

## Research Article

# Prediction of the RFID Identification Rate Based on the Neighborhood Rough Set and Random Forest for Robot Application Scenarios

Hong-Gang Wang,<sup>1</sup> Shan-Shan Wang<sup>1</sup>,,<sup>1</sup> Ruo-Yu Pan,<sup>1</sup> Sheng-Li Pang,<sup>1</sup> Xiao-Song Liu,<sup>2</sup> Zhi-Yong Luo,<sup>3</sup> and Sheng-Pei Zhou<sup>4</sup>

<sup>1</sup>School of Communication and Information Engineering & School of Artificial Intelligence, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi 710121, China

<sup>2</sup>Guangdong Zhongke Zhenheng Information Technology Co. Ltd., Foshan, Guangdong 528225, China

<sup>3</sup>School of Electronics and Communication Engineering, Sun Yat-Sen University, Guangzhou 510006, China

<sup>4</sup>Shenyang Institute of Automation (Guangzhou) Chinese Academy of Sciences, Guangzhou 511458, China

Correspondence should be addressed to Shan-Shan Wang; 15029070967@163.com

Received 3 September 2020; Revised 12 November 2020; Accepted 3 December 2020; Published 29 December 2020

Academic Editor: Rongxin Cui

Copyright © 2020 Hong-Gang Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of Internet of Things technology, RFID technology has been widely used in various fields. In order to optimize the RFID system hardware deployment strategy and improve the deployment efficiency, the prediction of the RFID system identification rate has become a new challenge. In this paper, a neighborhood rough set and random forest (NRS-RF) combination model is proposed to predict the identification rate of an RFID system. Firstly, the initial influencing factors of the RFID system identification rate are reduced using neighborhood rough set theory combined with the principle of heuristic attribute reduction of neighborhood weighted dependency, thus obtaining a kernel factor subset. Secondly, a random forest prediction model is established based on the kernel factor subset, and a confusion matrix is established using out-of-bag (OOB) data to evaluate the prediction results. The test is conducted under the constructed RFID experimental environment, whose results showed that the model can predict the identification rate of the RFID system in a fast and efficient way, and the classification accuracy can reach 90.5%. It can effectively guide the hardware deployment and communication parameter protocol setting of the system and improve the system performance. Compared with BP neural network (BPNN) and other prediction models, NRS-RF has shorter prediction time and faster calculation speed. Finally, the validity of the proposed model was verified by the RFID intelligent archives management platform.

## 1. Introduction

In recent years, ultra-high-frequency (UHF) passive RFID technology has been widely applied in applications of unmanned warehouse, industrial site, new retail store management, and other scenarios due to its excellent ability in long-distance and multitag reading [1]. With respect to the conventional quasistatic RFID system that is usually deployed in a fixed way in particular areas, its system architecture and parameter configuration are unadjustable to some extent, making it difficult to be applied to certain

practical situations. In response to the abovementioned circumstances, a novel mobile RFID system is, therefore, proposed. RFID robots are not only to simply assemble the robot with the RFID system but also to combine the RFID system with the mobile robot to form a unified system. The optimization and control of the RFID system need to fully consider the factors such as tag environment, space, moving speed, and other factors. Compared with the existing conventional quasistatic RFID system, the RFID system on the robot is a typical dynamic system. In the conventional quasistatic RFID system, the system deployment and reading

strategy are relatively fixed, while in mobile RFID robots, the RFID system needs to constantly adjust the parameters and control the robot moving to maximize the adaptation of the environment to obtain the best application performance. When the mobile RFID robot is in task areas, it can adaptively adjust protocol parameters and hardware deployment strategies to accomplish tag reading tasks reliably under dynamic scenarios, thereby improving identification efficiency of the system. In domains of project planning and designing, the RFID system identification rate is the key technical index measuring system quality. In order to improve efficiency of the system's architecture and engineering deployment, predicting the identification rate of the novel mobile RFID system is of critical priority.

Among the existing prediction models of the RFID system identification rate, Liu et al. successively proposed a logistic regression analysis model, learning vector quantization neural network, and other intelligent algorithms to predict the rate of RFID system identification, which achieved good prediction effect. Despite the good effect achieved through adopting the abovementioned proposals, certain shortcomings still occur in practical applications involving huge amount of computation and overfitting [2, 3]. By introducing neighborhood rough set theory, Qiao et al. conducted optimization works with respect to influencing factors of the RFID system identification rate to improve identification efficiency. However, those factors being selected from actual test scenario are rather subjective, which may impose potential disadvantageous impacts on the identification rate of the actual system [4].

All the aforementioned algorithms are used to predict the system identification rate in the conventional quasi-static RFID system, yet few research studies focus on the system in dynamic scenes so far. In response to various tag numbers, complex multipath channel interference, and other factors encountered in designing and deploying the RFID system in dynamic scenes currently, tremendous efforts have been made on validation tests and on reducing miss rate of tags to avoid adjusting parameter deployment to do a lot of testing. Through applying RFID technology to mobile robot while extending its corresponding application schemes to dynamic scenes, a novel mobile RFID system is, therefore, established, in which a new intelligent learning algorithm being referred to as the neighborhood rough set and random forest (NRS-RF) combination model is introduced to predict the system's identification rate.

From the perspective of RFID system hardware deployment, we comprehensively select initial influencing factors to explore the relationship between the system identification rate and each influencing factor. This approach avoids doing a lot of verification tests in order to obtain the optimal deployment strategy. The NRS-RF model predicts the system identification rate in a fast and efficient way, so as to reverse guide the hardware deployment and communication protocol parameters of the RFID system, improve the performance of the RFID system, and meet the engineering needs. This novel mobile RFID system breaks the conventional quasistatic RFID system design and

provides more in-depth scene perception and real-time read-write strategy optimization for practical engineering needs.

The specific steps of the NRS-RF model are as follows:

Firstly, multiple initial influencing factors that affect the identification rate of RFID system are identified comprehensively. In specific, NRS theory is adopted to reduce influencing factors and data redundancy in between these factors, by which the kernel factor subset is selected. Secondly, the bootstrap method is used to resample the training set to support training the random forest prediction model [5–7]. The NRS-RF model is compared with other prediction models such as the backpropagation neural network (BPNN) to verify its advantages in predicting the identification rate of the RFID system. The test results show that the NRS-RF model can accurately and quickly complete the prediction of the RFID system identification rate, and the classification accuracy can reach 90.5%. It effectively guides the project deployment and improves the performance of the RFID system.

Compared with other models such as the BPNN, the NRS-RF model has obvious advantages in terms of classification accuracy and training time. Last but not least, the model is applied to the project of the RFID intelligent archives management platform, thus validating and verifying effectiveness of the proposed model.

The remainder of this paper is organized as follows: Section 2 presents an overall review of related work. Section 3 highlights relevant theoretical methods. Section 4 outlines experimental testing and analysis. Section 5 analyzes simulation results in detail and engineering application. Section 6 summarizes conclusions.

## 2. Related Work

Generally speaking, the deployment environment of the conventional RFID system is usually located in fixed scenes such as the entrance and exit of a corridor or passageway, for which reason it is inconvenient to apply tag identification in these areas. Under the circumstance of the scene with large identification area, a multireader mechanism is generally adopted [8]. However, the expenditure for improving the mechanism will be huge and unaffordable, not to mention potential collision between readers. Currently, as diverse algorithms emerge, protocol algorithms of reader anticollision has attracted much attention, among which heartbeat algorithm [9] and color wave algorithm, as well as the improved version of corresponding algorithms, prosper [10, 11]. In this paper, as our study mainly concentrates on single mobile reader mechanism, the RFID system with multiple readers, therefore, does not need to be particularly introduced in detail.

The state-of-the-art mobile RFID system is suitable for tag identification in small and medium areas. This mobile RFID system that works at UHF does not require any power supply (passive), featuring with characteristics of long identification distance, small size, strong directionality, and outstanding robustness against environmental changes [12]. The system ensures that not only all tags are covered within

the identification range of readers' signal and can be read successfully but also corresponding information processing belongs to the extended applications of the conventional fixed RFID system. However, the aforementioned dynamic RFID system cannot achieve its optimal status due to the vulnerable identification rate being seriously affected by external interference that may be induced by subjectivity, experience, and real-time hardware deployment of the system. Therefore, the dynamic RFID application scene needs to guarantee more in-depth scene perception and real-time read-write strategy optimization, which poses new challenges for RFID read-write technology.

Under new technological strategies and formats such as artificial intelligence, big data, new-generation robots, intelligent manufacturing, and new retail, RFID technology is facilitating robots to complete automation and dehumanization of warehouse management, industrial site, and new retail management [13]. Minho et al. introduced relations between the RFID system identification rate and influencing factors in a mobile system, established a support vector machine (SVM) model, and predicted the RFID system identification rate. However, certain errors still existed between the prediction results and the actual identification rate. In the abovementioned methodology, factors that affect the system identification rate were selected inadequately, and the actual scene was neither validated nor verified [14].

