

Research Article

An Improved Stacking Model for Equipment Spare Parts Demand Forecasting Based on Scenario Analysis

Yi Yang,^{1,2} Wei Liu ,¹ Tingting Zeng,¹ Linhan Guo ,¹ Yong Qin ,³ and Xue Wang⁴

¹School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China

²State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

³Peng Cheng Laboratory, Shenzhen 518000, China

⁴Binzhou Polytechnic, Binzhou 256603, China

Correspondence should be addressed to Linhan Guo; linhanguo@buaa.edu.cn and Yong Qin; yqin@bjtu.edu.cn

Received 26 October 2021; Accepted 19 May 2022; Published 14 June 2022

Academic Editor: Liang Zou

Copyright © 2022 Yi Yang 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.

The purpose of spare parts management is to maximize the system's availability and minimize the economic costs. The problem of cost availability trade-off leads to the problem of spare parts demand prediction. Accurate and reasonable spare parts demand forecasting can realize the balance between cost and availability. So, this paper focuses on spare parts management during the equipment normal operation phase and tries to forecast the demand of spare parts in a specific inspection and replacement cycle. Firstly, the equipment operation and support scenarios are analyzed to obtain the supportability data related to spare parts requirements. Then, drawing on the idea of ensemble learning, a new feature selection method has been designed, which can overcome the limitations of a single feature selection method. In addition, an improved stacking model is proposed to predict the demand for spare parts. In the traditional stacking model, there are two levels of learning, base-learning, and meta-learning, in which the outputs of base learners are taken as the input of the meta learner. However, the proposed model brings the initial feature together with the output of the base learner layer as the input of the meta learner layer. And experiments have shown that the performance of the improved stacking model is better than the base learners and the traditional stacking model on the same data set.

1. Introduction

Accurate spare parts demand forecasting can ensure reasonable spare parts inventory, which can reduce the downtime loss or unnecessary inventory cost [1]. Different spare parts have different demand patterns, and the factors influencing spare parts demand in different phases differ a lot. Besides, the adoption of the prediction method is limited by data availability, so different prediction methods are required in different situations [2]. Generally, prediction methods can be divided into qualitative and quantitative. Qualitative methods are mainly based on the subjective judgment of decision-makers or experts, which often lack theoretical basis and are difficult to explain. In contrast, quantitative methods can be divided into time series and causal methods. The former is only based on historical demand data, while the latter considers other explanatory

variables that may affect the target value. And it can be subdivided into simple linear regression and machine learning methods. The main difference between linear regression and machine learning is that the former determines the model structure in advance, while the latter learns the relationship between input and output from data [3]. Compared with the statistical methods, machine learning is more suitable for fitting complex nonlinear relationships [4]. Machine learning has been widely used in the prediction field, such as hourly demand prediction of rental bikes [5], pharmacy drug demand estimation [6], and patients' demand prediction in hospitals [7]. In [8], clustering methods and XGBoost were used to assess vehicle driving and predict risk levels. In [9], grey neural networks, BP neural networks, SVM, and so on were used to train data and establish demand prediction models. There is a complex nonlinear relationship between spare parts demand and many

explanatory variables, so machine learning method is considered to predict spare parts demand.

In the task of spare parts demand forecasting based on machine learning, we need to first analyze the factors that may affect spare parts demand, and then make feature selection. Generally, the demand for spare parts comes from the process of operation and maintenance [10]. Demand for spares may be related to failure occurrence, maintenance activities, equipment age, and operating conditions, the environmental conditions, such as weather condition, past usage patterns such as fleet size, the annual budget from departments, flying hours, sorties flown, management, and reliability characteristics of spare parts should also be considered [11–13]. However, previous studies' analysis of influencing factors is not comprehensive and detailed. Many studies have mentioned that operating conditions and environment, maintenance activities, flight hours, takeoff and landing times, reliability, fleet size, management factors, and other factors may affect the demand for spare parts, but factors such as environmental variables and management are always omitted. Besides, the specific operation process of equipment, the maintenance support process of equipment spare parts, including the factors of spare parts themselves, may also affect the demand for spare parts. Therefore, through the analysis of the three-level maintenance support process of equipment spare parts, this paper tries to obtain more comprehensive potential factors affecting the demand of spare parts.

In the field of machine learning, feature selection is a data preprocessing technology. Selecting more compact feature subsets from existing features for modeling can improve the accuracy and generalization ability of the model and reduce the running time. [14, 15]. As demonstrated in Chandrashekar's study, the forecast model's performance depends not only on the model structure but also on the feature space [16]. Feature selection methods are generally divided into four categories, including filters [17–20], wrappers [21–24], and embedded [25] and hybrid methods [26, 27]. New feature selection methods based on ensemble learning have also emerged in recent years [28]. In the ensemble feature selection method, many different feature selectors are generated first, and then the outputs of these individual selectors are aggregated and returned as the final integrated result [29, 30]. The ensemble feature selection method also has been used in some fields, such as biomarker recognition in a cancer diagnosis process, and experiments have shown that it can significantly improve the stability and classification performance of biomarkers [31]. Taheri and Nezamabadipour [32] proposed a feature selection method of high-dimensional data based on the ensemble method. The performance of which has shown that it has more advantages than other methods in classification accuracy and feature reduction rate.

The purpose of researchers using ensemble learning is similar to that of people weighing various opinions and

making decisions in daily life. For example, in a democratic society, citizens vote to elect officials or make laws. Ensemble learning completes the learning tasks by constructing and combining multiple learners. First, a group of individual learners is generated, and then they are combined by some strategy. In 1979, Dasarathy and Sheela [33] first proposed the idea of ensemble learning. Since then, ensemble learning has become an important research field of machine learning. In 1992, Wolpert [34] proposed the stacking generalization model. Breiman [35] suggested a bagging algorithm in 1996. And Freund and Schapire [36] proposed the AdaBoost algorithm in 1997. Dietterich [37] explained three fundamental reasons for the success of the ensemble learning from a mathematical point of view: statistics, calculation, and representativeness.

Later, more ensemble learning algorithms have been proposed, and ensemble methods have been successfully applied in pattern classification, regression, and forecasting problems. For example, in [38], ensemble technique, bagging, and boosting were used to conduct anti-aircraft missile spare parts demand to forecast. Matrenin et al. [39] solved the power system medium-term load forecasting problem based on the ensemble machine learning method, in which the AdaBoost model has the best prediction performance. Gutierrez-Espinoza et al. [40] used the ensemble learning method to identify false information. The experimental results show that both the bagging ensemble and AdaBoost ensemble methods can detect deceptive information better than the traditional machine learning methods.

As a kind of ensemble learning technology, model fusion solves problems by combining multiple models in some way. In the era of big data, massive multimodal data exist widely. Data fusion method and model fusion method are essential means of multimodal data analysis and mining [41–43]. And the learning performance of model fusion is often better than that of the narrow sense ensemble learning model. Many papers have explained reasons from the perspectives of training basis, hypothetical space, and starting point of calculation [44, 45]. Theoretically, model fusion can reduce variance and deviation simultaneously and find the best advantage of variance deviation equilibrium.

Now, the widely used fusion methods include averaging, voting, stacking, and so on. The first two methods fuse the results of the base learners by voting or averaging, which can only improve the fusion results to a certain extent. However, stacking model uses meta learner to fuse the assumptions made by the base learner, which is equivalent to looking for the best fusion rules. Therefore, we can train the meta learner in the direction of minimizing the loss function, which is more reliable and efficient than the averaging and voting methods. Therefore, stacking is more effective than voting and averaging and has been widely used in various fields. For example, a stacking ensemble learning model was adopted to predict the short-term wireless network load and car-hailing demand [46, 47]. In [48], a stacking-based ensemble model

was deployed to forecast solar radiation strength. Liang et al. [49] and Al-Sarem et al. [50] adopted a stacking learning framework for genomic prediction and phishing website detection. In the stacking model, selecting the appropriate base models, optimizing the hyper-parameters of the model, and improving the generalization ability are the challenges we have to face, and many scholars have made attempts to solve the above problems. Al-Sarem et al. [50] adopted a genetic algorithm (GA) to tune the parameters of machine learning learners, Cui et al. [51] developed an improved swarm intelligence algorithm to optimize the critical parameters in the prediction model. In [52], the Pearson Correlation analysis was used to measure the correlation between base learners to select the basic learners with less correlation. In [53], the proposed deep ensemble learning framework adopted two sparse auto-encoders to reduce the correlation of attributes and diversify the base classifiers. Pari et al. [54] proposed a multi-tier stacked ensemble algorithm, in which the cross-validated predictions are combined to generate new features by different combination schemes.

Spare parts demand forecasting is an important issue in spare parts management, but in this field, ensemble learning and model fusion are rarely used, especially the stacking model, which has a large utilization space. Therefore, this paper introduces the stacking model in ensemble learning into the field of spare parts management to predict the demand for spare parts. Besides, the stacking model constructed in this paper has only three base models, which will lead to a small number of features of the meta model and may affect the prediction accuracy. Therefore, this paper proposes an improved stacking model which adds the initial features to the feature matrix of the meta model to solve this problem. And the main contributions of this paper are as follows:

- (1) Potential factors affecting spare parts demand that may not have been considered before are obtained and quantified based on the scenario analysis of the operation and support process of aviation equipment;
- (2) An integrated feature selection framework based on the three feature selection methods is proposed, which can reduce the one-sidedness and limitation of a single feature selection method;
- (3) In addition, the stacking model is introduced into the field of spare parts demand prediction, and the improved stacking model tries to solve the problem of insufficient number of meta learner features in the traditional stacking model, which effectively improves the generalization ability and prediction accuracy of the model.

2. Equipment Support Scenario Analysis

The so-called scenario was initially a concept in product design. From a macro perspective, it refers to the needs of users to use the product, such as analyzing what problems a

product solves for which target users; from the micro perspective, it is the specific use details, such as the user's specific search and browsing scenes, which reflects the interactive experience in the details of a product. Scenario modeling is a method to describe and analyze the user requirements. In the form of stories, it tells the whole process of target users using products to complete target tasks in a specific environment. The 5W1H method [55] is usually used to mine the six elements in the scenario, as shown in Table 1:

Therefore, we apply the concept of scenario modeling to the context of equipment integrated support, digitize the support scenario of equipment spare parts, and analyze the potential factors considered by the staff when planning the spare parts requirements.

By analyzing the spare parts supply and support process, we can summarize the potential factors affecting spare parts consumption, but this is not enough. Comprehensive and detailed analysis and digital description of equipment operation and support scenarios can get more detailed features.

Taking the annual flight hours as an example, many studies have taken the total flight hours as an influencing factor. However, different equipment utilization may result in various spare parts requirements when the total flight hours are determined. In this way, the flight time can be subdivided into flight time under different natural environments, different operating environments, different training subjects, different personnel operations, and different support and maintenance conditions. These differences may be the potential factors affecting the demand for spare parts under the premise of a certain total flight time.

And these differences may be the potential factors affecting the demand for spare parts under the premise of a certain total flight time.

2.1. The Three-Level Support System for Military Spare Parts. The flight regiment usually performs various training tasks under various working conditions.

Figure 1 shows the three-level maintenance support process of repairable parts of a flight regiment. The outfield spare parts warehouse and the outfield maintenance depot are responsible for the support and maintenance of LRU (Line Replaceable Unit), respectively. Spare parts will be demanded in the process of equipment operation and maintenance activities. Therefore, we will analyze the operation and support scenarios of equipment.

2.2. Quantify the Influencing Factors. The consumption of spare parts is related to the task quantity and the consumption law, while the former includes flight hours, takeoff and landing times, operation conditions, natural environment, and other factors, while the latter is related to the attributes of spare parts, such as loss sensitivity, and reliability. Besides, the replenishment of spare parts mainly comes from returning to the warehouse after repair, so factors related to maintenance should be considered. So, the 5W1H method is applied to the field of equipment support,

TABLE 1: Mining six elements of scenario by 5W1H method.

Element	Interpretation
Who	Who is the target user of the product
What	What tasks do users do with products, and what are their goals
When	When will users use the product
Where	The actual environment in which users use the product
Why	Potential rational and perceptual factors behind user behavior
How	What measures do users take to perform tasks

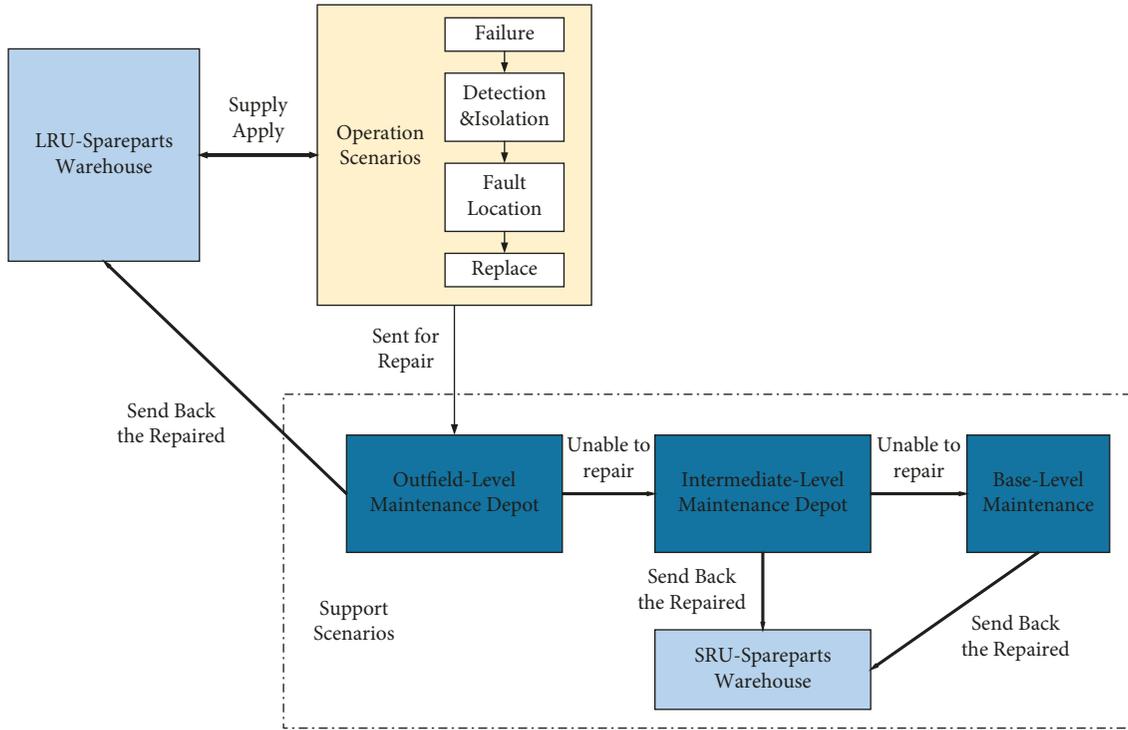


FIGURE 1: Three-level maintenance support process of repairable spare parts.

and the equipment operation and support scenarios are analyzed in Table 2:

- (1) According to the scenario analysis, the potential factors that may affect the demand for equipment spare parts are obtained, which are shown in Table 2.
- (2) Based on scenario analysis, we have obtained the relevant data of five kinds of spare parts from a department of the Chinese Air Force, which generated from 2011 to 2020. The data set has 34 attributes, corresponding to the attributes in Table 2, and the label is the monthly demand quantity of spare parts.
- (3) Due to military confidentiality requirements, the data used in this paper cannot be made public. To support analysis, the data have been standardized and normalized.

3. Model Framework

Based on the analysis of equipment spare parts use and support scenarios, the potential factors affecting spare parts demand are

obtained. The initial feature set is established after data cleaning and processing. After that, to eliminate irrelevant or redundant features, this paper proposes a feature selection method based on ensemble learning, which obtains the final feature set by aggregating the output of several individual feature selectors. And then, a stacking framework is established. First, G grid-search CV is used to tune the parameters of base learners, and then cross-validation is used to determine the best combination of base learners. The stacking model combines multiple prediction models to get a more accurate model to predict the demand for spare parts. In addition, an improved stacking model is proposed to solve the problem of less features of meta learners, and the effectiveness of the method is verified by comparing the performance of different models. And the model framework is shown in Figure 2.

3.1. Feature Selection. Inspired by ensemble learning methods, an ensemble feature selection method is proposed and shown in Figure 3, ensemble specifically means synthesizing the results of multiple feature selection methods to obtain the final feature set. And the three base feature

TABLE 2: Analysis of equipment operation and support scenarios based on the 5W1H method.

5W1H Scenarios	Who	When	Where	Why	What	How
Spare Part Attributes	Variety	Failure Time - Year	Maximum flight Altitude	Tough/ Special Tasks	Subject 1 flight hours	Flight days
	Number of Single Machine Installations					
Operation Scenario	Importance	Average Temperature	Average flight Altitude	General/ Daily Tasks	Subject 2 flight hours	Flight hours
	Loss Sensitivity		Plain flight hours		Subject 3 flight hours	Takeoff and Landing Times
	Price	Average Humidity	Plateau flight hours	Subject 4 flight hours	Sorties Rate	
		Bad flight Days				
Support Scenario	Maintenance Personnel Occupancy Rate	Mean Time to Repair	Outfield Repair Rate	Repair Rate	On-Site Rate of Maintenance Equipment Lead Time	

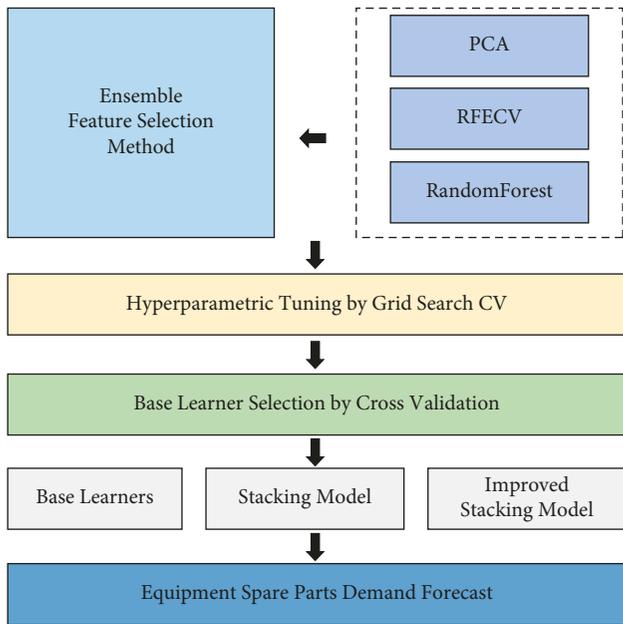


FIGURE 2: Model framework.

selection methods we chose are Principal Component Analysis (PCA), Recursive Feature Elimination and Cross-Validation (RFECV), and Random Forest (RF). PCA, RFECV, and RF are the three commonly used feature reduction methods, and in these three methods, we can find a parameter to measure the importance of features. In the PCA method, explained variance ratio represents the proportion of the variance value of each principal component in the total square difference after dimension reduction. Therefore, explained variance ratio is used to rank the feature importance. In the RFECV method, all features' importance evaluation is performed by the recursive feature elimination. The selected feature set is cross verified to determine the number of features with the highest average

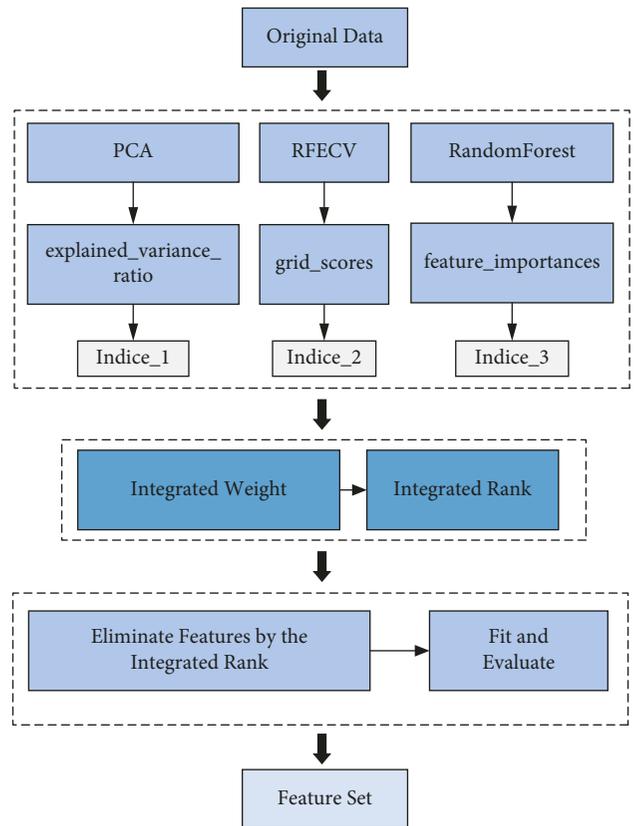


FIGURE 3: The proposed ensemble feature selection method.

score. In the RF method, the features are arranged in descending order through random forest, and a certain proportion of features are deleted to obtain a new feature set. The feature set with the lowest out-of-bag error rate is selected as the final feature set.

Then, according to the feature importance parameters, namely, explained variance ratio, grid scores, and feature

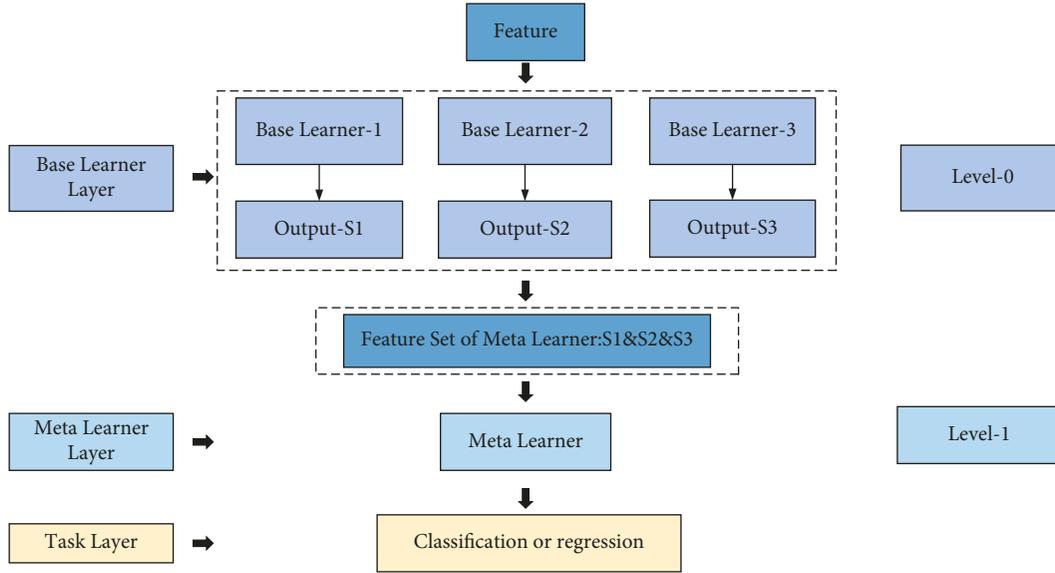


FIGURE 4: A typical stacking model with three base learners.

importance, the importance of features are sorted into three sequences, namely, $indice_1$, $indice_2$, and $indice_3$. Let i be a feature, the weight of feature i can be obtained by its position in the sequence, which is represented by $index(i)$. In the three sequences, the position of feature i may be different, which is represented by $indice_1.index(i)$, $indice_2.index(i)$, and $indice_3.index(i)$. So, the comprehensive weight of features can be obtained by synthesizing the position of features in the three methods. So, we define the comprehensive weight of feature i as $weight(i)$:

$$weight(i) = \frac{1}{indice_1.index(i) + 1} \times \alpha_1 + \frac{1}{indice_2.index(i) + 1} \times \alpha_2 + \frac{1}{indice_3.index(i) + 1} \times \alpha_3, \quad (1)$$

where α_1 , α_2 , α_3 represents the weight defined by different base feature selection methods according to the effect of feature selection, and $\alpha_1 + \alpha_2 + \alpha_3 = 1$.

Based on this, the features can be reordered by $weight(i)$, then, according to the feature importance, the features are removed in turn and put into the model training and testing. The optimal feature subset is determined by evaluating the model performance. And the model used in this paper is RandomForest, which can be called by Scikit-Learn.

3.2. Stacking Model. The core idea of stacking is to take the output of the base learner as the input of the meta learner. Base learners can also be called level 0 learners, and the meta learners can be called level 1 learners. Meta learners are stacked on top of base learners, which explains why this model is called stacking model. The structure of a simple

stacking model composed of three base models and one meta model is shown in Figure 4.

3.2.1. Cross-Validation. In the stacking model, the individual learner will train and predict on the original data, and then arrange the prediction results into a new characteristic matrix and put them into the meta learner for learning, which will inevitably lead to the problem of small sample size in the feature matrix of meta learner. When training the stacking model, the original data set is divided into three parts: training set, verification set, and test set. The test set is used to detect the effect of the whole fusion model, so it cannot be used in the training process. The training set is used to train individual learners, which has been wholly disclosed to individual learners, and the prediction result cannot represent the generalization ability of individual learners. Therefore, only the verification set can be used to predict and describe the real learning level of individual learners, leading to insufficient sample size in the feature matrix of the meta learner.

To solve this problem, cross-validation is used to expand the feature matrix of meta learners.

Figure 5 shows the stacking model composed of three base learners and one meta learner. Firstly, threefold cross-validation is performed on the three base learners, and the training data is divided into three parts. Two of them are used as the training set, and the left is used as the validation set. Each base learner is trained on the training set and performs prediction on the validation set and the whole test set. Repeat this process three times to get the prediction data of three validation sets and the prediction data of the test set.

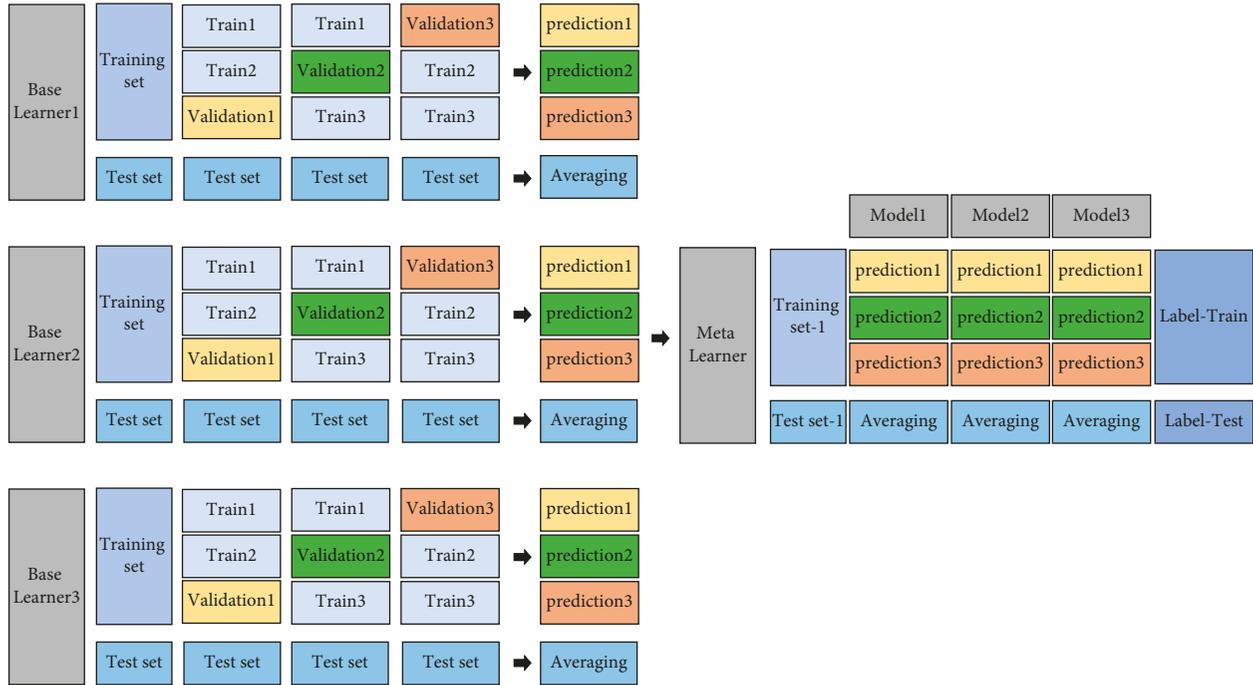


FIGURE 5: Stacking model and threefold cross-validation process.

Then, the prediction data of the three validation sets are stacked vertically, and the prediction data of the three test sets are averaged. The prediction data of the validation sets of the three base learners are spliced horizontally to obtain the training set of the meta learner, and the prediction data of the test sets of the three base learners are joined horizontally to get the test set of the meta learner. In this way, the problem of small sample size of meta learner is solved.

Generally speaking, before model fusion, all possible algorithms need to be run once, and the algorithm with better performance is selected as the basis of fusion.

Cross-Validation can be used to select the base learners in the stacking model. Specifically, a better model can be chosen by comparing the CV results of different models on the same data set. And the steps of model selection by cross-validation are as follows:

- (1) The total training set is divided into k parts, namely, S_1, S_2, \dots, S_k , taking S_j as the validation set while the others are the training sets;
- (2) For each model M_i , the algorithm executes k times, training the model M_i by the other nine training sets, and verifying M_i by S_j . And the generalization error of M_i can be obtained by averaging the errors of k times;
- (3) The model with minimum generalization error is selected as the final model.

In order to determine the best combination of base learners, eight candidate models are selected and shown in Table 3, all of which are regression models.

The cross-validation results are shown in Figure 6, and through Grid-search CV, the hyper-parameters of models

TABLE 3: Candidate models of stacking framework.

Candidate Models	model_xgbr	XGBRegressor
	model_gbr	GradientBoostingRegressor
	model_rf	RandomForestRegressor
	model_bagging	BaggingRegressor
	model_lr	LinearRegression
	model_dtr	DecisionTreeRegressor
	model_abr	AdaBoostRegressor
	model_ridge	Ridge

are determined, and the final base learners are shown in Table 4. Generally speaking, the meta learner needs to be a low complexity algorithm, so a simple linear model is selected as the meta learner.

3.2.2. *Stacking Model and Improved Model.* Based on the above analysis results, this paper builds a stacking model framework, as shown in Figure 7:

However, for the typical stacking model, it is inevitable that the number of features in the feature matrix of the meta learner is insufficient, because an individual learner can only output one set of prediction results, and the number of features in the new feature matrix is equal to the number of individual learners. Taking the stacking model built above as an example, since there are only three base learners, only three features are finally input to the meta learner. To solve this problem, we can consider adding the original feature matrix to the predicted values of individual learners to form a new high-dimensional feature matrix, which can be used as the input of meta learners. And the improved model structure is shown in Figure 8:

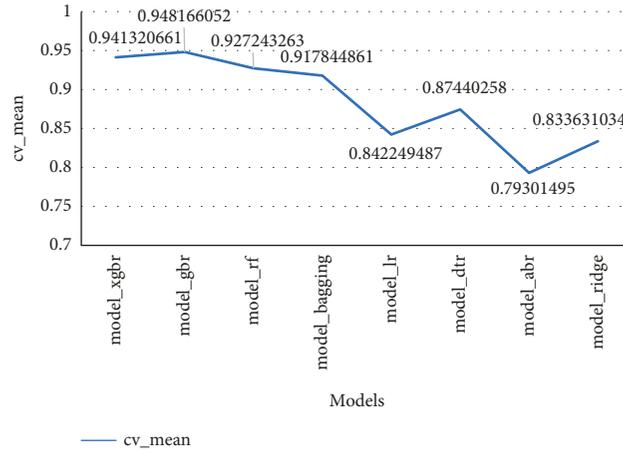


FIGURE 6: Stacking Model and threefold Cross-Validation Process. (cv_mean is the cross_val_score of models on the test set, which is R2_Score, the coefficient of determination.).

TABLE 4: Base and meta models of the stacking framework.

Base Learners	model_xgbr model_gbr model_rf	xgb.XGBRegressor(random_state = 42) ensemble.GradientBoostingRegressor(n_estimators = 100) ensemble.RandomForestRegressor(n_estimators = 20)
Meta Learner	model_lr	linear_model.LinearRegression()

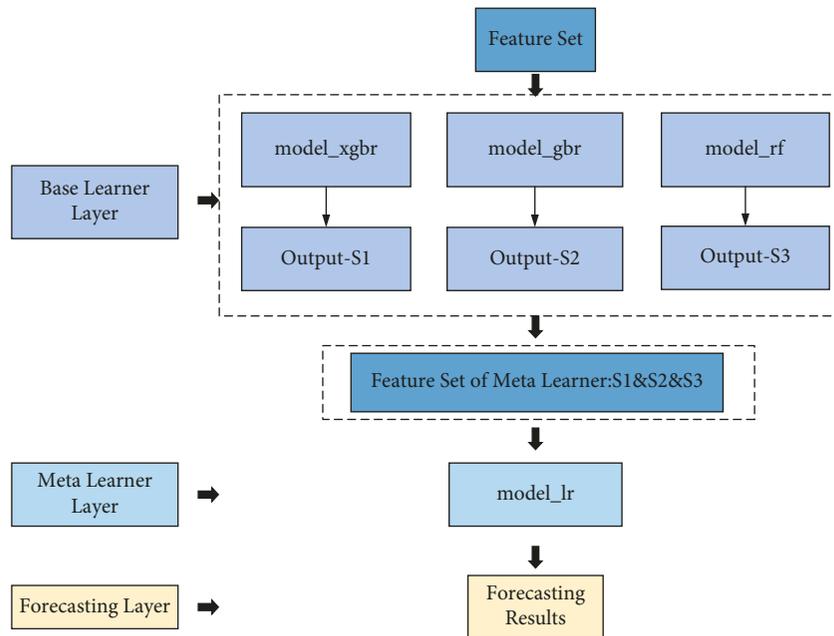


FIGURE 7: Stacking model.

4. Result Discussion

This section will discuss the results of feature selection and spare parts demand prediction based on the stacking model.

4.1. Feature Selection and Experimental Data. By analyzing the operation and support scenarios of equipment, the

factors that may affect the demand for spare parts are obtained. After the digital description, the digital features are obtained, which can be used as the input of the stacking model after feature selection. Figure 1 compares the feature selection method based on the idea of integrated learning proposed in this paper with the other three methods. We eliminate features in turn according to the increasing direction of feature importance, input the features into the

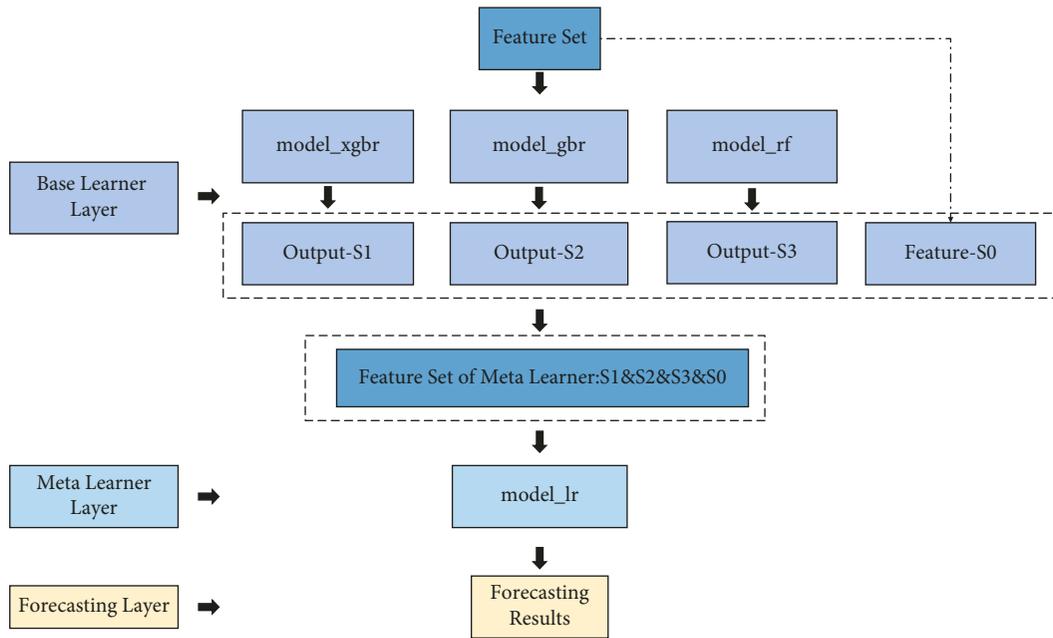
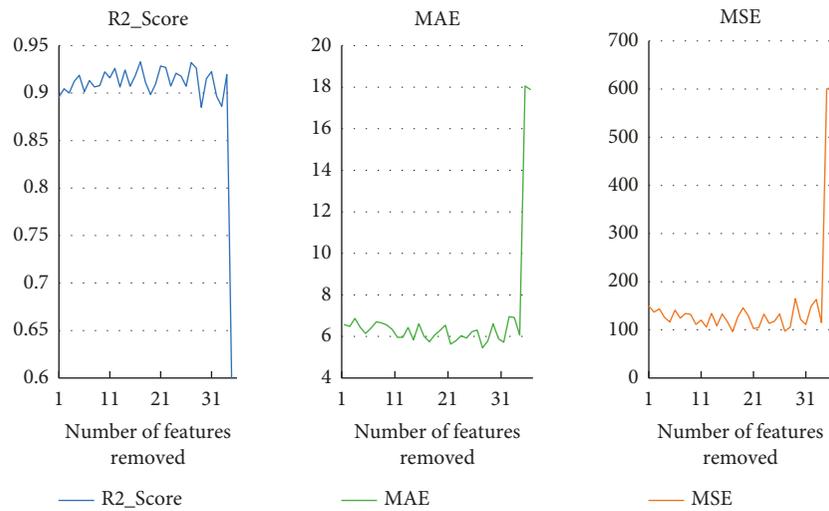
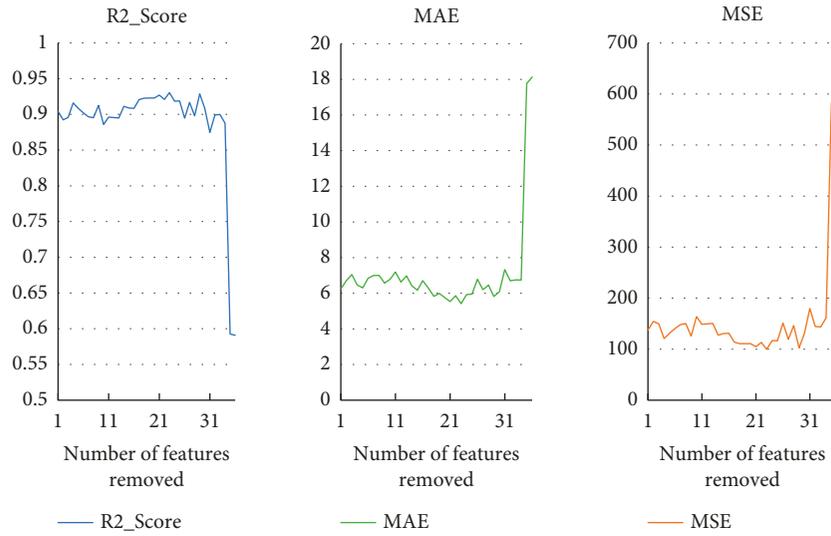


FIGURE 8: The improved stacking model.

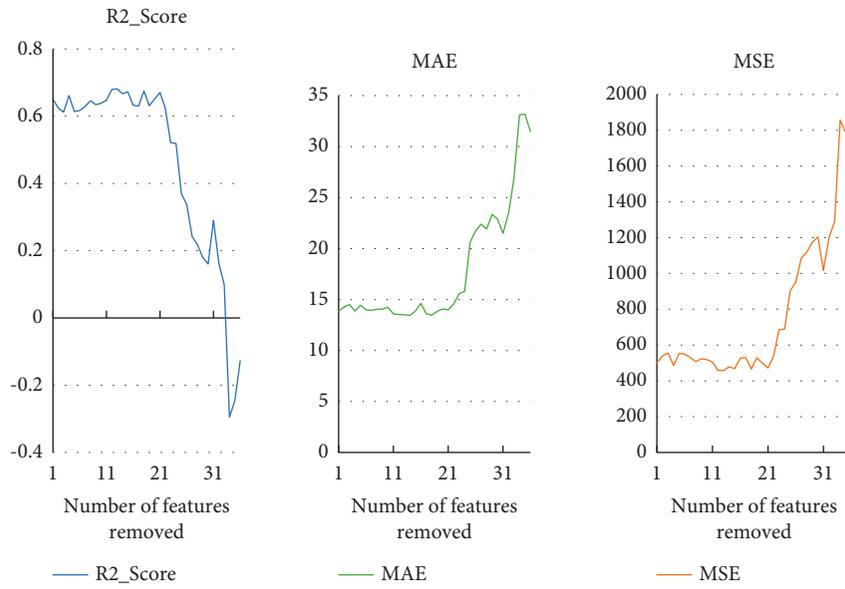


(a)

FIGURE 9: Continued.



(b)



(c)

FIGURE 9: Continued.

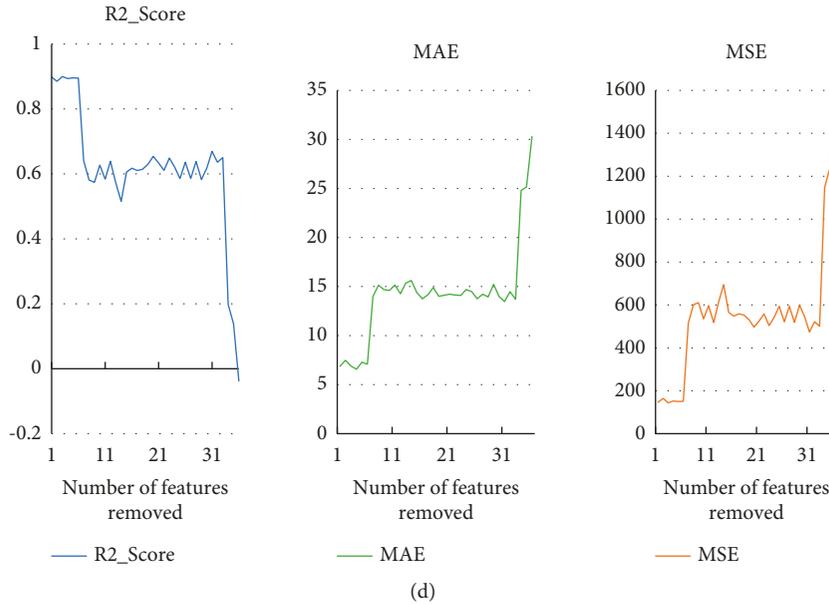


FIGURE 9: Performance of Different Feature Selection Methods. (a) The Proposed Ensemble Feature Selection Method. (b) RF Feature Selection Method. (c) PCA Feature Selection Method. (d) RFECV Feature Selection Method. (R2_Score is the coefficient of determination; MAE is mean_absolute_error, the expected value of the absolute error loss; MSE is mean_squared_error, the expected value of the squared (quadratic) error or loss.).

random forest model for training and testing, evaluate the feature selection method according to the model performance and prediction accuracy, and determine the most appropriate number of features. And the experimental results are shown in Figure 9: (1) The ensemble feature selection method proposed in this paper can optimize the model’s performance and minimize the prediction error. (2) After eliminating 17 features in turn according to the increasing direction of feature importance, the optimal feature set with the best performance of the model can be obtained.

Taking the features before and after feature selection as the input of the three base learners, comparing the model’s performance, it is found that the selected features can effectively improve the model’s performance according to Figure 10.

4.2. Performance and Result Discussion. Based on the selected features, the spare parts demand is predicted by the stacking model and the proposed improved stacking model, and compared with the prediction performance and error of the base learners. The results are shown in Figure 11.

For each individual evaluator in model fusion and the fusion model, we perform cross-validation on the training

set and final test on the test set. The goal we hope to achieve is that the results of cross-validation of the proposed model and the results on the test set should be as good as possible, which means that the model’s generalization ability is improved. According to the experimental results, we can draw the following conclusions:

It can be seen from figure (a) that the stacking model performs better than the other models in the score of cross-validation, and the proposed improved stacking model significantly improves this score, indicating that the generalization ability of the stacking model is much stronger. Figure (b) reflects the difference between the scores of different models in the training set and the test set, and according to figure (b), the base models have a certain degree of over-fitting. In contrast, the stacking model effectively reduces over-fitting, and the proposed improved model minimizes this gap. Similar to figure (a), the stacking model and improved stacking model in figure (c) can effectively improve the score on the test set, that is, it can effectively improve the generalization ability of the model. Figure (d) shows the prediction errors of different models on the test set, which can be effectively reduced by the stacking model and improved stacking model.

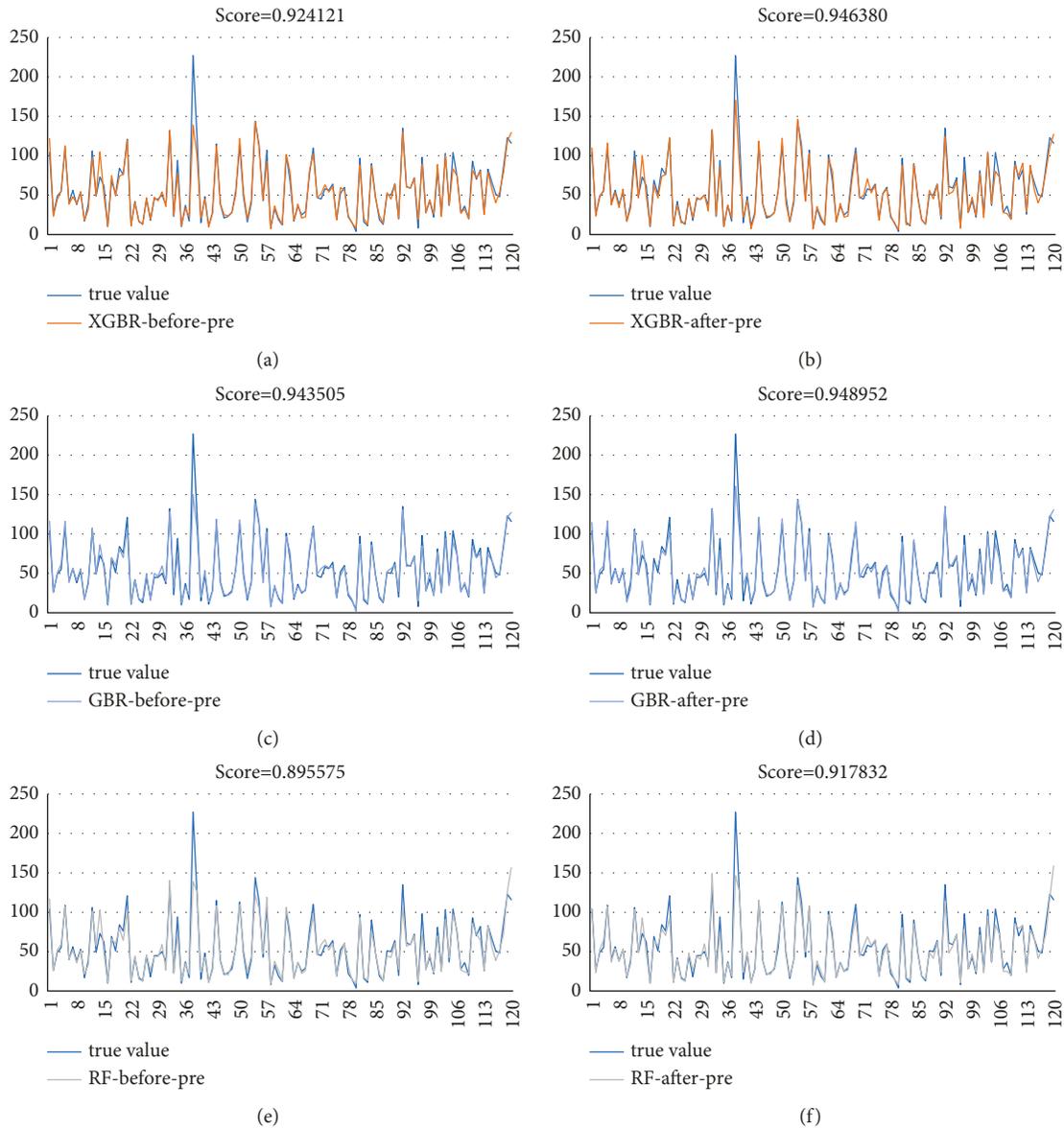


FIGURE 10: Comparison of Model Performance before and after Feature Selection. (a) Performance of model_xgbr before Feature Selection. (b) Performance of model_xgbr after Feature Selection. (c) Performance of model_gbr before Feature Selection. (d) Performance of model_gbr after Feature Selection. (e) Performance of model_rf before Feature Selection. (f) Performance of model_rf after Feature Selection. (score returns the coefficient of determination R2 of the prediction.).

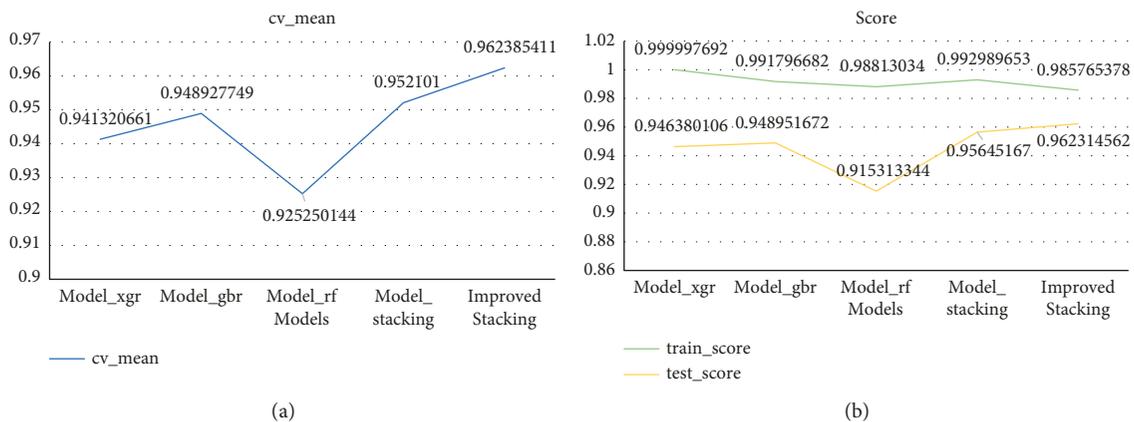


FIGURE 11: Continued.

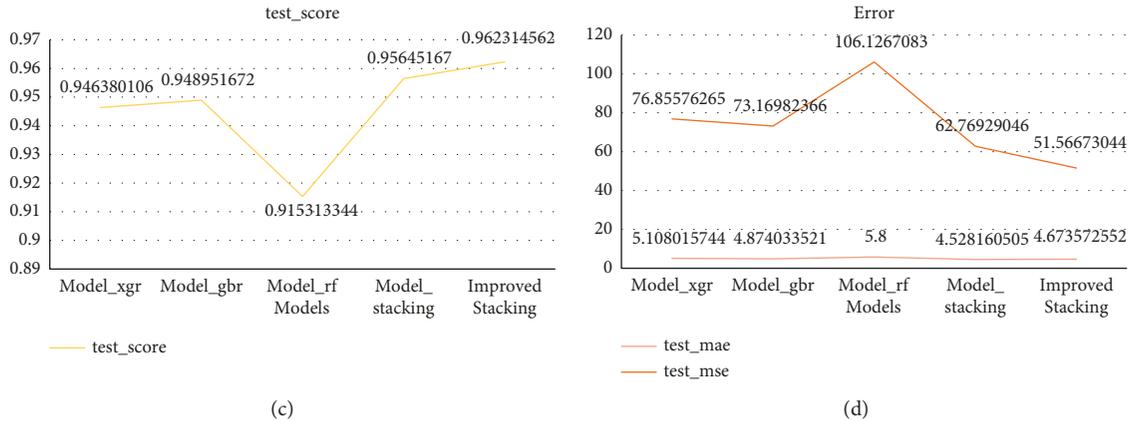


FIGURE 11: Performance of Different Models. (a) cv_mean of different models. (b) score of different models. (c) test_score of different models. (d) The error of different models. (cv_mean is the cross_val_score of models on the validation set, which is R2_Score; train_score is the cross_val_score of models on the training set, which is R2_Score; test_score returns the coefficient of determination R2 of the prediction; mean_absolute_error is the expected value of the absolute error loss; mean_squared_error is the expected value of the squared (quadratic error or loss).).

5. Conclusions

This paper attempts to predict the demand for equipment spare parts through the stacking model. By analyzing equipment operation and support scenarios, some factors that have rarely been considered by predecessors but may affect the demand for spare parts are obtained. In addition, drawing on the idea of ensemble learning, this paper proposes a new ensemble feature selection method, which can weaken the limitations of single feature selection. The experimental results show that the proposed feature selection method is better than the other three methods, and effectively improves the performance of the base models. In the past, only a few studies used ensemble learning in spare parts prediction, while fewer studies used a stacking model. In this paper, the stacking model in ensemble learning is used for spare parts demand prediction, and the problem of small number of features in the feature matrix of meta learner in the traditional stacking model is improved. The improved model added the original features as the meta learner’s input, which can improve the generalization ability and prediction accuracy. The experimental results show that the proposed model can improve the generalization ability and prediction accuracy, and performed well in the obtained data set. In the future, if the unstructured data generated during the operation and maintenance of equipment can be obtained and considered, for example, text information in maintenance history records, the data set will be expanded, and more influencing factors that have never been considered may be found.

Data Availability

The data supporting this study are available from the corresponding author through e-mail.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by National Science Foundation of China under Grants 61871013, 92167110, and 62073009, the State Key Laboratory of Rail Traffic Control and Safety (Contract No. RCS2021K004), Beijing Jiaotong University, and the National Defense Technology Basic Research Foundation.

References

- [1] Ç. Pinçe, L. Turrini, and J. Meissner, “Intermittent demand forecasting for spare parts: a critical review,” *Omega*, vol. 105, Article ID 102513, 2021.
- [2] Q. Hu, J. E. Boylan, H. Chen, and A. Labib, “OR in spare parts management: a review,” *European Journal of Operational Research*, vol. 266, no. 2, pp. 395–414, 2018.
- [3] T. Van Nguyen, L. Zhou, A. Y. L. Chong, B. Li, and X. Pu, “Predicting customer demand for remanufactured products: a data-mining approach,” *European Journal of Operational Research*, vol. 281, no. 3, pp. 543–558, 2020.
- [4] J. Huber, S. Müller, M. Fleischmann, and H. Stuckenschmidt, “A data-driven newsvendor problem: from data to decision,” *European Journal of Operational Research*, vol. 278, no. 3, pp. 904–915, 2019.
- [5] S. V. E, J. Park, and Y. Cho, “Using data mining techniques for bike sharing demand prediction in metropolitan city,” *Computer Communications*, vol. 153, pp. 353–366, 2020.
- [6] C. I. Papanagnou and O. Matthews-Amune, “Coping with demand volatility in retail pharmacies with the aid of big data exploration,” *Computers & Operations Research*, vol. 98, pp. 343–354, 2018.
- [7] S. Jiang, K.-S. Chin, L. Wang, G. Qu, and K. L. Tsui, “Modified genetic algorithm-based feature selection combined with pre-trained deep neural network for demand forecasting in outpatient department,” *Expert Systems with Applications*, vol. 82, pp. 216–230, 2017.
- [8] X. Shi, Y. D. Wong, M. Z. F. Li, C. Palanisamy, and C. Chai, “A feature learning approach based on XGBoost for driving assessment and risk prediction,” *Accident Analysis & Prevention*, vol. 129, pp. 170–179, 2019.

- [9] H. Song, C. Zhang, G. Liu, and W. Zhao, "Equipment spare parts demand forecasting model based on grey neural network," in *Proceedings of the International Conference on Quality, Reliability, Risk, Maintenance and Safety Engineering (QR2MSE)*, pp. 1274–1277, Chengdu, China, June 2012.
- [10] W. Wang and A. A. Syntetos, "Spare parts demand: linking forecasting to equipment maintenance," *Transportation Research Part E: Logistics and Transportation Review*, vol. 47, no. 6, pp. 1194–1209, 2011.
- [11] B. Choi and J. H. Suh, "Forecasting spare parts demand of military aircraft: comparisons of data mining techniques and managerial features from the case of South Korea," *Sustainability*, vol. 12, no. 15, p. 6045, 2020.
- [12] A. Bacchetti and N. Sacconi, "Spare parts classification and demand forecasting for stock control: investigating the gap between research and practice," *Omega*, vol. 40, no. 6, pp. 722–737, 2012.
- [13] A. Nouri Qarahasanlou, A. Barabadi, M. Ataei, and V. Einian, "Spare part requirement prediction under different maintenance strategies," *International Journal of Mining, Reclamation and Environment*, vol. 33, no. 3, pp. 169–182, 2019.
- [14] J. Li, K. Cheng, S. Wang et al., "Feature selection," *ACM Computing Surveys*, vol. 50, no. 6, pp. 1–45, 2018.
- [15] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: a new perspective," *Neurocomputing*, vol. 300, pp. 70–79, 2018.
- [16] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.
- [17] M. A. Hall and L. A. Smith, "Practical feature subset selection for machine learning," in *Proceedings of the 21st Australasian Computer Science Conference (ACSC 98)*, vol. 20, pp. 181–191, Perth, Australia, February 1998.
- [18] Q. Gu, Z. Li, and J. Han, "Generalized Fisher score for feature selection," in *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence*, February 2012.
- [19] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural Computing & Applications*, vol. 24, no. 1, pp. 175–186, 2014.
- [20] X. Cui, Y. Li, J. Fan, and T. Wang, "A novel filter feature selection algorithm based on relief," *Applied Intelligence*, vol. 52, no. 5, pp. 5063–5081, 2022.
- [21] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997.
- [22] L. Ma, M. Li, Y. Gao, T. Chen, X. Ma, and L. Qu, "A novel wrapper approach for feature selection in object-based image classification using polygon-based cross-validation," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 3, pp. 409–413, 2017.
- [23] A. A. Shanab, T. M. Khoshgoftaar, and R. Wald, "Evaluation of wrapper-based feature selection using hard, moderate, and easy bioinformatics data," *IEEE International Conference on Bioinformatics and Bioengineering (BIBE)*, pp. 149–155, November 2014.
- [24] S. Karasu, A. Altan, S. Bekiros, and W. Ahmad, "A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series," *Energy*, vol. 212, Article ID 118750, 2020.
- [25] K. Shima, N. Hossein, and N. Bahareh, "Multilabel feature selection: a comprehensive review and guiding experiments," *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery*, vol. 8, Article ID e1240, 2018.
- [26] G. Wei, J. Zhao, Y. Feng, A. He, and J. Yu, "A novel hybrid feature selection method based on dynamic feature importance," *Applied Soft Computing*, vol. 93, Article ID 106337, 2020.
- [27] A. Got, A. Moussaoui, and D. Zouache, "Hybrid filter-wrapper feature selection using whale optimization algorithm: a multi-objective approach," *Expert Systems with Applications*, vol. 183, Article ID 115312, 2021.
- [28] C.-F. Tsai and Y.-T. Sung, "Ensemble feature selection in high dimension, low sample size datasets: parallel and serial combination approaches," *Knowledge-Based Systems*, vol. 203, Article ID 106097, 2020.
- [29] Y. Li, T. Li, and H. Liu, "Recent advances in feature selection and its applications," *Knowledge and Information Systems*, vol. 53, no. 3, pp. 551–577, 2017.
- [30] V. Bolón-Canedo and A. Alonso-Betanzos, "Ensembles for feature selection: a review and future trends," *Information Fusion*, vol. 52, pp. 1–12, 2019.
- [31] T. Abeel, T. Helleputte, Y. Van de Peer, P. Dupont, and Y. Saeys, "Robust biomarker identification for cancer diagnosis with ensemble feature selection methods," *Bioinformatics*, vol. 26, no. 3, pp. 392–398, 2009.
- [32] N. Taheri and H. Nezamabadipour, "A hybrid feature selection method for high-dimensional data," in *Proceedings of the 2014 4th International Econference on Computer & Knowledge Engineering (ICCKE)*, pp. 141–145, Mashhad, Iran, October 2014.
- [33] B. V. Dasarathy and B. V. Sheela, "A composite classifier system design: concepts and methodology," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 708–713, 1979.
- [34] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.
- [35] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [36] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.
- [37] T. G. Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2000.
- [38] J. Kim, H. Lee, and S. Choi, "Machine learning based approach for demand forecasting anti-aircraft missiles," in *Proceedings of the 5th International Conference on Industrial Engineering and Applications (ICIEA)*, pp. 367–372, Singapore, Singapore, April 2018.
- [39] P. Matrenin, M. Safaraliev, S. Dmitriev, S. Kokin, A. Ghulomzoda, and S. Mitrofanov, "Medium-term load forecasting in isolated power systems based on ensemble machine learning models," *Energy Reports*, vol. 8, pp. 612–618, 2022.
- [40] L. Gutierrez-Espinoza, F. Abri, A. S. Namin, K. S. Jones, and D. R. W. Sears, "Ensemble learning for detecting fake reviews," in *Proceedings of the 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC2020)*, pp. 1320–1325, Madrid, Spain, July 2020.
- [41] L. Zhao, C. Mo, T. Sun, W. Huang, and X. Wan, "Aero engine gas-path fault diagnose based on multimodal deep neural networks," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 8891595, 10 pages, 2020.
- [42] L. Zhao, Z. Chen, and Y. Yang, "Parameter-free incremental co-clustering for multi-modal data in cyber-physical-social systems," *IEEE Access*, vol. 5, Article ID 21852, 2017.

- [43] L. Zhao, T. Yang, J. Zhang, Z. Chen, Y. Yang, and Z. J. Wang, "Co-learning non-negative correlated and uncorrelated features for multi-view data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 4, pp. 1486–1496, 2021.
- [44] E. Bauer and R. Kohavi, "An empirical comparison of voting classification algorithms: bagging, boosting, and variants," *Machine Learning*, vol. 36, no. 1/2, pp. 105–139, 1999.
- [45] D. W. Opitz and R. F. Maclin, "An empirical evaluation of bagging and boosting for artificial neural networks," in *Proceedings of the 1997 IEEE International Conference on Neural Networks*, pp. 1401–1405, Houston, TX, USA, June 1997.
- [46] F. Divina, A. Gilson, F. Gómez-Vela, M. García Torres, and J. Torres, "Stacking ensemble learning for short-term electricity consumption forecasting," *Energies*, vol. 11, no. 4, p. 949, 2018.
- [47] Y. Jin, X. Ye, Q. Ye, T. Wang, J. Cheng, and X. Yan, "Demand forecasting of online car-hailing with stacking ensemble learning approach and large-scale datasets," *IEEE Access*, vol. 8, Article ID 199513, 2020.
- [48] R. Al-Hajj, A. Assi, and M. M. Fouad, "Forecasting solar radiation strength using machine learning ensemble," in *Proceedings of the 7th International Conference on Renewable Energy Research and Applications (ICRERA)*, pp. 184–188, Paris, France, October 2018.
- [49] M. Liang, T. Chang, B. An et al., "A stacking ensemble learning framework for genomic prediction," *Frontiers in Genetics*, vol. 12, Article ID 600040, 2021.
- [50] M. Al-Sarem, F. Saeed, Z. G. Al-Mekhlafi, B. A. Mohammed, T. Al-Hadhrani, and M. T. A. T. S. Alshammari, "An optimized stacking ensemble model for phishing websites detection," *Electronics*, vol. 10, no. 11, p. 1285, 2021.
- [51] S. Cui, Y. Yin, D. Wang, Z. Li, and Y. Wang, "A stacking-based ensemble learning method for earthquake casualty prediction," *Applied Soft Computing*, vol. 101, Article ID 107038, 2021.
- [52] X. Hu, H. Zhang, H. Mei, D. Xiao, Y. Li, and M. Li, "Landslide susceptibility mapping using the stacking ensemble machine learning method in Lushui, southwest China," *Applied Sciences*, vol. 10, no. 11, p. 4016, 2020.
- [53] N. An, H. Ding, J. Yang, R. Au, T. F. A. Ang, and R. Au, "Deep ensemble learning for Alzheimer's disease classification," *Journal of Biomedical Informatics*, vol. 105, Article ID 103411, 2020.
- [54] R. Pari, M. Sandhya, and S. Sankar, "A multitier stacked ensemble algorithm for improving classification accuracy," *Computing in Science & Engineering*, vol. 22, no. 4, pp. 74–85, 2020.
- [55] W. Wang, D. Zhao, L. Zou, D. Wang, and W. Zheng, "Extracting 5W1H event semantic elements from Chinese online news," in *International Conference on Web-Age Information Management*, Springer Berlin Heidelberg, Berlin, Germany, 2010.