https://doi.org/10.1007/s41230-024-3090-1

Data-driven casting defect prediction model for sand casting based on random forest classification algorithm

Bang Guan[1], *Dong-hong Wang[1,2], **Da Shu[1,2], Shou-qin Zhu[3], Xiao-yuan Ji[4], and Bao-de Sun[1,2]

1. Shanghai Key Lab of Advanced High-Temperature Materials and Precision Forming, School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

2. State Key Lab of Metal Matrix Composites, Shanghai Jiao Tong University, Shanghai 200240, China

3. Hefei Casting and Forging Factory of Anhui Heli Co., Ltd, Hefei 230000, China

4. State Key Laboratory of Materials Processing and Die & Mould Technology, Huazhong University of Science and Technology, Wuhan 430074, China

Abstract: The complex sand-casting process combined with the interactions between process parameters makes it difficult to control the casting quality, resulting in a high scrap rate. A strategy based on a data-driven model was proposed to reduce casting defects and improve production efficiency, which includes the random forest (RF) classification model, the feature importance analysis, and the process parameters optimization with Monte Carlo simulation. The collected data includes four types of defects and corresponding process parameters were used to construct the RF model. Classification results show a recall rate above 90% for all categories. The Gini Index was used to assess the importance of the process parameters in the formation of various defects in the RF model. Finally, the classification model was applied to different production conditions for quality prediction. In the case of process parameters optimization for gas porosity defects, this model serves as an experimental process in the Monte Carlo method to estimate a better temperature distribution. The prediction model, when applied to the factory, greatly improved the efficiency of defect detection. Results show that the scrap rate decreased from 10.16% to 6.68%.

Keywords: sand casting process; data-driven method; classification model; quality prediction; feature importance

CLC numbers: TP391.9 **Document code:** A **Article ID:** 1672-6421(2024)02-137-10

1 Introduction

Foundry industries need to produce quality castings with a minimum number of rejections to meet the market demand, however, casting defects caused by a combination of many factors are inevitable due to the involvement of the number of process parameters in the casting process [1]. To reduce defects in sand casting, it is necessary to understand the process parameters as well as their effects on the final castings. The trial and error method is usually used in foundries, which causes amounts of failure in maintaining a satisfactory quality control level [2]. Smart manufacturing that integrates highly digitized production facilities, including the Internet of Things, Cyber Physics Systems, Big Data, Machine Learning, etc., makes the production process more intelligent and efficient [3]. Using various quality control tools to control process variability can effectively solve quality problems in the manufacturing industry. Combining casting knowledge and big data models to analyze, predict, and improve casting quality shows great potential in reducing scrap rate [4].

An efficient defect prediction model requires a detailed consideration of the production process. It is confirmed that some typical defects are related to the process parameters in sand casting. In cast iron foundries, the causes of cold shut include low pouring temperature, low fluidity, very hard mould, and high

***Dong-hong Wang**

Male. His research interests mainly focus on material genetic engineering and intelligent thermal manufacturing.

E-mail: wangdh2009@sjtu.edu.cn

****Da Shu**

E-mail: dshu@sjtu.edu.cn

Received: 2023-07-21; Accepted: 2024-01-16

moisture; the shrinkage porosity defect is related to incorrect feeding, unsuitable carbon equivalent, and high pouring temperature; the sand hole defect is related to the weak mould surface, loose sand, and handling of mould after closing friable sand; the gas porosity defect is related to incorrect feeding, weak mould, and high phosphorus content [1, 5-7]. It is difficult to analyze the specific impact on the actual production castings due to complex multi-parametric relationships. Data mining is a novel strategy to determine the significant process parameters, which contains defect data set preparation, data-driven model building, and feature importance analysis [8].

Data-driven techniques have been widely used in manufacturing industries, such as industrial process monitoring [9], process planning [10-12], and decision-making for production systems [13]. In research related to the manufacturing industry, the majority of data-driven models utilize a supervised learning method to construct regression or classification models [14]. Babu et al. [15] constructed Naive Bayes and Support Vector Machine (SVM) classifier models based on a simpler tabular data input consisting of morphological and mean gray values of inclusions to distinguish the types of nonmetallic inclusions. Boto et al. [16] presented three data-driven models based on process data to estimate different indicators related to process performance in a steel production process and developed a new approach with feature selection methods and four state-of-the-art regression approximations (random forest, gradient boosting, xgboost and neural networks). Zhao et al. [17] took die-casting pressure, aluminum liquid components, and 25 other parameters as the prediction basis (input) and the geometrical dimensions as the goal of quality prediction to train the historical data in the die-casting smart factory using the BP neural network. Liu et al. [18] proposed a machine learning method to realize the real-time quality prediction in the die-casting process and the appearance defect quality prediction after processing, respectively. Another widely used data-driven method is to collect process parameters for product defect classification and prediction in the die-casting production process [19, 20]. The production data from cast steel and iron foundries can be used to create data-driven models for predicting casting surface-related defects and the model was successfully applied to the prediction of surface defects and the understanding of relative parameters [8]. However, this study collected data from many types of castings, the correlation between defects and process parameters is different for various casting models. Therefore, collecting effective data on specific castings and defect types are beneficial to construct models with high accuracy. Method of combining Monte Carlo (MC) models and data-driven models have been proposed to optimize process parameters [21, 22], which randomly generated a data pool following either a normal or uniform distribution, and then the corresponding results are calculated according to the mathematical model, so as to select the optimal preset parameter distribution.

This study presents an approach for the application of machine learning in the prediction and understanding of casting related defects. The Steering Bridge casting data were collected for the creation of a classification model in a foundry for half a year. The model analysis provided valuable insights into how process parameters affect defects. Furthermore, the application of Monte Carlo simulation was demonstrated effectively for defect prediction in different stages of the casting process, including the calculation of defect occurrence probability.