In January 2016, Keonn technology company of United States presented an RFID robot being referred to as *advanrobot* and applied it to a clothing retail scene to achieve fast and accurate mobile reading [15]. In April 2016, a well-known United States manufacturer named *Thingmagic* introduced adaptive duty cycle technology in the reader to minimize the reader's working time with respect to tag numbers, thus reducing power consumption [16]. Wang et al. proposed an efficient energy detection and calculation method for the RFID system in a dynamic scene, which is different from conventional anticollision algorithm. The tag helps the reader to judge whether the tags in the identification area collide with each other or not by sending a PBD burst time. If collision occurs, the collision problem will be solved by recursive polling, thereby improving the tag identification rate of the system [17]. In August 2016, *Impinj* proposed a scheme of the *Speedway Revolution* RFID reader, to which automatic performance setting is introduced based on environmental noise detection and on automatic dynamic antenna switching technology to optimize read-write time and efficiency [18].

In this paper, aiming at promoting sustainable development of RFID technology, research studies on the RFID robot and on developing new generation of adaptive read-write technology are, therefore, conducted, satisfying specific demand of niche market while accelerating technological progress of the industry. In order to improve the system identification rate from a physical perspective of hardware deployment, an intelligent learning algorithm named RFID system identification rate prediction is proposed based on the NRS-RF model. By using our proposed method, not only is the relation mined between diverse influencing factors and the system identification rate but

also intelligent scene perception is realized through model matching instead of using conventional methods, thereby improving the prediction accuracy. Moreover, by combining the novel mobile RFID system, the optimal combination of hardware deployment configuration scheme is obtained to improve the RFID system's identification rate, thus maximizing effectiveness and efficiency of the hardware deployment while ensuring cost efficiency in terms of labor force and resources.

The philosophy of using the NRS-RF model can be summarized as follows:

- (1) Influencing factors of the RFID system identification rate are selected as the sample data, in which the NRS theory is used to reduce the attribute of these factors, to select the kernel factor subset that affects the identification rate, and to reduce the input dimension of nonlinear mapping.
- (2) Based on the kernel factor set, the prediction model is constructed featuring with the 2-classification random forests RFID system identification rate, upon which a novel mobile RFID experimental test platform is established accordingly. Due comparison analysis is performed between the NRS-RF model and the BPNN and other prediction models in terms of OA, Kappa coefficient, RMSE, MAE, training time and prediction time, and correlation. The test results show the superiority of the NRS-RF model.
- (3) The prediction model is applied to the intelligent archives management platform, and the importance distribution of influencing factors to RFID system identification rate classification prediction is analyzed, verifying effectiveness and efficiency of the proposed model.

### 3. Methods

**3.1. Reduction Feature Factors.** Emerging as an innovation in classical rough set theory, neighborhood rough set (NRS) theory was put forward by Lin in 1988 [19, 20]. The idea of NRS algorithm is that, in the real space, each data point will form a neighborhood  $\delta_B(x_i)$  and the data in the neighborhood family will constitute the basic information particles [21–23]. NRS solves the problem of numerical data set that is not easy to be processed in classical rough set theory, removes redundant data features, and selects the key factors that affect the identification rate of the RFID system [24, 25].

In the RFID system, the information system  $W$  is composed of quad-tuple  $W = (U, Y, V, f)$ , where  $U$  is the sample number set of the identification rate,  $Y$  is the sample set of the identification rate ( $Y = C \cup D$ ),  $C$  is the influencing factors of the identification rate functioning as the attribute set, and  $D$  is the classification level of the identification rate functioning as the decision attribute. This quad-tuple information system  $W$  is called the decision table, within which  $V$  denotes the value field of attribute and  $f$  represents the mapping relation used to specify the property value of sample  $x$ , that is,  $f = U \times Y \rightarrow V$ .

If the sample  $x_i \in U$ , the neighborhood condition of  $x_i$  needs to satisfy  $\delta_B(x_i) = \{x_j | x_j \in U, \Delta_B(x_i, x_j) \leq \delta\}$ , where  $\Delta$  denotes distance function, for any  $x_1, x_2, x_3 \in U$ , and  $\Delta$  satisfies

$$\begin{cases} \Delta(x_1, x_2) = \Delta(x_2, x_1), \\ \Delta(x_1, x_2) \geq 0, \Delta(x_1, x_2) = 0; \quad \text{at this time } x_1 = x_2, \\ \Delta(x_1, x_3) \leq \Delta(x_1, x_2) + \Delta(x_2, x_3). \end{cases} \quad (1)$$

For any attribute set  $L$ , when it is classified, the indiscernible data will be grouped into one class. They belong to the indiscernible relation, which can be given by the following equation:

$$\text{IND}(L) = \{(x, y) \in U \times U: f(x, a) = f(y, a), a \in P\}. \quad (2)$$

For the indistinguishable relationship group  $H \subseteq Y$ ,  $a \in H$ , if the relationship  $\text{IND}(L) = \text{IND}(H - \{h\})$  is satisfied, it is considered that  $\{h\}$  is the redundant data on  $H$ , which can be reduced. It can be defined that the value of theory field is  $V$ ,  $R$  is the equivalent relation on  $V$ , to which the upper approximation, the lower approximation of neighborhood rough set, and the boundary field of subset  $x$  satisfy the following equations, respectively:

$$\bar{R}(X) = \cup \{Y \in U | R: Y \cap X \neq \emptyset\}, \quad (3)$$

$$R(X) = \cup \{Y \in U | R: Y \subseteq X\}, \quad (4)$$

$$BN_R(X) = \bar{R}(X) - R(X), \quad (5)$$

where  $\text{Pos}_R(X) = R(X)$  is the positive domain of subset  $X$  and  $\text{Neg}_R(X) = U - \bar{R}(X)$  is the negative domain. For any  $c \in C$ , the dependence degree of decision attribute  $D$  on condition attribute  $c$  is defined as

$$\gamma_a = \frac{\text{Pos}_a(D)}{U}. \quad (6)$$

If two random variables are defined, the correlation degree calculated using mutual information measure can be satisfied by the following equation:

$$I(X, Y) = - \sum_{x \in V_x} \sum_{y \in V_y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}. \quad (7)$$

**3.2. Random Forest Prediction.** Random forest (RF) algorithm is based on the decision tree as a learning machine to build bagging integration [26, 27], thus further introducing the selection of random attributes. Specifically, conventional decision tree algorithm selects an optimal attribute in the current attribute set when selecting the partition attribute, whereas the RF algorithm randomly selects a subset containing  $K$ -th attributes in the attribute set, through which an optimal attribute is selected from the subset selection for classification. Using this kind of random selection, the random forest can avoid disadvantages of overfitting,

exhibiting excellent antinoise performance. Outperforming other intelligent algorithms, it only requires simple computation while maintaining cost efficiency. The principle of RF algorithm is given below, as shown in Figure 1.

The random forest algorithm adopts an integrated algorithm; the classification accuracy of the algorithm itself is much higher than other single algorithms, so the accuracy is higher. The random forest algorithm can handle high-dimensional data without any feature selection. When bootstrap sampling is performed on training samples, out-of-bag data will be generated. Unbiased estimates of true errors can be obtained in the process of model generation without loss of training data. Due to the simple implementation, high accuracy, and strong antioverfitting ability of the algorithm, when faced with nonlinear data such as the identification rate of the RFID system, the model shows high classification accuracy and is also suitable as a benchmark model.

Due to the change of tag numbers and the complex multipath channel interference in the architecture and deployment of the RFID system, the prediction value of the RFID system identification rate is, therefore, discretized. In order to avoid adjusting the parameters to obtain optimal deployment strategy and to do a lot of testing and verification, it is necessary to comprehensively select the influencing factors, thus mining the nonlinear relationship between the RFID system identification rate and its influencing factors. From the perspective of hardware deployment, the RFID hardware deployment and communication protocol parameters should be optimized and adjusted with the purpose to reduce the missed rate of tags and to improve the system performance. The relation between the influencing factors and the identification rate is obtained using RF classification prediction algorithm, to which its mathematical model can be expressed by the following equations:

$$\text{RFP} = \{Tn, M\}, \quad (8)$$

$$d = f(x_1, x_2, \dots, x_n, \text{RFP}), \quad (i = 1, 2, \dots, n). \quad (9)$$

In equation (8), RFP is the parameter set of the random forest prediction model,  $Tn$  is the number of regression trees in the model, and  $M$  is the number of influencing factors. In equation (9), the prediction method of the RFID system identification rate is defined, where  $f$  is the uncertainty function relation of random forest classification algorithm,  $d$  is the system identification rate, and  $x_i$  is the  $i$ -th index factor affecting the identification rate, including number of tags, number of antennas, reading distance, and other parameters.  $n$  is the number of influencing factors of the system identification rate. Another advantage of the random forest algorithm is that the influencing factors participating in the algorithm can measure the importance degree of the classification. The contribution value of the influencing factors can be determined by calculating the information gain rate of the dataset. The information gain rate is positively correlated with certainty of the influencing factor, indicating that the higher the information gain rate, the stronger the certainty of the influencing factor. Calculation of the information gain is given by the following equations:



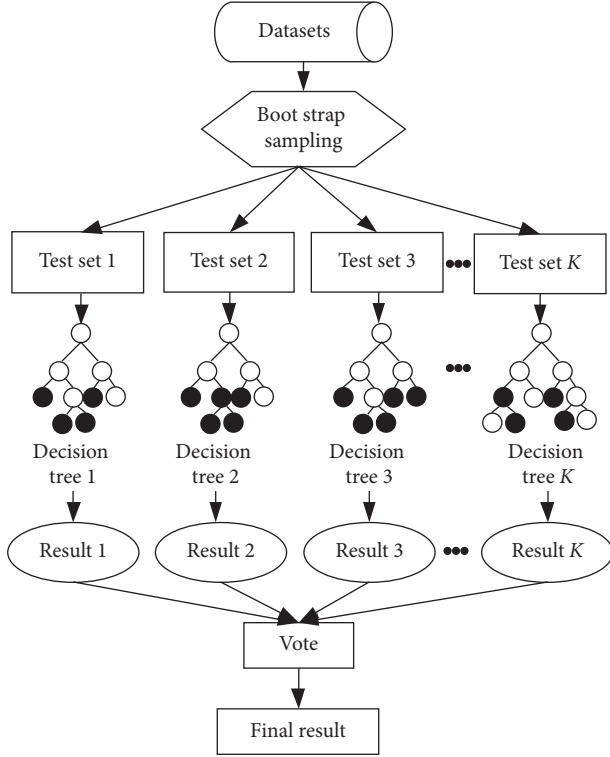


FIGURE 1: Schematic diagram of the random forest algorithm.

$$g(D, X_i) = H(D) - H(D|X_i), \quad (10)$$

where  $H(D)$  is the information entropy of dataset  $D$ , namely,

$$H(D) = - \sum_{i=1}^n D_i \log_2 D_i, \quad (11)$$

where  $H(D|X_i)$  is the entropy change brought by  $X_i$  variables under the condition of dataset  $D$ .

$$H(D|X_i) = H(D, X_i) - H(X_i). \quad (12)$$

## 4. Experiments

**4.1. Selection of Influencing Factors.** We use the UHF passive RFID system as the experimental background to build a novel mobile UHF passive RFID test platform. This type of RFID system has many advantages that users cannot refuse. This technology can achieve no human intervention and without-contact identification. This technology can realize nonhuman intervention and noncontact identification. The system can be used in many harsh environments, such as automatic container terminal yard operation system identification, high-speed moving object identification, multitag identification, and other scenarios at low cost.

For the novel mobile RFID system, because of its mobility, the tags in the identification area have the opportunity to be identified and processed, which can linearly expand the identification range of the reader. Under the same test conditions, the novel mobile RFID system can identify more tags and obtain the system identification rate faster than the

conventional quasistatic RFID system. However, when the system is put into actual engineering deployment, it will encounter factors such as changes in the number of tags and complex multipath channel interference. When we adjust the hardware deployment of the system, there will be subjectivity and timeliness, resulting in the identification range of the reader. In the blind area, we can reduce the error of the RFID system identification rate by optimizing the hardware configuration reasonably, so as to maximize the performance of the system.

In RFID system project planning and designing, the system identification rate that represents the number of tags being successfully read in the inventory process accounts for the total number of tags. It is an important indicator in measuring the system performance. Generally, the system identification rate correlates with controllable factors involving moving speed of the reader, number of tags, and other factors and with uncontrollable factors including multipath channel interference, Doppler effect, and other factors, as shown in Figure 2. During the experiment, eleven controllable factors are selected as conditions, including height of the antenna  $P1$ , number of tags  $P2$ , horizontal distance between the tag and shelf  $P3$ , number of reader polling cycles  $P4$ , moving speed of the antenna  $P5$ , horizontal angle of the antenna  $P6$ , vertical angle of the antenna  $P7$ , number of antennas  $P8$ , light intensity (the sunlight and the dark room without light are selected under same experimental conditions)  $P9$ , shelf height  $P10$ , and reader transmitting power  $P11$ . The level of the RFID system identification rate is regarded as the decision attribute.

In order to verify prediction accuracy of the proposed intelligent prediction method for the RFID system, firstly, orthogonal experiments are conducted using a variable-controlling approach. The antenna height is set to 0.6 m, 0.9 m, and 1.2 m, respectively; the number of tags is set to 90, 130, 150, and 180, respectively; the distance between tags and antennas is set to 0.8 m, 1.5 m, and 1.9 m, respectively; the number of polling turns of the reader is set to 1, 2, and 3, respectively; the moving speed of the antenna is set to 0.3 m/s, 0.6 m/s, and 0.9 m/s, respectively; the horizontal angle of antenna is set to  $0^\circ$  and  $30^\circ$ , respectively; the vertical angle of antenna is set to  $0^\circ$  and  $30^\circ$ , respectively; the number of antennas of the same specification is set to 1 and 2, respectively; the light intensity is set to 0 (sunlight) and 1 (darkroom), respectively; the shelf height is set to 0.6 m and 1.2 m, respectively; and the reader transmitting power is set to 18 dBm, 23 dBm, and 28 dBm, respectively.

We need to traverse 10368 sets of cross experiments and record the identification rate of the RFID system, in which each group of influencing factors is used to conduct 3 experiments, and both the average value of the RFID system identification rate and the influence factors can be taken as the sample data.

In practical engineering applications, the new type of portable RFID system should be expected to reach 100% identification rate on tagged goods inventory. In practical applications, however, due to complex constraining factors such as multipath effect, it is not ideal to rely on the current hardware. Therefore, the threshold of the identification rate

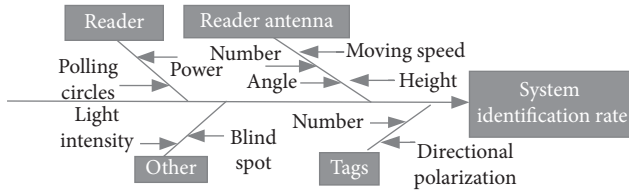


FIGURE 2: Influencing factors and the system identification rate.

can only reach 95%, which needs to be adjusted with respect to differing projects according to actual situations. We believe that whether the identification rate is qualified or not lies in satisfying the requirements of actual situations. If the identification rate of the system exceeds 95%, it is regarded as high rate and can be categorized into type 1; otherwise, it is labelled as unqualified type 2 if the identification rate is less than the threshold value of 95%.

**4.2. Comprehensive Analysis of the Test Platform.** According to the proposed prediction model of the RFID system identification rate, a novel mobile UHF passive RFID system is constructed combining a mobile robotic car reader antenna and other devices as an RFID experimental test platform. Featuring 9 m length, 4 m width, and 3.75 m height, the platform is built in the corridor of an open classroom, as shown in Figure 3(a). The product of water drop robot of *Beijing Yunji technology company* is selected as the robotic vehicle used in the experiment, which maneuvers with abilities of highly sensitive intelligent perception and positioning navigation.

As shown in Figure 3(b), the robotic vehicle that is enabled with a path scanning function can adjust its moving speed following the experimental requirements. The vehicle scans the test site, where the white color represents the area it walks within, the gray color represents the unexplored area, and the black solid line represents the obstacle information established in the map. The scanning results are shown in Figure 4. The mobile robotic vehicle is equipped with an adjustable tripod, reader, and antenna. The detailed experimental specifications are introduced as follows.

An RFID reader (MODEL Mercury6, *ThingMagic*) is adopted, which has stable read-write performance supporting reading ISO 18000-6C protocol standard in wired mode with a 9 dBi circular polarization reader antenna. The four-layer book shelf on the left side is 1.5 m \* 2 m, with a certain number of books evenly distributed on each layer and with archive tags pasted on the side of each book. The tag used is UHF passive and its working frequency mainly ranges from 860 MHz to 920 MHz, exhibiting excellent directionality and satisfactory read distance.

The novel mobile RFID experimental system has the following characteristics:

- (1) The novel mobile RFID system overcomes the traditional quasistatic RFID system because the identification area is static, there is always the problem of blind spots in identification, and it avoids the disadvantages of manual handheld readers that cause

large errors in the identification rate of the RFID system

- (2) In order to identify all tags in the area, the novel mobile RFID system requires a mobile robot to poll the tags in the reading area, and the tags are covered by the reader signal area, at least, once
- (3) Some tags will repeatedly enter and leave the signal area of the reader
- (4) Some tags will be identified multiple times by the reader

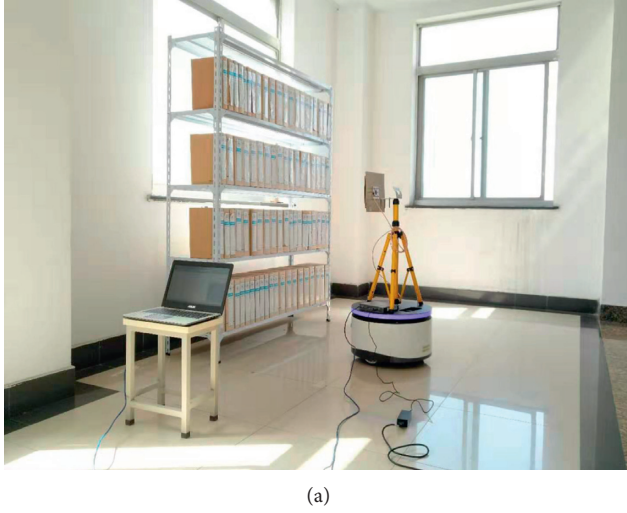
The experiment builds a novel mobile UHF passive RFID test platform. Because only a movable single reader mechanism can be used to realize multitag identification, the cost is much less than the RFID system with multireader mechanism, which also saves multiple readings. It also saves equipment installation and wiring costs. However, the system still has the following limitations.

Due to the sector-shaped antenna radiation field of RFID mobile robotic vehicle, the echo signal of tags on both sides of the sector is weak and vulnerable to interference. With the range of sector coverage, the echo signal of far end tag A is weak, whereas that of near end tag B is strong. During the process of identifying tags, the tags on both ends of the sector may block the reader antenna if the vehicle moves too fast, thus resulting in information loss.

In order to facilitate simple and direct identification, different moving speeds of the vehicle are selected as initial influencing factors for the RFID system. As shown in Figure 5, it is required that the vehicle must move along the linear path with the signal radiation radius  $R$  at a reasonable and constant speed under specified conditions to obtain the optimal identification rate of the RFID system. Our future research will be focusing on the optimal disk point path to obtain the maximum identification rate of the system.

Throughout the process of identifying tags, only by adjusting the parameters from the physical perspective of hardware deployment can the optimal deployment scheme of combination be obtained. In the abovementioned experiment, however, the ideal identification rate of 100% cannot be achieved merely through adjusting hardware configurations. Admittedly, there are still other detrimental encumbrances that may potentially induce blind area problems, to which limitation involving antenna polarization mismatching and multipath fading can be attributed largely.

In response to the situation that the number of tags on the shelf is positively correlated with the size of storage area, only small-size tags are selected to be uniformly placed in the identification area without considering the serious tag collision problem in the scene with dense multiple UHF tags. With respect to other initial factors selected in Section 4.1, such as antenna height, polling circle of the reader, and horizontal distance between the antenna and shelf, a reasonable test range should be conscientiously selected. Furthermore, there are still many important tasks to be accomplished, which include, but not limited to, conducting orthogonal combination of controllable variables and testing



(a)



(b)

FIGURE 3: Test environment, (a) experimental test platform, and (b) water droplet robot.



FIGURE 4: Two-dimensional scanning image of the test environment.

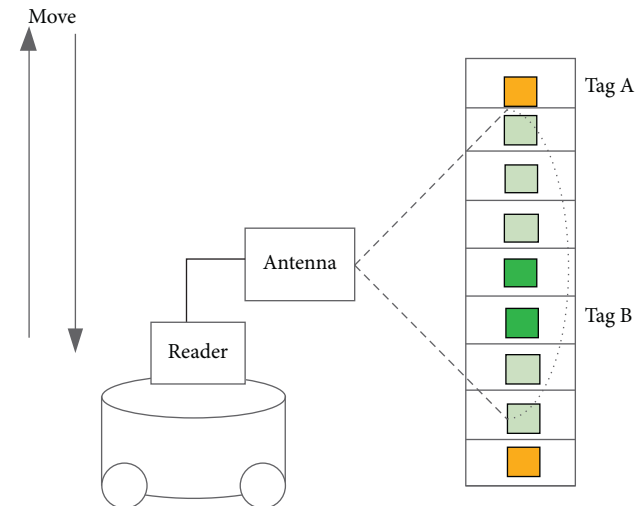


FIGURE 5: Schematic diagram of the novel mobile RFID system with the robotic vehicle.

diverse combinations of influencing factors, as well as recording the RFID system identification rate.

#### 4.3. Proposed Algorithm

**4.3.1. Reducing Influencing Factors.** In Section 4.1, 11 controllable factors (see Section 4.1) are selected as condition attributes, and the level of the RFID system identification rate is taken as a decision attribute.

In order to reduce the redundancy in between the aforementioned 11 groups of influencing factors and to improve the prediction accuracy of system identification, heuristic reduction algorithm of neighborhood weighted dependency is, therefore, adopted. The purpose of using this algorithm aims on reducing the attribute reduction of influencing factors and on obtaining the kernel factor subset  $R$  [28, 29]. The algorithm is a forward greedy attribute reduction algorithm based on the attribute importance of weighted dependence (Algorithm 1). Detailed explanation of the algorithm is described as follows [30, 31].

In a neighbourhood decision system  $W = (U, C \cup D, V, f, \varepsilon)$ , for  $Z \subset C$ , we define attribute  $x \in Z - B$  and define  $w$  as adjustment parameter and  $w\gamma_B^\delta$  as weighted dependency, so the importance degree of  $Z$  and  $D$  based on the weighted dependence degree satisfies the following equation:

$$\text{sig}(x, Z, D) = w\gamma_{B \cup x}^\varepsilon(D) - w\gamma_B^\varepsilon(D). \quad (13)$$

The neighborhood radius  $\varepsilon = 0.28$  was selected, the attribute subset was selected from the empty set, and the reduction subset was selected in turn to build an ordered reduction attribute subset. The dependence of the 11 groups of influencing factors increased as the important attributes of the reduction subset increase, thereby finally obtaining the decision table M1, as shown in Table 1. From Table 1, it is obvious that there are 500 sets of data after reducing sample set  $U$  which is composed of  $P1, P2, P3, P4, P5, P8$ , and  $P11$ . Compared with the initial influencing factors, after performing heuristic reduction algorithm of neighborhood

**Input:** Neighborhood decision system  $W=(U, C \cup D, V, f, \varepsilon)$ ,  $\varepsilon$  is neighborhood threshold,  $T$  is temporary subset.

**Output:** Reduction subset  $R$ .

**Steps:**

- (1)  $\phi \longrightarrow R; T \longrightarrow C - R;$
- (2)  $\forall x_i \in T$ , Calculate the attribute importance sig of the sample  $\text{sig}(x_i);$
- (3)  $x_k = \arg(\max(\text{sig}(x_i)))$ , Find the attribute with the most important attribute of attribute reduction attribute subset  $x_i;$
- (4) **If** ( $\text{sig}(x_i) > 0$ )
- (5)      $R \longleftarrow R \cup \{x_k\}, T \longleftarrow T - \{x_k\};$
- (6) **Else**
- (7)     return  $R;$
- (8) **End for**

ALGORITHM 1: Forward greedy attribute reduction.

rough set weighted dependency, some redundant factors are removed, and the input features are changed from the original 11 sets to 7 sets.

**4.3.2. Random Forest Prediction.** The prediction model in this paper is based on the random forest toolbox developed by the University of Colorado [32], and the corresponding codes were written in the operating environment of MATLAB7.1. The RFID identification rate prediction model based on the NRS-RF includes 6 steps. Specific prediction steps of the NRS-RF are demonstrated as follows:

- (1) *Normalized Input Data.* The kernel factor subset {antenna height  $P1$ , number of tags  $P2$ , distance between tags and shelf  $P3$ , number of polling cycles  $P4$ , moving speed of the antenna  $P5$ , number of antennas  $P8$ , and transmission power of the reader  $P11$ } obtained by attribute reduction of neighborhood rough set theory is used to construct  $500 * 7$ -dimensional sample data as input variables of the random forest model. The sample data are processed by the following equation:

$$x'_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}. \quad (14)$$

The predicted value of the identification rate of the system is transformed by

$$x = x'(x_{\max} - x_{\min}) + x_{\min}, \quad (15)$$

where  $x$  is the initial RFID system identification rate value and  $x_{\max}$  and  $x_{\min}$  are the maximum and minimum value of the system identification rate, respectively.

- (2) *Bootstrap Sampling Training Subsets and Decision Tree.* The bootstrap method was used to perform  $n$  times of resampling from the sample set  $S$  and to randomly generate  $n$  training subsets  $S_i$  with the same number of samples. During bootstrap sampling of training samples, 1/3 out-of-bag (OOB) data will be left behind. The OOB precision estimation of each decision tree can be obtained through out-of-pocket

TABLE 1: Decision table M1.

| $W$   | 1 | 2 | ... | 251 | 252 | ... | 351 | ... | 500 |
|-------|---|---|-----|-----|-----|-----|-----|-----|-----|
| $P1$  | 1 | 3 | ... | 2   | 3   | ... | 2   | ... | 6   |
| $P2$  | 3 | 7 | ... | 5   | 2   | ... | 2   | ... | 4   |
| $P3$  | 5 | 6 | ... | 2   | 1   | ... | 7   | ... | 2   |
| $P4$  | 3 | 5 | ... | 4   | 7   | ... | 4   | ... | 2   |
| $P5$  | 5 | 2 | ... | 5   | 2   | ... | 1   | ... | 6   |
| $P8$  | 3 | 6 | ... | 5   | 3   | ... | 6   | ... | 2   |
| $P11$ | 2 | 3 | ... | 1   | 4   | ... | 5   | ... | 3   |
| $V$   | 1 | 1 | ... | 1   | 1   | ... | 1   | ... | 1   |

samples. The OOB precision estimation of left and right decision trees in the forest can be averaged to obtain the generalization precision estimation of random forest. For all the sample subsets  $S_i$ , the CART algorithm is performed to construct decision trees, thus combining these trees to form a random forest, which is expressed as  $\{C_1, C_2, \dots, C_i\}$ .

- (3) *Node Split Growth.* When the nodes of the decision tree are splitting and growing, the input parameters in the Mtry block prediction model are randomly taken as the split subset of the current node. The value of Mtry represents the disturbance degree of the model attributes because the value in the model is sensitive which directly affects the prediction accuracy of the model. The value can be given according to the empirical equations:

$$M_{\text{try}} = \left\lceil \frac{M}{3} \right\rceil, \quad (16)$$

$$M_{\text{try}} = \lceil \log_2 M \rceil, \quad (17)$$

where  $M$  is the number of input variables which is 6 in this study. Hence, according to equations (16) and (17), the  $M_{\text{try}}$  value is 2. When the nodes are divided in the subset, the Gini index in CART algorithm is taken as the minimum principle to select the optimal split influencing factor and optimal split value. During the splitting process, none of pruning operations is performed, and the  $M_{\text{try}}$  block degree remains constant. The Gini system is defined by the following equation:



$$\text{Gini}(t_i) = 1 - \sum_{K=1}^K P_k^2, \quad (18)$$

where  $t_i$  represents the current influencing factor,  $K$  represents the number of groups of the influencing factors  $t_i$ , and  $P_k$  represents the probability that the sample point belongs to the  $K$  class. After determining the optimal splitting influencing factor  $t_i$ , if a subset  $S_i$  is split into two subsets  $S_{i1}$  and  $S_{i2}$  with respect to  $t_i$ , the optimal splitting value “ $a$ ” can be calculated by the following equation:

$$\min \text{Gini}(t_i, a) = \frac{|S_{i1}|}{|S_i|} \text{Gini}(S_{i1}) + \frac{|S_{i2}|}{|S_i|} \text{Gini}(S_{i2}), \quad (19)$$

where  $|S_i|$ ,  $|S_{i1}|$ , and  $|S_{i2}|$ , are the samples of  $S_i$ ,  $S_{i1}$ , and  $S_{i2}$ , respectively.

- (4) *Prediction Sample Category.* After each decision tree is constructed from the bottom to the top, the integrity of the tree is preserved without performing pruning operations, and all the decision trees are tested with test set  $X$  to obtain the test sample of the prediction category of the RFID system identification rate, which can be expressed as  $\{C_1(X), C_2(X), \dots, C_n(X)\}$ .
- (5) *Final Prediction Classification.* After training, the sample data  $x$  are input from test set data  $X$  into the model to obtain the prediction classification results and to select the final classification results of test set by voting mechanism. The principle of voting mechanism can be expressed by the following equation:

$$f_{\text{RF}}(x) = \arg \max_{i=1,2,\dots,c} \{I(f_i^{\text{tree}}(x) = i)\}. \quad (20)$$

- (6) *Evaluation Model.* The confusion matrix is established by the validation set of OOB data, and the classification results are evaluated. Based on confusion matrix, four evaluation indexes are selected, including overall accuracy (OA), Kappa coefficient, root mean square error (RMSE), and mean absolute error (MAE). The final prediction results of the RFID system identification rate are compared with the threshold condition of the system identification rate in the actual project. Therefore, it can be judged whether the hardware deployment scheme of the RFID system can meet the application requirements. The OA, Kappa coefficient, RMSE, and MAE are expressed by the following equations, respectively:

$$\text{Overall accuracy} = \frac{\sum_{i=1}^n X_{ii}}{N}, \quad (21)$$

$$\text{Kappa} = \frac{N \sum_{i=1}^n X_{ii} - \sum_{i=1}^n (X_{i+} \times X_{+i})}{N^2 - \sum_{i=1}^n (X_{i+} \times X_{+i})}, \quad (22)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}, \quad (23)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{Y}_i - Y_i|. \quad (24)$$

**4.3.3. Time Complexity Analysis.** This paper proposes a neighborhood rough sets and random forest combination model of the identification rate of the RFID system prediction model. The essence is to reduce the dimension of  $X$  samples and  $Y$  initial influencing factors. Firstly, the initial influencing factors are reduced using neighborhood rough set theory combined with the principle of heuristic attribute reduction of neighborhood weighted dependence, thus obtaining a kernel factor subset. So, the dimension  $Y$  is reduced to  $V$ . At this time, the time complexity of calculating the kernel factor set of the neighborhood rough set is  $O(V^2 X \log X)$ . Then, the selected kernel factor subset is taken as the input of the random forest model to establish the RFID system identification rate prediction model. At this time, the time complexity of the model is  $O(KVs (\log s)^2)$ , where  $K$  represents the number of basic classifier CART and  $s$  represents the number of training sets in the random forest algorithm. It is obvious that the time complexity at this time is lower than that of the random forest directly dealing with the initial influencing factor. After all, the sample has been dimensionally reduced,  $V \leq Y$ .

## 5. Results and Discussion

**5.1. Optimizing N-Tree and Constructing the Decision Tree.** Before performing the random forest algorithm, it is necessary to optimize the super parameter N-Tree which is the number of decision trees. The 500 \* 7-dimensional sample data are input variables of the random forest model. By changing the N-Tree value, the OOB precision corresponding to different N-Tree values can be calculated. The number of decision trees can be estimated by OOB precision, as shown in Figure 6.

From Figure 6, it can be seen that the value of N-Tree increases as the progress of model classification proceeds. When the value of N-Tree is greater than 500, the accuracy accordingly increases, to which the increasing tendency is not obvious but declines instead. Therefore, taking the model's identification classification accuracy and the classification time as reference standards, the final value of N-Tree is 500. Once the number of decision trees is determined, each tree is divided by the Gini coefficient expression and optimal splitting value given in Section 4.3.2 from the root node until each tree accomplishes growing. Here, we select a decision tree to observe its splitting and growing process. The optimal splitting influencing factor and optimal splitting value of each node in the splitting process of the decision tree are shown in Table 2. The complete construction process of the decision tree is drawn from the root node to the bottom, as shown in Figure 7.

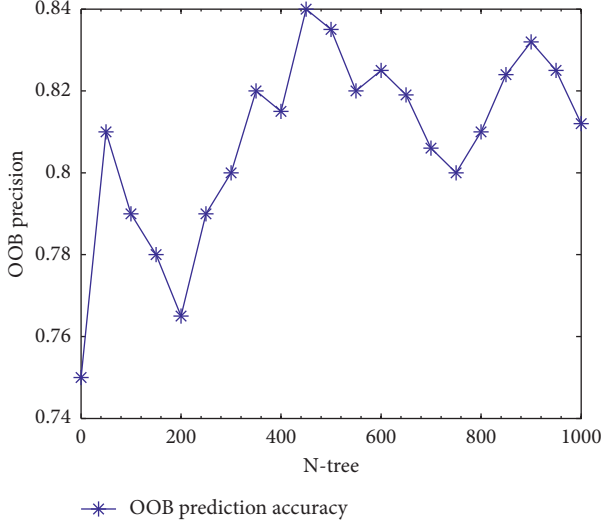


FIGURE 6: Relationship between N-tree and OOB.

TABLE 2: Optimal splitting influencing factor and the optimal splitting value of nodes.

| Optimal splitting influencing factor | Optimal splitting value |
|--------------------------------------|-------------------------|
| P1                                   | 0.364                   |
| P2                                   | 0.854                   |
| P3                                   | 0.648                   |
| P4                                   | 0.985                   |
| P5                                   | 0.751                   |
| P8                                   | 0.651                   |
| P11                                  | 0.528                   |

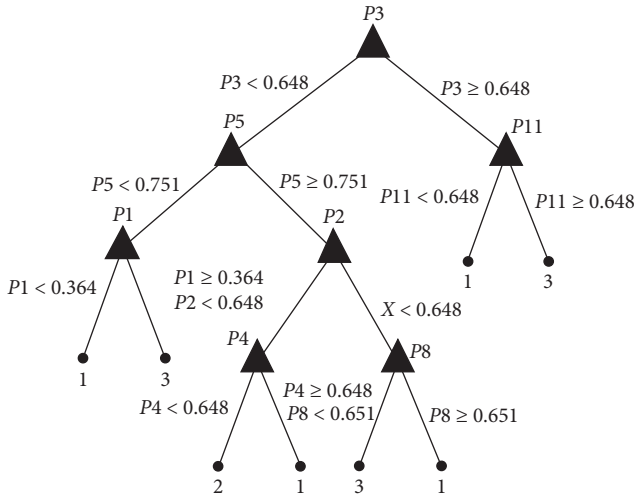


FIGURE 7: Decision tree structure.

**5.2. Test Results.** The test set is used to verify the classification accuracy of the constructed random forest model. The identification rate of the RFID system is obtained through simulation, as shown in Figure 8. According to the prediction results in Figure 8, the classification accuracy is

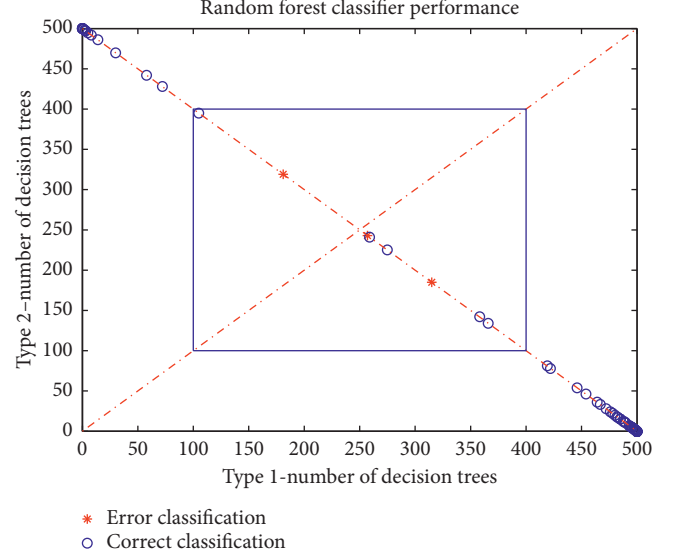


FIGURE 8: Predicting classification chart.

90.5%. The horizontal and vertical axis represent 500 groups of data, the red asterisk shape represents the error classification sample, and the blue circle denotes the correct classification sample. The data in the blue square represent the error-prone range, the closer the sample approaches 250 decision trees, the more difficult it is to make decisions, whereas the easier it is to make classification errors.

From the perspective of sample data, among the 500 sets of sample data, there are 357 sets of qualified samples that meet the threshold condition, which means the RFID system identification rate is higher than 95%. There are 143 sets of unqualified samples that do not meet the threshold condition; that said, the RFID system identification rate is less than 95%. Among 100 sets of data in the test set, there are 80 groups of qualified identification rate and the accurate prediction rate is about 96.25%, with an average misjudgment of 3 data groups. Among the unqualified identification rate, there are 20 groups with an accuracy rate of 90% and an average misjudgment of 2 data groups. At this time, the NRS-RF model exhibits excellent performance on predicting the RFID system identification rate.

In order to verify that the prediction accuracy can be improved through using a neighborhood rough set to reduce the initial influencing factor set, the relation between the prediction accuracy and the number of influencing factors is, therefore, analyzed adopting variable-controlling comparison experiment. First, accuracy verification was added to the test samples one by one according to the importance of influencing factors, as shown in Figure 9. It can be seen from Figure 9 that, under the condition of ensuring consistency of other parameters, the classification prediction accuracy was significantly improved as the number of influencing factors increases.

When the number of influencing factors reached 5, the overall prediction accuracy of the test sample increased slowly. When the number increased to 7, the accuracy reached 90.5% and then stabilized following with a small

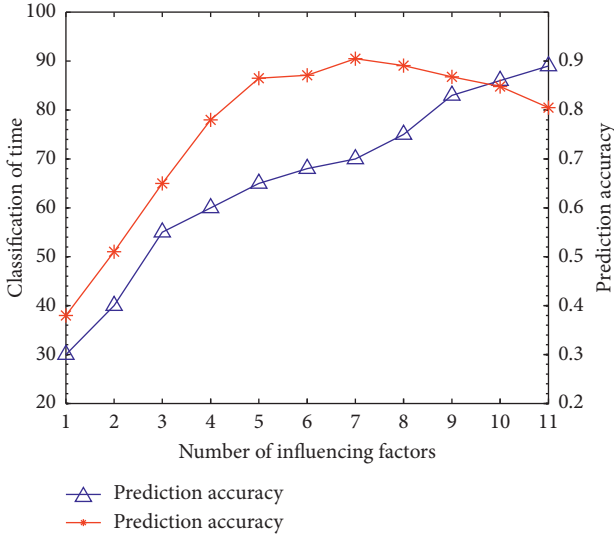


FIGURE 9: Influencing factors reduction verification.

decline. The classification time was on an upward trend as the number of influencing factors increased. When the influencing factors reached 11, the accuracy increased more significantly. Considering the combination of prediction accuracy and classification time, the 7 influencing factors obtained by attribute reduction are the optimal feature combination which effectively improve the classification accuracy while reducing the classification time.

In order to further verify the advantages of the NRS-RF model in predicting the RFID system identification rate, two new prediction algorithms are selected to construct the prediction network model and make a comparative analysis with the proposed NRS-RF model, including the  $K$ -nearest neighbor-naive Bayesian (KNN-NB) and backpropagation neural network (BPNN) [33–35]. These three models predict 300 groups of RFID system identification rate sample data, being compared in terms of OA, Kappa coefficient, RMSE, MAE, training time and prediction time, and correlation. The prediction results are shown in Table 3 and Figures 10–12.

The Kappa statistic is a measurement value for evaluating consistency, which indicates whether there is consistency between the predicted results of the model and the actual results. When the Kappa coefficient is greater than 0.75, it indicates that the model is better and has a certain value. It can be seen from Table 3 that the NRS-RF model with the highest classification accuracy has OA and Kappa coefficients of 88.5% and 0.875, which are much higher than those of other two models. It shows that the NRS-RF method can effectively screen out the nuclear factor set of system identification rate and improve the classification accuracy of the model.

Admittedly, featuring with excellent performance on applicability, the NRS-RF method can effectively eliminate the influencing factors affecting the RFID system identification rate. In Figure 10, the RMSE and MAE values of the NRS-RF model are relatively small, the prediction error is small, and the classification accuracy is higher. Comparing

the training set of the three models with the prediction set in terms of operating time, the NRS-RF model requires shorter time while ensuring lower computational complexity and higher calculation speed, thereby better satisfying engineering applications, as shown in Figure 11.

Different from KNN-NB and BPNN models, the random forest model is a kernel subset composed of 500 sample sets as the input of the model. However, due to its simple bifurcation structure of the base learner decision tree, its learning time is less than 6 s.

At the same time, when constructing the decision tree, it randomly selects part of the features as the classification basis of the tree growth, and when constructing the internal base learning device, it adopts the random sampling to put back the training samples, which ensures the generalization ability of the final model. Therefore, the random forest algorithm model has higher OA.

The correlation between the predicted value of the identification rate of the three models and the actual value is analyzed, as shown in Figure 12. The correlation coefficient  $R$  of the three models is  $R_{\text{NRS-RF}} = 0.891$ ,  $R_{\text{KNN-NB}} = 0.824$ , and  $R_{\text{BPNN}} = 0.798$ , respectively. The  $R$  value of the NRS-RF model is closer to 1, which indicates that the prediction value of the model is closer to the actual measurement value, exhibiting better prediction effect.

The NRS-RF model compared with the other two kinds of prediction model shows great advantages, mainly because the initial influencing factors of redundant attributes is more, and they not only increase the classifier identification time but also make the classification accuracy of the RFID system significantly decreased. Through the reduction of NRS algorithm, the kernel factor subset is obtained. The classification of kernel factors set contains stronger characteristic sensitivity, improves the prediction precision, and reduces the computational complexity of the model.

The KNN-NB combination algorithm uses the KNN algorithm to calculate the distance between the sample data to be tested and the sample set. The selected sample data are used as the training sample of the NB algorithm, and then, the NB model is used for prediction and classification, where  $K=3$ . Because the NB model needs to know the prior probability, the prior probability often depends on the hypothetical model. However, there are many kinds of hypothetical models, so in some cases, the prediction effect will be poor due to the hypothetical prior model, so there is a certain error rate in classification decisions.

We use a three-layer BPNN model. The number of hidden layer nodes is set to 6, and the number of output layer nodes is 1. Since the BPNN model is essentially a gradient descent method, the objective function to be optimized is more complicated and prone to the “sawtooth phenomenon,” which makes the convergence speed of the BPNN model slow and affects the final prediction classification accuracy of the test set.

**5.3. Engineering Application.** As intelligent archives management inventory technology rapidly develops [36, 37], the UHF passive RFID technology liberates the archives

TABLE 3: Prediction accuracy comparison.

| Prediction model | Sample data | Correct classification | Error classification | OA (%) | Kappa coefficient |
|------------------|-------------|------------------------|----------------------|--------|-------------------|
| NRS-RF           | 300         | 265                    | 35                   | 88.5   | 0.875             |
| KNN-NB           | 300         | 254                    | 46                   | 84.7   | 0.805             |
| BPNN             | 300         | 245                    | 35                   | 82.3   | 0.784             |

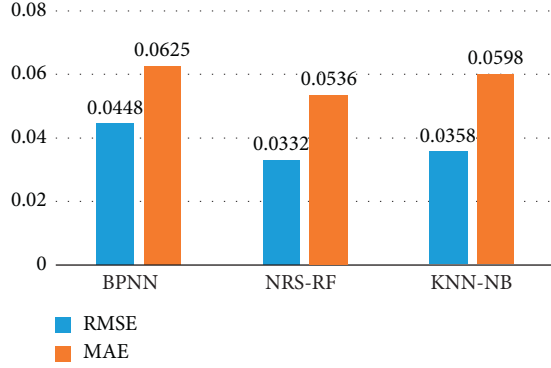


FIGURE 10: Prediction error.

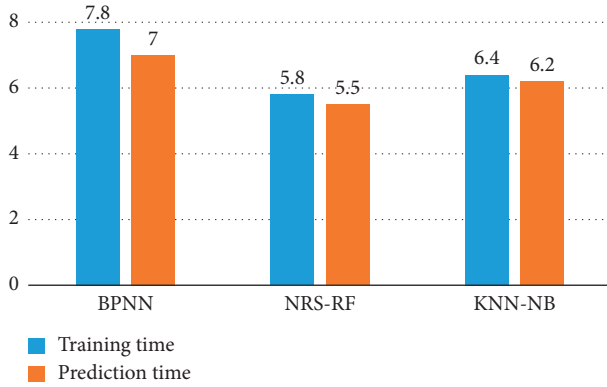


FIGURE 11: Comparison of training time and prediction time in different models.

management from the relatively “stereotyped” impression brought by barcodes. In order to further improve reading efficiency of the archive inventory, the RFID system is mounted on a mobile robotic vehicle, and the effectiveness of the NRS-RF model is verified in archives management applications. Under diverse hardware deployment conditions, this study selects 300 sets for the RFID system identification rate and for influencing factors as the sample data, on which heuristic attribute reduction of the initial influencing factors is conducted using neighborhood rough set theory. The importance distribution of influencing factors is given in this paper, which is obtained by neighborhood rough set reduction and by OOB error analysis of random forest.

The high importance score indicates that the influencing factor has greater impacts and contributions on classification results, as shown in Figure 13. It can be seen that the importance scores all exceed 5, involving antenna height  $P1$ , reader transmission power  $P2$ , distance between the tag and

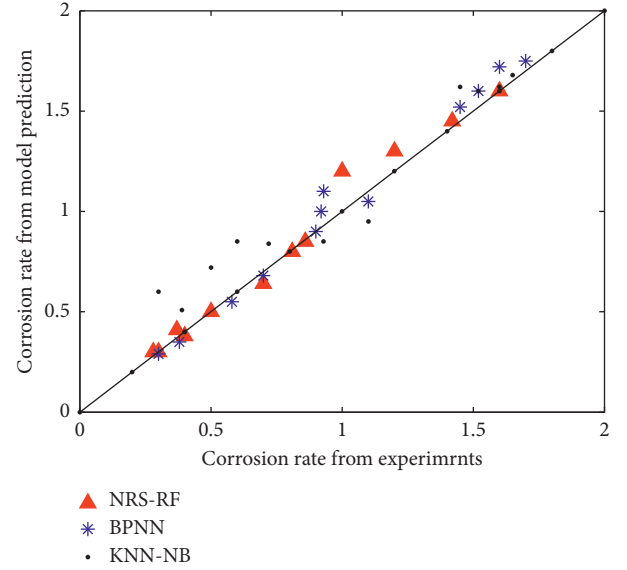


FIGURE 12: Correlation between the predicted value and actual value.

antenna  $P3$ , and other 4 influencing factors. Finally, the set of kernel factors is selected, which include antenna height  $P1$ , number of tags  $P2$ , distance between tags and antennas  $P3$ , reader polling cycles  $P4$ , antenna moving speed  $P5$ , antenna number  $P8$ , and reader transmitting power  $P11$ .

The average decline accuracy rate and average decline Gini coefficient of 7 groups of influencing factors are obtained through analyzing the Gini coefficient. Both the abovementioned average decline accuracy rate and average decline Gini coefficient can represent the degree of decline in accuracy when the influencing factor is replaced, both of which are positively correlated with the importance of the influencing factors. As shown in Figure 14, the larger the values of both the abovementioned rate and coefficient, the higher the importance of the influencing factor. In addition, some influencing samples of the RFID system identification rate are shown in Table 4.

From the 300 sample data, 240 groups were selected as training sets to train the model, and the remaining 60 groups were test set data for verification and prediction. Ten groups of test data were randomly selected from the test set samples, and scatter plots were made, as shown in Table 5 and Figure 15. The RMSE is 0.548, and the correlation coefficient  $R$  is 0.951, indicating outstanding prediction accuracy of the model.

It can be seen from Table 5 that the predicted classification level of 10 groups of test sample data is basically consistent with the actual classification, satisfying the engineering requirements. By comparing the third and fourth



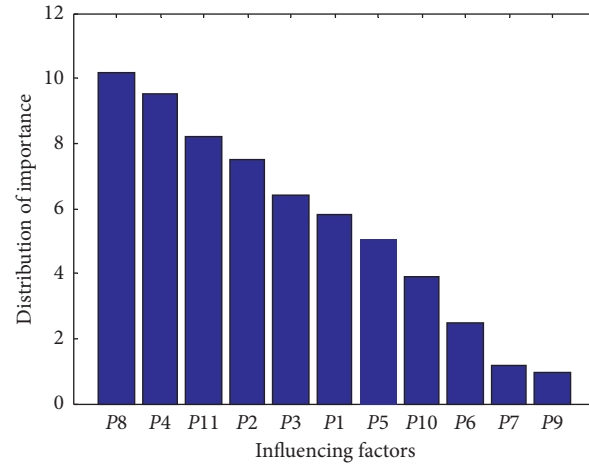


FIGURE 13: Importance distribution of influencing factors.

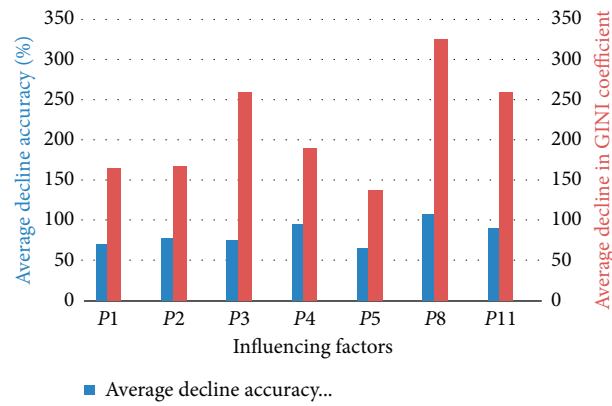


FIGURE 14: Average decline accuracy and average decline Gini coefficient.

TABLE 4: Samples of the RFID system identification rate.

| Sample number | P1  | P2  | P3  | P4 | P5  | P8 | P11 | RFID system identification rate level |
|---------------|-----|-----|-----|----|-----|----|-----|---------------------------------------|
| 1             | 0.6 | 90  | 0.8 | 1  | 0.3 | 2  | 18  | 1                                     |
| 2             | 0.9 | 150 | 1.5 | 2  | 0.6 | 1  | 23  | 2                                     |
| 3             | 0.6 | 90  | 1.9 | 3  | 0.9 | 2  | 28  | 1                                     |
| 4             | 1.2 | 130 | 0.8 | 3  | 0.9 | 1  | 18  | 1                                     |
| 5             | 0.9 | 110 | 1.5 | 2  | 0.3 | 2  | 23  | 1                                     |
| 6             | 1.2 | 180 | 1.9 | 1  | 0.6 | 1  | 28  | 2                                     |

TABLE 5: NRS-RF model RFID system identification rate level prediction.

| Sample number | P1  | P2  | P3  | P4 | P5  | P8 | P11 | Forecast level | Actual level |
|---------------|-----|-----|-----|----|-----|----|-----|----------------|--------------|
| 1             | 0.6 | 90  | 0.8 | 1  | 0.6 | 1  | 18  | 2              | 2            |
| 2             | 0.9 | 150 | 1.5 | 2  | 0.3 | 2  | 23  | 2              | 1            |
| 3             | 1.2 | 130 | 0.8 | 3  | 0.6 | 1  | 18  | 1              | 1            |
| 4             | 1.2 | 130 | 0.8 | 2  | 0.9 | 1  | 18  | 1              | 2            |
| 5             | 0.9 | 180 | 1.5 | 1  | 0.9 | 2  | 23  | 2              | 2            |
| 6             | 1.2 | 90  | 1.5 | 3  | 0.9 | 1  | 28  | 1              | 1            |
| 7             | 0.6 | 180 | 0.8 | 1  | 0.3 | 1  | 18  | 2              | 2            |
| 8             | 0.9 | 180 | 1.5 | 2  | 0.6 | 2  | 23  | 2              | 2            |
| 9             | 1.2 | 90  | 1.9 | 3  | 0.9 | 2  | 28  | 1              | 1            |
| 10            | 0.6 | 150 | 0.8 | 1  | 0.3 | 2  | 18  | 1              | 2            |

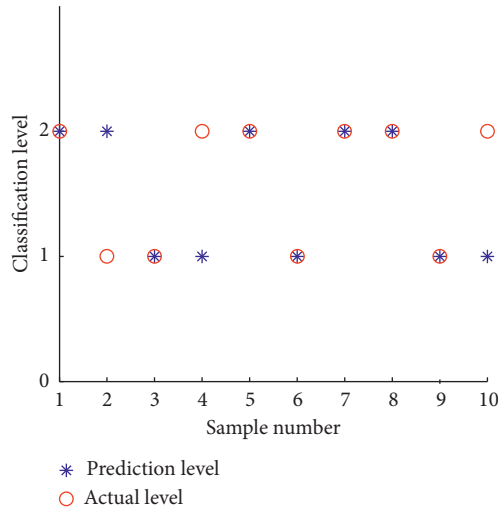


FIGURE 15: Test sample classification.

groups of samples in Table 5, it can be analyzed that the system identification rate can be improved to type 1 by reducing the moving speed of the antenna or increasing the number of reader polling cycles when other influencing factors are consistent. Comparative analysis of the second and fifth groups of samples in Table 5 shows that the system identification rate can be improved to type 1 when the number of tags is reduced or the number of reader polling cycles is increased, thereby accomplishing tag reading to the largest extent. The RFID system identification rate can, therefore, be predicted in a prompt and effective way by mining the relation between the identification rate and the influencing factors. Furthermore, the system identification rate can be improved by purposefully optimizing and adjusting the corresponding hardware deployment, through which the application requirements of more engineering inventories will be satisfied.

## 6. Conclusions

In order to optimize the hardware deployment of the RFID system and improve the system identification rate, a prediction model of the RFID system identification rate based on the combination model of neighborhood rough set and random forest is proposed through mining the relation between relevant influencing factors and the system identification rate. This study uses neighborhood rough set theory to conduct heuristic attribute reduction of weighted dependence of initial influencing factors and takes the kernel factor set as the input variable of random forest model for model training. The model is validated and verified in the RFID experimental test platform. Simulation results suggest that the fitting accuracy of the NRS-RF model is higher than that of the BPNN and other prediction models. Finally, the proposed model is applied to the RFID intelligent archives management platform, thus proving the excellent performance of the NRS-RF model. The proposed model can reversely configure the parameter setting of RFID hardware

deployment, and the system identification rate is, therefore, improved to satisfy the requirements of engineering applications.

Despite the abovementioned findings, the influence of antenna polarization mismatch, multipath fading, or other possible blind zone restrictions is not fully considered in the process of mobile robotic vehicle inventory tags, which may potentially affect the process of tag inventory. Our future research will be focusing on in-depth exploration of the automatic tag counting technology and on realizing the function of automatic tracking and path planning for mobile robots and robotic vehicles, paving way for the future development of automation in tag reading and writing.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This research was funded by the National Social Science Fund of Education Department of Shaanxi Province (no. 2018JK0704), the Science and Technology Plan Project of Xi'an (nos. 201805040YD18CG24-3 and 2019GX173), the Key Research and Development Plan of Shaanxi Province (nos. 2018ZDXM-GY-041 and 2018GY-150), the Science and Technology Research Plan Project of Xianyang (nos. 2018ZDXM-GY-041 and 2018GY-150), and the Team Project of Foshan Entrepreneurship and Innovation (2017IT100032).

## References

- [1] K. Kapucu and C. Dehollain, "A passive UHF RFID system with a low-power capacitive sensor interface," in *Proceedings of the 2014 IEEE RFID Technology and Applications Conference (RFID-TA)*, Tampere, Finland, September 2014.
- [2] Y. L. Liu, S. B. Hu, and F. Y. Yan, "Analysis of key factors of embedded RFID system identification rate," *Journal of Engineering Design*, vol. 12, no. 45, pp. 479–484, 2012.
- [3] Y. Deng, Y. Zhu, Y. F. Yang, X. H. Cheng, and C. Q. Li, "Research on RFID indoor positioning algorithm based on BP neural network," *Miniature Microcomputer System*, vol. 40, no. 8, pp. 1707–1712, 2019.
- [4] L. Y. Qiao, J. G. Yang, H. Zhou, S. S. Dai, and P. C. Fu, "Optimization of key index of reading rate of RFID system based on rough set theory," *Manufacturing Automation*, vol. 36, no. 45, pp. 13–16, 2014.
- [5] M. A. Faraz, S. Yasir, M. Roberto, and C. Noel, "A comparative analysis of machine/deep learning models for parking space availability prediction," *Sensors*, vol. 20, no. 1, pp. 1–17, 2020.
- [6] N. T. Rafia, E. P. Alexander, and M. K. Jong, "Bearing fault diagnosis of induction motors using a genetic algorithm and machine learning classifiers," *Sensors*, vol. 20, no. 1, pp. 1504–1523, 2020.

- [7] K. Ashima, M. Rajni, N. Anand, K. Jeamin, G. K. Byeong, and C. A. Naveen, "Novel deep-learning-based bug severity classification technique using convolutional neural networks and random forest with boosting," *Sensors*, vol. 19, no. 3, pp. 1329–1351, 2019.
- [8] Y. Kang, M. Kim, and H. A. Lee, "A hierarchical structure based reader anti-collision protocol for dense RFID reader networks," in *Proceedings of the 13th International Conference on Advanced Communication Technology*, Seoul, Republic of Korea, December 2011.
- [9] S. Kwangcheol and S. Wonil, "RAC-multi: reader anti-collision algorithm for multichannel mobile RFID networks," *Sensors*, vol. 10, no. 1, pp. 1329–1351, 2010.
- [10] Z. Li, H. Yang, J. Li, C. He, and J. Zhou, "An enhanced neighbor-friendly reader anti-collision algorithm in mobile RFID networks," in *Proceedings of the 2014 4th IEEE International Conference on Information Science and Technology*, Shenzhen, China, April 2014.
- [11] N. Nguyen, B. Liu, and V. A. Pham, "Dynamic-range-based algorithm for reader-tag collision avoidance deployment in RFID networks," in *Proceedings of the 2016 International Conference on Electronics, Information, and Communications (ICEIC)*, Da Nang, Vietnam, January 2016.
- [12] V. Ferro, A. Luz, and A. Lucrécio, "Small long range UHF tag for metal applications," in *Proceedings of the 2013 IEEE International Conference on RFID Technologies and Applications (RFID-TA)*, Johor Bahru, Malaysia, September 2013.
- [13] G. Cicirelli, A. Milella, and D. Di, "Supervised learning of RFID sensor model using a mobile robot," in *Proceedings of the 2011 IEEE International Conference on RFID Technologies and Applications*, Sitges, Spain, September 2011.
- [14] J. Minho, Y. Yonu, S. Cha, and H. Choo, "Mobile RFID tag detection influence factors and predict detectability," *Sensors*, vol. 9, no. 2, pp. 112–119, 2009.
- [15] Keonn, "Shop keonn UHF RFID products," 2020, <https://www.atlasrfidstore.com/keonn-rfid-readers-antennas/>.
- [16] J. Liu, W. Li, Z. Wen, S. Xiao, L. Yang, and F. Tan, "An energy efficiency approach via self-adaptive duty-cycle in social contacts scenario," in *Proceedings of the 25th International Conference on Parallel and Distributed Systems (ICPADS)*, Tianjin, China, December 2019.
- [17] H. Wang, S. Wang, J. Yao et al., "Effective anti-collision algorithms for RFID robots system," *Assembly Automation*, vol. 40, no. 1, pp. 55–64, 2019.
- [18] Impinj speedway connect software for Impinj RFID readers, <http://www.idshi.com/ProductView.aspx?Id=84>, 2020.
- [19] Z. H. Zhou, *Machine Learning*, pp. 121–146, Tsinghua University Press, Beijing, China, 2016.
- [20] Z. Deng, Z. Zheng, D. Deng, T. Wang, Y. He, and D. Zhang, "Feature selection for multi-label learning based on F-neighborhood rough sets," *IEEE Access*, vol. 8, no. 1, pp. 39678–39688, 2020.
- [21] C. Wang, Y. Huang, M. Shao, and X. Fan, "Fuzzy rough set-based attribute reduction using distance measures," *Knowledge-Based Systems*, vol. 164, no. 1, pp. 205–212, 2019.
- [22] C. Wang, Y. Wang, M. Shao, Y. Qian, and D. Chen, "Fuzzy rough attribute reduction for categorical data," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 5, pp. 818–830, 2019.
- [23] C. Wang, Y. Huang, M. Shao, Q. Hu, and D. Chen, "Feature selection based on neighborhood self-information," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 4031–4042, 2020.
- [24] J. Zhao and X. Pan, "Remote sensing image feature selection based on rough set theory and multi-agent system," in *Proceedings of the 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, Zhangjiajie, China, August 2015.
- [25] M. Ren, Y. Qu, and A. Deng, "Covering rough set-based three-way decision feature selection," in *Proceedings of the 2018 Tenth International Conference on Advanced Computational Intelligence (ICACI)*, Xiamen, China, March 2018.
- [26] V. K. Y. Wu and N. H. Vaidya, "RFID trees: a distributed RFID tag storage infrastructure for forest search and rescue," in *Proceedings of 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, Boston, MA, USA, July 2010.
- [27] P. N. Samarakoon, N. E. P. Promayon, and C. Fouard, "Light random regression forests for automatic multi-organ localization in CT images," in *Proceedings of the 14th International Symposium on Biomedical Imaging (ISBI 2017)*, Melbourne, Australia, June 2017.
- [28] N. Spolaôr and M. C. Monard, "Evaluating ReliefF-based multi-label feature selection algorithm," *Advances in Artificial Intelligence—IBERAMIA 2014*, vol. 40, no. 1, pp. 194–205, 2014.
- [29] Y. Lin, Q. Hu, J. Liu, and J. Duan, "Multi-label feature selection based on max-dependency and min-redundancy," *Neurocomputing*, vol. 168, no. 1, pp. 92–103, 2015.
- [30] M. C. Yao and M. F. Chao, "A greedy algorithm for attribute reduction," *Journal of Jiamusi University (Natural Science Edition)*, vol. 6, no. 12, pp. 125–130, 2013.
- [31] L. I. Mina and C. W. Donga, "Exploration greedy algorithm with attribute reduction," *Computer Engineering*, vol. 38, no. 19, pp. 163–166, 2013.
- [32] Random forest toolbox, [https://download.csdn.net/download/u011727648/10502334?utm\\_source=bbsseo](https://download.csdn.net/download/u011727648/10502334?utm_source=bbsseo), 2020.
- [33] B. Sanniv, K. Anwesha, B. Rohit, and D. N. Tibarewala, "Classification of lower limb motor imagery using K-nearest neighbor and naïve-bayesian classifier," in *Proceedings of the 3rd International Conference on Recent Advances in Information Technology (RAIT)*, Dhanbad, India, July 2016.
- [34] Y. Yu and L. Zhang, "WSN location method based on BP neural network in NLOS environment," in *Proceedings of the International Conference on Wireless Communication and Sensor Network*, Wuhan, China, December 2014.
- [35] D. Xu, Y. Wang, P. Peng, S. Beilun, Z. Deng, and H. Guo, "Real-time road traffic state prediction based on kernel-KNN," *Transportmetrica A: Transport Science*, vol. 16, no. 1, pp. 104–118, 2020.
- [36] Q. L. Shen, B. Shao, and L. J. Chen, "Design and implementation of book inventory robot based on UHF RFID," *Library Science Research*, vol. 7, no. 1, pp. 24–28, 2016.
- [37] N. Zhang and Y. L. Gou, "Application of new insect-proof RFID tags in smart archive management," *Lantai World*, vol. 7, no. 1, pp. 103–105, 2018.