and Y. N. Hu, "Construction of ecological security pattern in the agro-pastoral ecotone based on surface humid index: A case study of Hangjin banner, inner mongolia autonomous region, China," *J. Appl. Ecol.*, vol. 29, no. 6, pp. 1990–1998, 2018.

[24] I. M. Jacobson, J. K. Lim, and M. W. Fried, "American gastroenterological association institute clinical practice update—Expert review: Care of patients who have achieved a sustained virologic response after antiviral therapy for chronic hepatitis C infection," *Gastroenterology*, vol. 152, pp. 1578–1587, May 2017.

[25] S. Cocco, R. Monasson, L. Posani, and G. Tavoni, "Functional networks from inverse modeling of neural population activity," *Current Opinion Syst. Biol.*, vol. 3, pp. 103–110, Jun. 2017.

[26] S. Yingying, H. Lianjuan, W. Jianan, and W. Huimin, "Quantum-behaved RS-PSO-LSSVM method for quality prediction in parts production processes," *Concurrency Comput., Pract. Exper.*, vol. 9, Sep. 2019, Art. no. e5522.

[27] C.-W. Shen, M. Chen, and C.-C. Wang, "Analyzing the trend of O2O commerce by bilingual text mining on social media," *Comput. Hum. Behav.*, vol. 101, pp. 474–483, Dec. 2019, doi: 10.1016/j.chb.2018.09.031.

[28] C.-W. Shen, J.-T. Ho, P. T. M. Ly, and T.-C. Kuo, "Behavioural intentions of using virtual reality in learning: Perspectives of acceptance of information technology and learning style," *Virtual Reality*, vol. 23, no. 3, pp. 313–324, Sep. 2019, doi: 10.1007/s10055-018-0348-1.

ZHAOYA CHEN was born in Shunping, Hebei, China, in 1990. He takes a successive postgraduate and doctoral program. He is currently pursuing the Ph.D. degree with the College of Resources and Environmental Sciences, Hebei Agricultural University.

He has published many academic articles in the field of research in domestic and foreign journals. He has participated in relevant academic conferences and academic salons for many times. He is the author of one book and more than 10 articles. His research interests include land ecology, land evaluation, land planning, sustainable use of soil, and land resources.

LIJUN MA was born in Chengde, Hebei, China, in 1980. He received the master's degree in land resource management from Hebei Agricultural University, in 2006.

Since 2015, he has been the Deputy Director of the Department of Land Engineering Technology, Hebei Agricultural University. Since 2016, he has been an Assistant Professor with the School of Land and Resources, Hebei Agricultural University. He is the author of one book and more than 10 articles. His research interests include land safety evaluation, rural revitalization, and land use.

He is a Senior Member of the Chinese Society of Agricultural Engineering.

FENGYU GUO received the B.S. degree in land resource management from Hebei Agricultural University, in 2003, and the master's degree and the Ph.D. degree in management science and engineering management from China Agricultural University, in 2006 and 2009, respectively.

Since 2015, she has been an Associate Professor and a Master's Tutor with Hebei Agricultural University, majoring in land ecology and land economics. She has successfully undertaken more than ten scientific research projects, including the Ministry of Education, the Hebei Social Science Association, the Hebei Social Science Fund, the Baoding Social Science Association, Hebei Agricultural University, and so on. She has conducted many researches on engineering management and land economics. She has published more than 20 articles in important academic journals. She has edited and participated in one scientific and technological work and three training materials.

YAHENG CHEN was born in Shunping, Hebei, China, in 1973. He received the B.S., M.S., and Ph.D. degrees in soil science from Hebei Agricultural University.

Since 2015, he has been the Director of the Department of Land Engineering and Technology, Hebei Agricultural University. He is the First Batch of Young Academic Leaders with Hebei Agricultural University. His main research interests include land consolidation and land use planning. He has received two provincial second prizes, four third prizes, two software copyright registrations, three works, four textbooks, more than 80 academic articles, and five projects at or above the provincial level.

He is a Senior Member of the Chinese Society of Agricultural Engineering.

● ● ●