2 Methods

A novel strategy for a quality prediction model and process optimization was constructed, and the brief flow chart for the overall methodology is shown in Fig. 1. This framework is mainly divided into four parts: data acquisition, pre-processing, data classification model construction, quality prediction and process optimization. The specific methods for each of these four parts are introduced in the following sections.

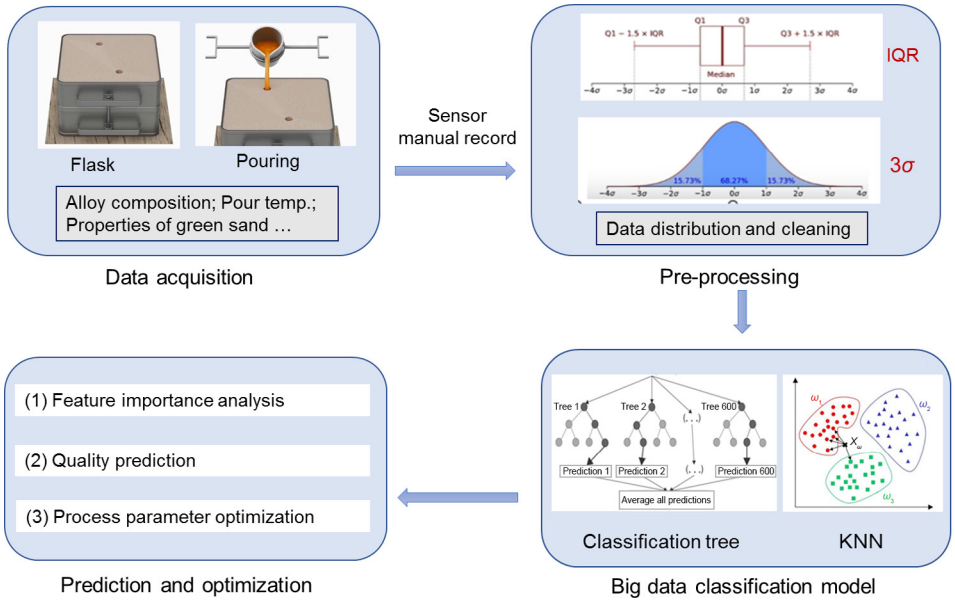

Fig. 1: Overview of the methodology for the quality prediction model and process optimization

2.1 Data acquisition

The casting data has been collected in a foundry for half a year. The foundry has implemented a casting ERP system incorporating a Single-Piece management method to achieve the recording of process parameter data and the requirement on product single-piece traceability [23]. The steering bridge is an important part of the forklift, as shown in Fig. 2(a). The casting overall dimensions are 840 mm×224 mm×370 mm, with a critical position thickness of 27 mm. The maximum thickness of the casting is 60 mm, while the minimum thickness is 12 mm. The weight of the casting is 51 kg. The metallurgical defect is inevitable due to process fluctuations and complex casting structures, such as the sand hole shown in Fig. 2(b). Considering the importance and accessibility of parameters in the manufacturing process of sand mould and casting, 18 process parameters listed in Table 1 were collected through manual records, sensors, and analysis equipment, and the corresponding casting quality was detected. According to the quality detection results, there are five types of casting defects: gas porosity, sand hole, cold shut, shrinkage porosity, and good sample. The dataset is constructed containing 6,382 sample data collected by the factory, including 400 samples of gas porosity, 359 samples of sand hole, 273 samples of cold shut, 148 samples of shrinkage porosity, and 5,202 good samples.

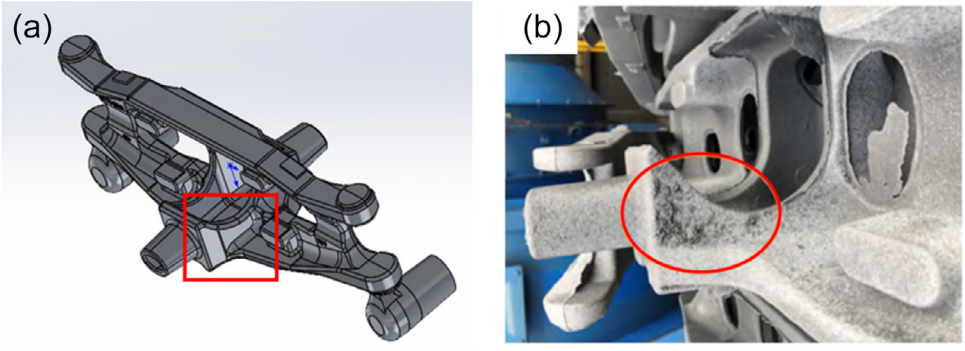

Fig. 2: CAD model (a) and steering bridge casting (b) showing the position prone to surface defects in red circle

Table 1: Statistics of sand casting process parameters

Feature	Process parameter	Max.	Min.	Mean	Std. deviation
1	Pouring temperature (°C)	1,415	1,385.000	1,401.413	6.220
2	C (wt.%)	3.85	3.610	3.761	0.056
3	Si (wt.%)	2.92	2.600	2.710	0.053
4	Mn (wt.%)	0.66	0.380	0.516	0.050
5	P (wt.%)	0.047	0.013	0.027	0.005
6	S (wt.%)	0.018	0.006	0.012	0.002
7	Mg (wt.%)	0.057	0.034	0.045	0.005
8	Al (wt.%)	0.04	0.017	0.025	0.004
9	Pouring weight (kg)	145	128.000	134.929	2.157
10	Pouring time (s)	27.2	11.900	17.293	2.006
11	Inoculation amount (g)	92	24.000	49.562	9.694
12	Moulding sand compactability (%)	48.82	35.070	39.816	1.277
13	Moulding sand shear strength (kPa)	60	2.000	5.033	5.760
14	Used sand temperature (°C)	48.8	33.400	41.268	2.732
15	Moisture of used sand (%)	2.94	1.380	1.993	0.195
16	Bentonite (%)	58.5	12.500	23.006	1.891
17	Mixed clay (%)	13.9	9.800	11.851	0.620
18	New sand (%)	40	0.000	10.641	12.488

2.2 Pre-processing

Pre-processing involves data cleaning, data verifying, and data formatting into a usable dataset, which helps build a machine learning model more accurately. According to the process parameters given in Table 1, the sample elements are 18-dimensional feature vectors.

Abnormalities in the collected data include abnormal sensor records, manual miss-recording, and missing data. Box-plot and normal distribution graphs are commonly used to deal with data abnormalities. The elimination of the 3σ rule requires the original data distribution to be close to the normal distribution. Furthermore, a Quantile-Quantile plot is used to judge whether the sample distribution is suitable [24]. The bottom and top points of the box plot are the 25th and 75th percentile. The process parameters are considered to be an input of feature vectors. The abnormal data is concentrated on Feature 8, Feature 12, Feature 13, and Feature 17, therefore, the corresponding 36, 52, 12, and 8 groups of data are reduced respectively. The Quantile-Quantile plots of those features are shown in Fig. 3, some sample values are deviated from the normal range [Fig. 3(a) and Fig. 3(d)], and some samples have obvious loss or error [Fig. 3(b) and Fig. 3(c)]. Finally, 6,276 sample data are used for model creation.

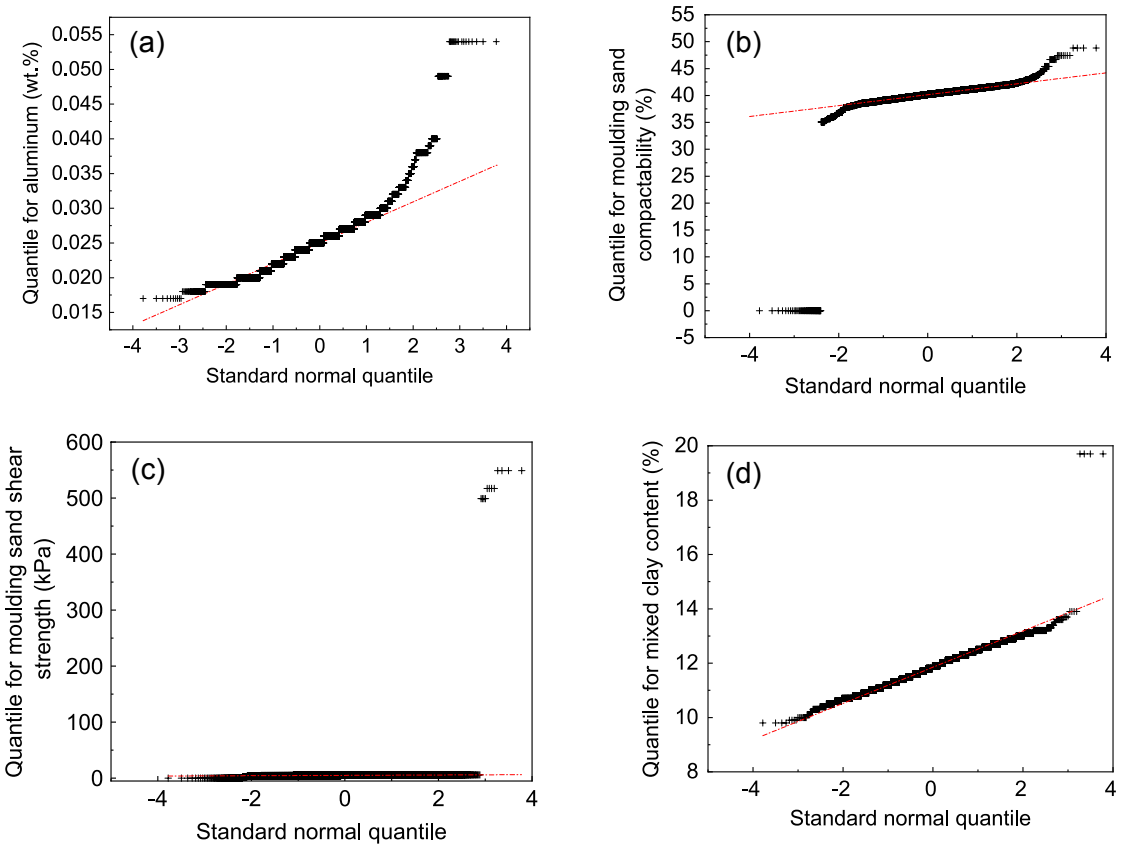

Fig. 3: Quantile-Quantile plot of features: (a) aluminum content; (b) moulding sand compactability; (c) moulding sand shear strength; (d) mixed clay content

Learning algorithms benefit from the normalization of datasets, but the collected data has different ranges, which makes it difficult for many machine learning models to reach the optimal computational state. Z-score normalization is used to scale the data so that it falls into a specific interval, which obeys normal distribution with a mean of 0 and a standard deviation of 1 [25].

2.3 Classification model

Machine learning methods of Random Forest (RF), K-nearest neighbor (KNN), Support Vector Machine (SVM), and Neural Networks (NNs) were used to construct classification models. As an integrated learning algorithm, the accuracy rate of a random forest is higher than that of the decision tree, mainly for a large number of table data. Neural network can achieve a higher accuracy than that of random forest when the amount of data is large enough, but its interpretability is relatively low, and it is difficult to configure appropriate hyper-parameters. Since the casting defect data belongs to one-dimensional tabular data, it is better to choose the random forest algorithm for integrated learning.

The construction process of a random forest is shown in Fig. 4, which is mainly divided into three key steps. Firstly, the training sample set was built based on the bagging method, and then feature attributes were selected based on the randomness of feature subsets to construct a decision tree, finally, a majority voting method was used to make decisions on the results of each decision tree.

(1) Building training sample set

To construct the training sample set, assuming that the original dataset contains N samples, the elements in the sample set are L-dimensional attribute vectors. According to the

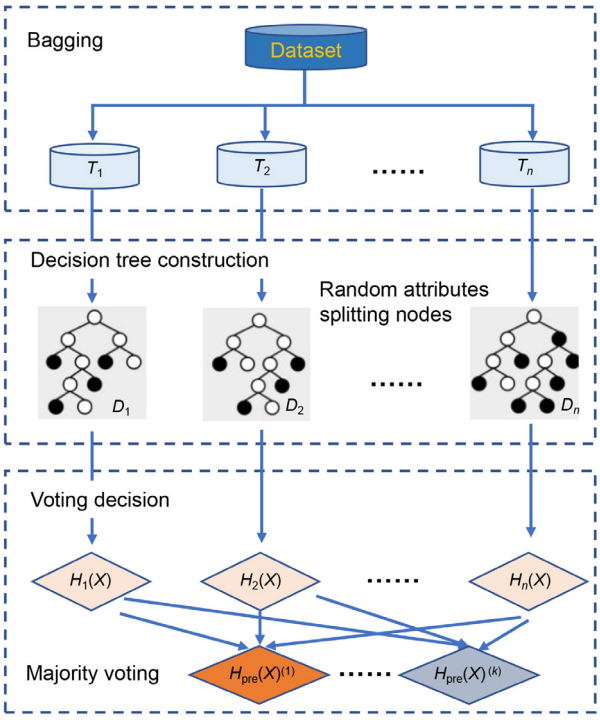

Fig. 4: Construction of random forest classification model

process parameters given in Table 1, the sample elements are 18-dimensional feature vectors. The bagging method was used for sampling, and after N times of sampling, a new training set containing N samples was obtained. Then, this process was repeated K times, and the training sample set of K*N samples can be obtained. Notably, the process of sampling is an independent event, each sample has the same probability of being selected at each sampling instance. Therefore, the newly constructed sample set may have duplicate samples, and some samples may not be selected during every iteration. Initially, the weight of each sample was set to the same value to test the model, then, the prediction accuracy of the model for each sample was obtained by increasing the weight of samples with low prediction accuracy and decreasing the weight of samples with high prediction accuracy. Finally, an optimal set of weights can be obtained to maximize the prediction accuracy by continuously selecting the weight of the samples.

(2) Construct the decision tree

The decision tree was constructed for each training sample set. When nodes were split in each decision tree, some attributes were randomly selected from the 18 feature attributes to form an attribute subset for node splitting. The purpose of decision tree node splitting is to find the path with the maximum drop rate and quickly filter and classify sample data. Purity describes the consistency of data categories in a node, the greater the purity, the more similar the data filtered through the decision path. This indicates that the attributes selected for node splitting in the decision tree are appropriate. In contrast to purity, impurity describes the inconsistency of data categories in a node. The Gini impurity was used as the evaluation standard, it can be calculated as:

$$\mathrm{Gini}(D)=\sum_{k=1}^{|y|}\sum_{k'\neq k}p_k p_{k'}=1-\sum_{k=1}^{|y|}p_k^2 \quad (1)$$

where D is the sample set, p_k is the proportion of samples in category k, and y is the total number of categories. Gini Index (GI) is the best principle to judge the classification quality in the random forest [26], the lower GI value implies the higher quality of classification based on the optimal attribute.

(3) Majority voting for the result

After training, each decision tree predicts the classification category. As shown in Fig. 4, there are N decision trees with K decision results for one sample, and the decision-making results are not the same. To obtain the final decision-making result of the random forest, the majority voting method was used to select the prediction result. The principle of the majority voting method is that when a prediction label obtains more than half of the decision tree votes, the final prediction is the prediction label, otherwise, the prediction label is rejected.

2.4 Quality prediction and process parameters optimization

The application of the quality prediction model is shown in Fig. 5, which consists of 3 cases: (1) Determine the prediction results of quality classification after all process parameters were determined. (2) During the casting process, if some parameters were not determined, the pending parameters should be assigned random values following the historical dataset. (3) Set the undetermined parameters as random values from a new distribution, and then changed the distribution of input parameters, to select the optimal preset parameter distribution.

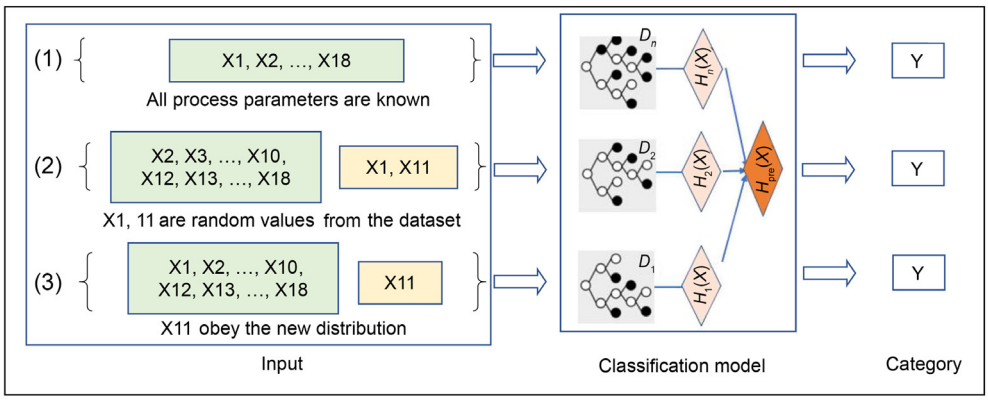

Fig. 5: Quality prediction in different cases

For Case 2 and Case 3, the Monte Carlo simulation method was used to calculate the input parameters which selected random values from the statistical distribution of undetermined parameters and then calculate the classification results using the trained random forest model. In the casting process, features 1 to 18 are replaced by X1 to X18. Considering the pouring temperature and inoculation amount are the final parameters, which is mapped to the inputs X1 and X11, the classification results can be affected by changing these two values, especially when them are of high GI value. In case 2, directly setting pending parameters as random values following the historical dataset ensures they obey the real distribution.

The optimization process is to determine the optimal distribution according to the results of statistical analysis. Table 1 shows the temperature difference of pouring temperatures is 30 K, assuming a normal distribution (μ, σ^2), in which the variance σ^2 is set to a fixed value of 5 and μ varies from the minimum to the maximum. It should be noted that as a reference for this case, the actual variance is difficult to obtain. Then, the optimal distribution is selected according to the predicted result.

3 Results and discussion

3.1 Classification result

When training the classification model, 5-fold cross-validation was used, and the classification performance was preliminarily evaluated with two evaluation indexes: (1) Accuracy: Measure the percentage of all samples that are correctly classified; (2) F_1 score: Harmonic mean of precision rate and recall rate. The recall and precision of defective castings are the main performance indicators in the industry [27]. Recall refers to the ratio between the amount of relevant information detected from the database and the total amount. Precision is the ratio of the number of relevant samples in the recognition or retrieval results to the total number of samples in the results, which can be expressed as:

$$TPR = TP/(TP+FN) \quad (2)$$

$$PPV = TP/(TP+FP) \quad (3)$$

where TPR is the recall, and PPV is the accuracy. False positive (FP) is the number of negative samples incorrectly predicted as positive samples. False negative (FN) is the number of positive samples incorrectly predicted as negative samples. True positive (TP) is the number of positive samples correctly predicted.

Table 2 shows the classification performance of selected models. In the RF model, the average recall for classification is 94.08 and the average precision is 90.72, indicating this model has great classification performance. Compared with three other common classification algorithms, the RF model is the best.

The accuracy of this model is 97.1%, but the model with high accuracy may not have a good classification effect, it is necessary to calculate and analyze the observed values of the confusion matrix. Figure 6 shows the confusion matrix of random forest multi-classification results. Categories 0, 1, 2, 3, and 4 correspond to good, gas porosity, sand hole, cold shut, and shrinkage porosity samples, respectively.

Table 2: Performance comparison of classification models

Process parameter	Accuracy	F_1 score	Recall (%)	Precision (%)
RF	97.1	92.37	94.08	90.72
KNN	95.8	90.25	92.52	88.08
SVM	92.3	77.51	74.4	80.90
NNs	94.3	86.84	91.44	82.68

Figure 7(a) shows the recall of the defect is more than 90%, indicating that more than 90% of the defects have been correctly detected. Most samples with the wrong classification of defects are considered to be zero defects. This indicates that there are fewer cases where multiple defects occur simultaneously on the same casting. According to the data analysis, there is a certain error between category 1 (gas porosity) and category 3 (cold shut), which indicates that the process parameters of porosity and cold shut are more similar than other defects. The accuracy shown in Fig. 7(b) is lower than the recall rate. The training process has a great impact on the result. When the cost of misclassifying defect samples as good samples is preset to be higher, the recall rate is increased and the accuracy rate is reduced. On the contrary, if this error cost is set lower, the recall rate decreases and the accuracy rate increases. In this study, the recall rate is considered a more important indicator, the cost function is set reasonably.

3.2 Feature importance

To determine the impact of each specific process parameter on specific defects, the GI value of the classification models of the four types of defects are calculated, and the results are shown in Fig. 8. The influence of inoculation amount and pouring temperature (11, 1) on gas porosity is more important, because that high pouring temperature and excess inoculation amount are easy to cause porosity defects. This is consistent with the

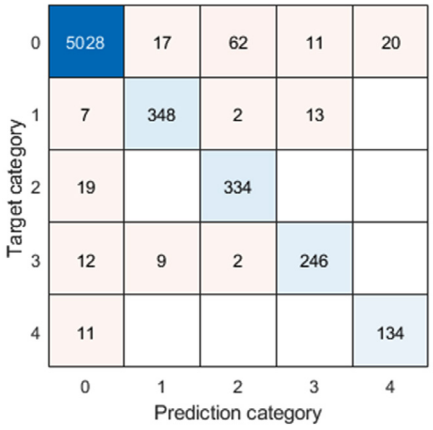

Fig. 6: The confusion matrix of random forest multi-classification results

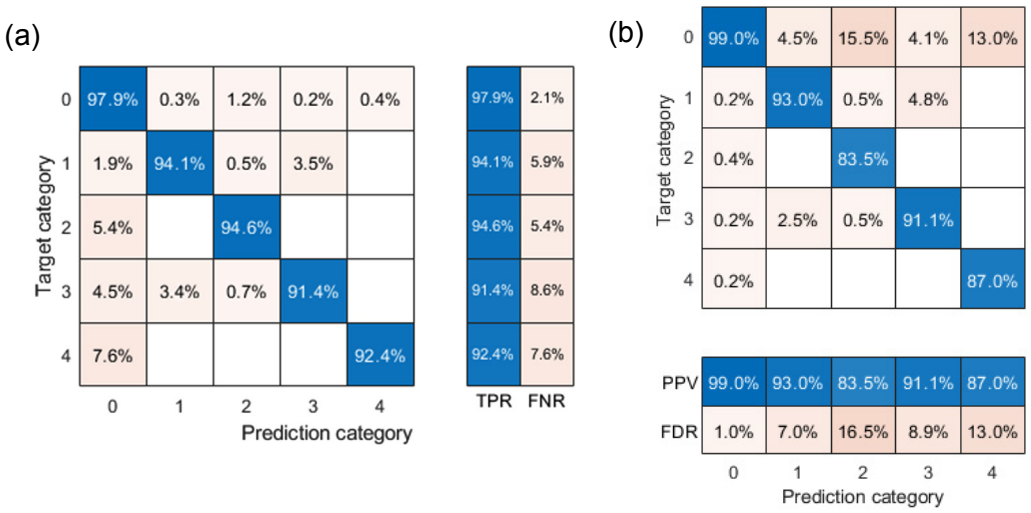

Fig. 7: Results of classification: (a) TPR (recall) of the confusion matrix; (b) PPV (accuracy) of the confusion matrix

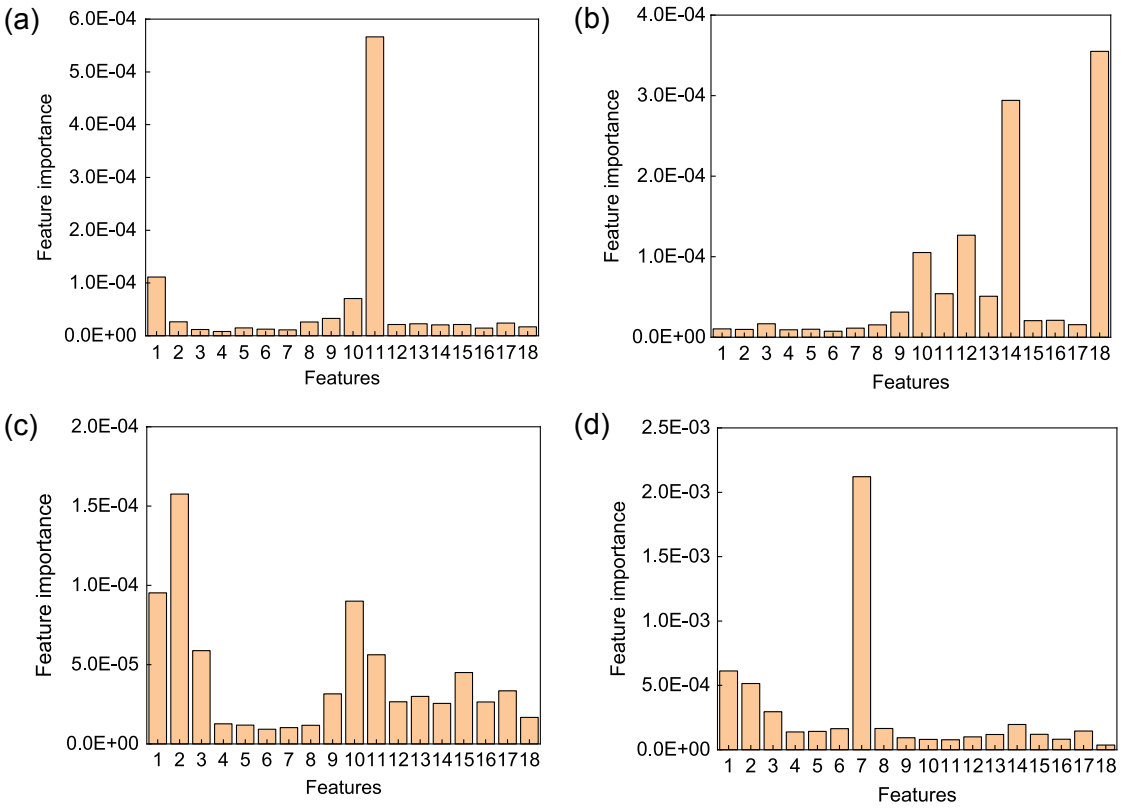

Fig. 8: Feature importance of each defect in classification model: (a) gas porosity; (b) sand hole; (c) cold shut; (d) shrinkage porosity

features importance analysis results. It is recommended to adjust inoculation amount and reduce temperature wherever possible. For sand hole defects, the important features are new sand contents, used sand temperature, moulding sand compactability, and pouring time (18, 14, 12, 10) in sequence. For cold shut defect, the important features are carbon content, pouring temperature, pouring time, silicon content, and inoculation amount (2, 1, 10, 3, 11) in sequence. There are many factors with similar contributions that influence cold shut. This indicates that the process parameters collected in this study cannot be well distinguished from cold shut defects. For shrinkage porosity defects, the most important feature is magnesium content. In order to obtain a better nodulizing quality, the residual Mg content is generally higher than 0.03%, but it increases the shrinkage tendency of casting. Taking into account the difficulty of process control, it is generally controlled at 0.035%–0.04% in actual production for spheroidal graphite cast iron [28].

As for the result of data mining, the unexpected results are more worth considering for the foundries instead of the reasonable ones. In addition, the possible process of defect occurrence can be inferred through feature importance analysis to improve detection efficiency.

3.3 Quality prediction

The product quality prediction model can effectively control product quality with real-time data interaction. According to the above method, the predicted process of Cases 1–3 was discussed in detail with examples.

(1) Completion of the casting process

The random forest classifier trained above was used for classification. The input was the vector of features 1-18, and the output result was the predicted classification category. The score of each category for the corresponding output was also calculated, which was obtained by voting from the post-verification probability of each weak classifier. Table 3 shows the classification scores of some samples. The category with the highest score is the predicted output, and the higher the score, the higher the reliability. When the scores are close, there is a high likelihood of misjudgment in the classification result.

(2) Process in progress

In the casting process, the pouring temperature and inoculation amount are the final process parameters, which are set as random values following the distribution of the historical dataset, to form a complete input feature vector. The 6,276 samples with the same number of samples as the dataset are constructed, as shown in Table 4, where 16 features of them are fixed values.

Statistically averaging the predicted output class scores for each sample, the probability of predicting each type of defect is 0.9267 (good), 0.0487 (gas porosity), 0.0019 (sand hole), 0.0201 (cold shut), and 0.0025 (shrinkage porosity). The result shows there is a 92.67% probability to obtain good products by selecting random values in Feature 1 and Feature 11, indicating that the previous 16 process parameters are acceptable. Next, fixed the inoculation amount value with a larger value of 70, the probability of predicting each type of defect is 0.1003 (good), 0.8479 (gas porosity), 0.0399 (sand hole), 0.0104 (cold shut), and 0.0016 (shrinkage porosity). Obviously, the Feature 11 greatly affects the result of gas porosity and the value of 70 is inappropriate. The same conclusion can also be obtained from feature importance analysis. This predictive method can make real-time decisions for the product production process and reduce the scrap rate.

(3) Process parameter optimization

The optimization goal is to obtain the appropriate pouring temperature when all other parameters are determined. The input feature vector is shown in Table 4, and the inoculation

Table 3: Some samples output scores of prediction category

Sample	Good	Gas porosity	Sand hole	Cold shut	Shrinkage porosity
1	0	0.9500	0	0.0167	0.0333
2	0	0	0	1	0
3	0	0.0333	0	0.9667	0
4	0	0.0033	0	0.9667	0
5	0.0333	0.0367	0	0.9300	0
6	0	0.0367	0	0.9633	0
7	0	0.7033	0	0.2967	0
8	0	0.8685	0	0.1315	0
9	0	0.6843	0	0.3157	0
10	0	0.1643	0	0.8357	0

Table 4: Values of input parameter

Feature	Process parameter	Value	Feature	Process parameter	Value
1	Pouring temperature (°C)	Random	10	Pouring time (s)	15.9
2	C (wt.%)	3.7	11	Inoculation amount (g)	Random
3	Si (wt.%)	2.73	12	Mould sand compactability (%)	41.51
4	Mn (wt.%)	0.52	13	Mould Sand shear strength (kPa)	5.56
5	P (wt.%)	0.014	14	Used sand temperature (°C)	40.2
6	S (wt.%)	0.007	15	Moisture of used sand (%)	1.92
7	Mg (wt.%)	0.035	16	Bentonite (%)	24.1
8	Al (wt.%)	0.026	17	Mixed clay (%)	11.7
9	Pouring weight (kg)	132	18	New sand (%)	20

amount value is set to a larger value of 70. The pouring temperature was preset as 30 normal distributions, and the Monte Carlo simulation method was used to simulate 50,000 times within each temperature distribution interval to calculate the defect probability for each temperature segment. The curve of the probability of various defects within the temperature range is plotted in Fig. 9. It can be seen that the gas porosity defect rate has a maximum value at 1,403 °C. However, at low temperatures, cold shut defects are prone to occur. Considering the defect rate comprehensively, a higher pouring temperature is recommended in this case.

The above functionality can be integrated into a quality prediction and decision-making system, as shown in Fig. 10, which can perform real-time quality classification of ongoing casting processes. This application software is deployed on PCs on the site, providing computational services to quality data management personnel. The results show that the scrap rate has dropped from 10.16% to 6.68%.

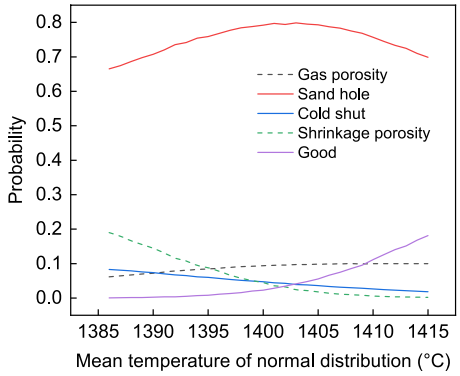

Fig. 9: Curve of the probability of various defects with the temperature range

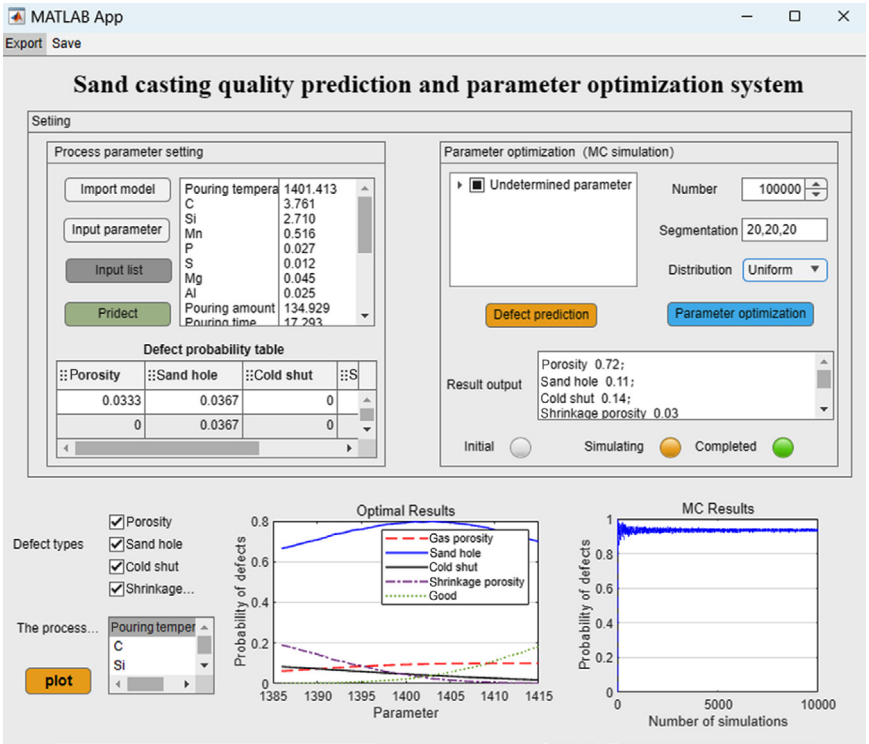

Fig. 10: Quality prediction and decision-making system

4 Conclusion

This study collected the data on relevant process parameters in the sand casting process and established a data-driven casting defect classification model. The recall rate of all the defects is more than 90%. In this model, the influence of process parameters on gas porosity and cold shut defect is similar according to classification error result. Based on the classification model, the importance of defect-influencing factors was analyzed. The influence of pouring temperature and inoculation amount on gas porosity is more important, high pouring temperature and excess inoculation amount are easy to cause gas porosity defects. For sand hole defects, the important parameters are new sand contents, used sand temperature, moulding sand compactability, and pouring time in sequence. For shrinkage porosity defect, the most important feature is magnesium content.

Quality prediction model applied in three situations was implemented. In the first case, defects are predicted when all processes were completed, and the score of each category for the corresponding output was also calculated. In the second case, the prediction method can determine whether the preset parameters are appropriate in the process. Finally, the classification model is used as the mathematical model of Monte Carlo simulation to determine the optimal process parameters. In this case, the gas porosity defect rate has a maximum value at 1,403 °C and a higher pouring temperature is recommended. This data-driven method for quality prediction and process optimization shows great potential in reducing scrap rates.

Acknowledgments

This work was financially supported by the National Key Research and Development Program of China (2022YFB3706800, 2020YFB1710100), and the National Natural Science Foundation of China (51821001, 52090042, 52074183).

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Sertucha J, Lacaze J. Casting defects in sand-mold cast irons–An illustrated review with emphasis on spheroidal graphite cast irons. Metals, 2022, 12(3): 504–584.
[2] Giannetti C, Ransing R S, Ransing M R, et al. Knowledge management and knowledge discovery for process improvement and sustainable manufacturing: A foundry case study. Proceedings of the Sustainable Design and Manufacturing, 2014: 537–548.
[3] Tao F, Qi Q, Liu A, et al. Data-driven smart manufacturing. Journal of Manufacturing Systems, 2018, 48: 157–169.
[4] Antony J, Mcdermott O, Sony M. Revisiting Ishikawa's original seven basic tools of quality control: A global study and some new insights. IEEE Transactions on Engineering Management, 2021, 99: 1–16.
[5] Sai T V, Vinod T, Sowmya G. A critical review on casting types and defects. Engineering and Technology, 2017, 3(2): 463–468.
[6] Natarajan N K. Review analysis of casting defects with respect to Indian standards in cast iron foundry. Journal of Chemical and Pharmaceutical Sciences, 2016, 2: 63–68.
[7] Suthar J, Persis J, Gupta R. Predictive modeling of quality characteristics – A case study with the casting industry. Computers in Industry, 2023, 146: 1–16.
[8] Chen S, Kaufmann T. Development of data-driven machine learning models for the prediction of casting surface defects. Metals, 2021, 12(1): 1–15.
[9] Zhang Y, Zhang R, Wang Y, et al. Big data driven decision-making for batch-based production systems. Procedia CIRP, 2019, 83: 814–818.
[10] Lundgren M, Hedlind M, Kjellberg T. Model-driven process planning and quality assurance. Procedia CIRP, 2015, 33: 209–214
[11] Yin S, Ding S X, Xie X, et al. A review on basic data-driven approaches for industrial process monitoring. IEEE Transactions on Industrial Electronics, 2014, 61(11): 6418–6428.
[12] Ktari A, El Mansori M. Intelligent approach based on FEM simulations and soft computing techniques for filling system design optimisation in sand casting processes. The International Journal of Advanced Manufacturing Technology, 2021, 114(3–4): 981–995.
[13] Ge Z. Review on data-driven modeling and monitoring for plant-wide industrial processes. Chemometrics and Intelligent Laboratory Systems, 2017, 171: 16–25.
[14] Tao F, Cheng J, Qi Q, et al. Digital twin-driven product design, manufacturing and service with big data. The International Journal of Advanced Manufacturing Technology, 2017, 94(9–12): 3563–3576.
[15] Babu S R, Musi R, Thiele K, et al. Classification of nonmetallic inclusions in steel by data-driven machine learning methods. Steel Research International, 2022, 94(1): 2200617.
[16] Boto F, Murua M, Gutierrez T, et al. Data driven performance prediction in steel making. Metals, 2022, 12(2): 172–191.
[17] Zhao Y, Qian F, Gao Y. Data driven die casting smart factory solution. Recent Advances in Intelligent Manufacturing: First International Conference on Intelligent Manufacturing and Internet of Things and 5th International Conference on Computing for Sustainable Energy and Environment, IMIOT and ICSEE 2018, Chongqing, China, 2018, 923: 13–21.
[18] Liu D, Du Y, Chai W, et al. Digital twin and data-driven quality prediction of complex die-casting manufacturing. IEEE Transactions on Industrial Informatics, 2022, 18(11): 8119–8128.
[19] Bak C, Roy A G, Son H. Quality prediction for aluminum diecasting process based on shallow neural network and data feature selection technique. CIRP Journal of Manufacturing Science and Technology, 2021, 33: 327–338.
[20] Chakrabarti A, Sukumar R P, Jarke M, et al. Efficient modeling of digital shadows for production processes: A case study for quality prediction in high pressure die casting processes. In: Proc. 2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA), Porto, Portugal, 2021: 1–9.
[21] Fang Y, Ma L, Yao Z, et al. Process optimization of biomass gasification with a Monte Carlo approach and random forest algorithm. Energy Conversion and Management, 2022, 264: 115734.
[22] Kozlovsky V N, