Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 6, No. 1, 2023

The Mechanical Performance Prediction of Steel Materials based on Random Forest

Shihao Wang *, Xiangxiang Wu

School of Mechanical Engineering, Tianjin University of Technology and Education, Tianjin 300222, China
* **Corresponding author**: Shihao Wang (Email: 554615076@qq.com)

Abstract: The mechanical performance of steel materials is crucial for the design, selection, and application of materials. In order to better predict the mechanical performance through chemical composition and process parameters, this paper establishes a predictive model for the mechanical properties of steel materials based on the random forest algorithm. The model predicts yield strength, tensile strength, and elongation based on chemical composition and process parameters. The results indicate that the random forest algorithm model demonstrates excellent performance in predicting the mechanical properties of steel materials.

Keywords: Yield Strength; Tensile Strength; Elongation; Random Forest Algorithm.

1. Introduction

The mechanical properties of steel materials are crucial characteristics that measure their resistance and deformation behavior. These performance indicators are essential for the design, selection, and application of materials. Key parameters among them include tensile strength, yield strength, and elongation. The mechanical properties of steel materials hold profound significance in engineering and manufacturing. These performance indicators directly impact the safety, reliability, cost-effectiveness, and sustainability of material usage. High-quality steel materials with stable mechanical properties enhance the safety and reliability of engineering structures, reduce maintenance costs, minimize resource waste, and promote environmental protection and sustainable development. Furthermore, research on mechanical properties drives the development of new materials, offering possibilities for innovation and the advancement of new fields. Therefore, a comprehensive understanding and optimization of the mechanical properties of steel materials are crucial for the modern engineering and manufacturing industries.

In the manufacturing process of steel materials, factors influencing key mechanical performance indicators can generally be categorized into two main types. Firstly, the material's chemical composition, which refers to the proportions of trace elements added apart from iron. Secondly, process parameters during production, including quenching temperature, tempering temperature, and so on. Multiple factors intertwine, complexly affecting the mechanical performance indicators mentioned above, and there exist intricate relationships of mutual constraints and influences among these performance indicators.

Traditionally, methods for seeking material mechanical properties include experimental and theoretical calculation approaches. Experimental methods have drawbacks such as being expensive, time-consuming, and sample-consuming. Theoretical calculations also suffer from limitations like high computational complexity and limited precision. In recent years, with the development of technologies such as artificial intelligence and machine learning, data-driven methods have emerged as new avenues for studying material mechanical properties. For example, machine learning algorithms are used to analyze and model large amounts of material data, predicting material performance indicators. These methods effectively combine the advantages of experiments and theoretical calculations, offering benefits such as speed, accuracy, and low cost [1-3].

Machine learning, born in the 1950s, has been widely applied in fields like computer vision, data mining, and bioinformatics. As a branch of artificial intelligence, machine learning utilizes large amounts of data to continuously optimize models, making reasonable predictions under the guidance of algorithms [4]. Machine learning can rapidly process large amounts of data and accurately identify relationships between variables [5]. Therefore, machine learning has become a trend in the research and development of new materials, leading to numerous scientific research achievements.

Mukhopadhyay et al. [6] developed a steel material mechanical performance prediction model based on artificial neural networks (ANN), selecting a network structure with a 7-19-3 topology. Experimental results showed a reliability of 90.91% for ultimate tensile strength (UTS) and 100% for elongation (EL), and the model has been applied to Tata Steel in India. Yang Wei et al. [7] employed the random forest algorithm, based on a large amount of collected real data from the hot rolling production process, obtaining importance rankings for various influencing factors on mechanical performance. They established a series of mechanical performance prediction models based on this ranking, filtering out more important influencing factors based on the trend of prediction error variations, and ultimately built a model with a small number of most crucial influencing factors as independent variables. Wang Ling et al. [8] proposed a steel material mechanical performance prediction method based on support vector regression. To avoid the blindness in selecting parameters for the support vector regression algorithm, they used a genetic algorithm to optimize the selection of parameters for this model, reducing model complexity while improving the prediction accuracy of support vector regression modeling.

2. Model Establishment

2.1. Random Forest Algorithm

Random Forest [9,10] is built on the foundation of constructing a Bagging ensemble with decision trees as base learners. Furthermore, it introduces the selection of random attributes during the training process of decision trees. A forest is created by independently establishing multiple decision trees, and when they are brought together, it forms a forest. These decision trees are built to address the same task, sharing a consistent final objective. The ultimate goal is to average their results, as illustrated in Figure 3. The regression decision tree follows the principle of minimizing the mean squared error. In other words, for any feature A and determining split point s, it aims to minimize the mean squared errors for the resulting datasets D1 and D2 individually, while also minimizing the sum of the mean squared errors for D1 and D2. The corresponding feature is identified based on this criterion. The expression is given by:

$$min_{(A,S)}[min_{c_1}\sum_{x_i\in D_1(A,S)}(y_i-c_1)^2 + min_{c_2}\sum_{x_i\in D_2(A,S)}(y_i-c_2)^2] \quad (1)$$

Where: c_1 represents the sample output mean of dataset D_1, c_2 represents the sample output mean of dataset D_2, $y_i$represents the observed value of the sample data. Random Forest is comprised of a group of tree-like decision trees, and each decision tree is considered equal during the model training process. The prediction result of a CART tree is the mean of the leaf nodes. Therefore, the prediction result of a Random Forest is the average of the predicted values from all trees.

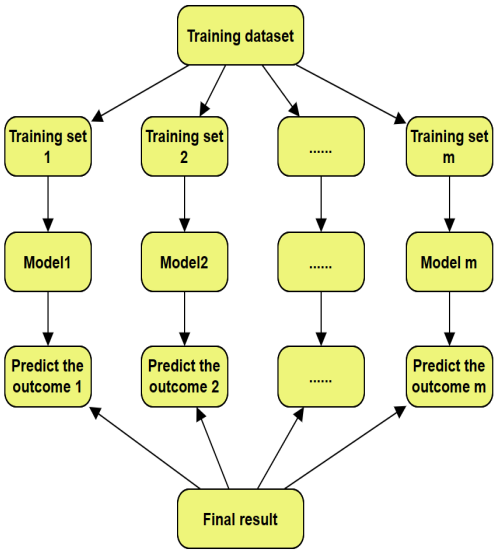

Figure 1. Diagram of the Random Forest Principle

2.2. Evaluation Metrics

Format Based on the collected data, the dataset is subjected to standardization, and subsequently, data is partitioned. Eighty percent of the data is selected as the training set, while the remaining 20% constitutes the test set. Commonly used methods for evaluating the quality of machine learning models include Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the determination coefficient R-square (R2), expressed in Equations 2, 3, and 4, respectively. Here, 'n' represents the total number of samples, 'y_i' denotes the true values, and 'y' represents the machine learning predicted values.

$$RMSE = \sqrt{\sum_{i=1}^{n}(y_i-y)^2/n} \quad (2)$$

$$MAPE = \frac{100\%}{n}\frac{\sum_{i=1}^{n}|(y_i-y)|}{|y_i|} \quad (3)$$

$$R2 = 1-\frac{\sum_i(y_i-\hat{y}_i)^2}{\sum_i(y_i-\bar{y}_i)^2} \quad (4)$$

The MAPE and RMSE values indicate the errors between predicted and true values, with smaller errors suggesting closer proximity between predicted and true values. The R2 value represents the degree of closeness between the model's predicted values and the true values. Therefore, a higher R2 value signifies more accurate predictions.

In Table 1, it can be observed that the Random Forest (RF) algorithm exhibits favorable predictive performance for yield strength, tensile strength, and elongation. The R2 performance is particularly noteworthy for yield strength, reaching an impressive 0.95, followed by 0.90 for elongation, and the least favorable result is for tensile strength at 0.87. The RMSE shows relatively larger errors for yield strength and tensile strength, attributed to the limited dataset size. Future efforts should focus on increasing the data volume for yield strength and tensile strength in steel materials. The MAPE demonstrates excellent predictive performance for all three mechanical properties, with yield strength having the most outstanding absolute percentage error at 4%, followed by 4.2% for elongation, and the least favorable being 8% for tensile strength.

Table 1. Evaluation criteria

	R2	RMSE	MAPE
Yield strength	0.95	70.34MPa	4%
Tensile strength	0.87	86.55 MPa	8%
Elongation	0.90	0.96%	4.2%

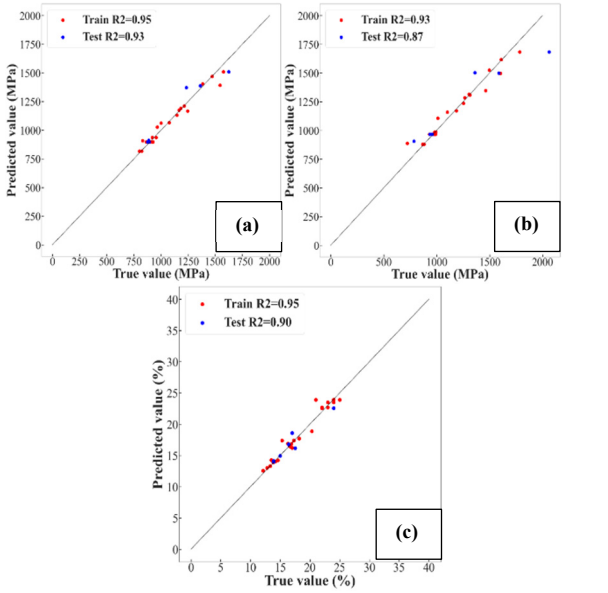

Figure 2. Comparison between the Fitting Results of the Random Forest (RF) Algorithm for Three Mechanical Properties Using the Training Set and Test Set: (a) Yield Strength; (b) Tensile Strength; (c) Elongation.

Figure 2 illustrates the fitting results of the Random Forest (RF) algorithm for the three mechanical properties, comparing the predictions from the training set and the test

set with the actual values. For yield strength, the prediction accuracy is higher in the range of 500 MPa to 1250 MPa, while it is lower in the range of 1250 MPa to 1750 MPa. Analysis indicates that the lower prediction accuracy in the range of 1250 MPa to 1750 MPa is attributed to the limited data availability for yield strength in this interval. Future efforts should focus on adding corresponding data in this range.

3. Reliability Testing of the Model

Although the model's training set and test set were randomly selected from the total dataset without overlapping data, there may be similarities between some data due to the common sources of origin. To mitigate the potential impact of this situation on the accuracy of the model testing results, two additional sets of data were chosen for model validation. The alloy compositions and processes for the new validation datasets are outlined in Table 2.

Table 2 Alloy Compositions and Processing Methods in the Validation Dataset

	1	2
C (%)	0.22	0.28
Si (%)	0.24	0.18
Mn (%)	0.48	0.56
S (%)	0.003	0.003
P (%)	0.011	0.007
Cr (%)	0.96	0.82
Mo (%)	0.69	0.51
Ni (%)	0	2.35
QT (℃)	910	880
TT (℃)	660	680
YS (MPa)	990	964
TS (MPa)	1039	1012
EL (%)	16.3	17

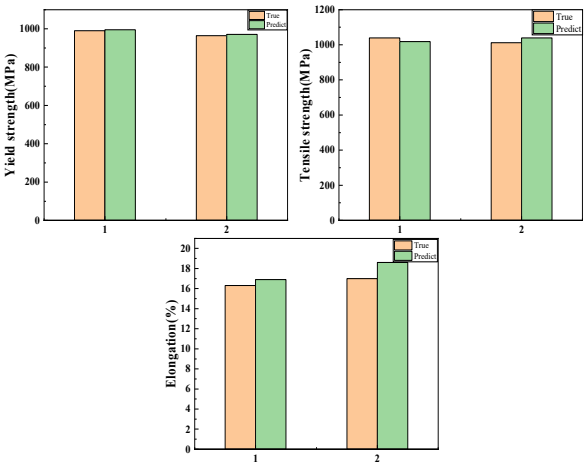

Figure 3. Actual and Predicted Values of the New Dataset Using the RF Model

Figure 3 displays the prediction results of the Random Forest (RF) algorithm model for the validation dataset. It can be observed that despite the variations in chemical composition, quenching temperature, and tempering temperature between these two sets of data, the RF algorithm model can achieve relatively accurate predictions. The average prediction errors are 6.25 MPa for yield strength, 24.07 MPa for tensile strength, and 1.09% for elongation.

4. Conclusion

Based on the Random Forest (RF) algorithm, a predictive model for the mechanical properties of steel materials, including yield strength, tensile strength, and elongation, was established. Chemical composition and processing methods were selected as features for prediction. The predictive accuracy of RF for the three mechanical properties was demonstrated using R2, RMSE, and MAPE standards. Finally, two sets of data with different chemical compositions and processing methods were chosen for validation, assessing the reliability of the RF algorithm model. The results indicate an average prediction error of 6.25 MPa for yield strength, 24.07 MPa for tensile strength, and 1.09% for elongation. The RF algorithm model proves to be reliable for predicting the mechanical properties of steel materials.

Acknowledgments

Natural Science Foundation.

References

[1] JUAN Y, DAI Y, YANG Y, et al. Accelerating materials discovery using machine learning [J]. Journal of Materials Science & Technology, 2021, 79: 178-190.W.-K. Chen, Linear Networks and Systems (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[2] WEI J, CHU X, SUN X Y, et al. Machine learning in materials science [J]. InfoMat, 2019, 1(3): 338-358.B. Smith, “An approach to graphs of linear forms (Unpublished work style),” unpublished.

[3] HOU Tengyue;SUN Yanhui;SUN Shupeng, et al. A Review of the Application of Machine Learning in Material Structure and Performance Prediction [J]. Materials Reports, 2022, 36(06): 165-176.

[4] WEI J, CHU X, SUN X Y, et al. Machine learning in materials science [J]. InfoMat, 2019, 1(3): 338-358.C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.

[5] LIU Y, SUN J-B, LIU S-J, et al. Optimization of Ultra-High and High Manganese Steel Based on Artificial Neural Network and Genetic Algorithm [J]. Journal of Materials Engineering and Performance, 2023: 1-11.M. Young, The Techincal Writers Handbook. Mill Valley, CA: University Science, 1989.

[6] MUKHOPADHYAY A, IQBAL A. Prediction of mechanical properties of hot rolled, low-carbon steel strips using artificial neural network [J]. Materials and Manufacturing Processes, 2005, 20(5): 793-812.

[7] YANG Wei, LI Wei-gang, ZHAO Yun-tao, et al. Mechanical property prediction of steel and influence factors selection based on random forests [J]. Iron & Steel, 2018, 53(03): 44-49.

[8] WANG Ling, FU Dongmei, MU Zhichun. GA-based on SVR Method in Prediction of Mechanical Property of Steel Materials [J]. Journal of System Simulation, 2009, 21(04): 1192-1194.

[9] BREIMAN L. Random forests [J]. Machine learning, 2001, 45: 5-32.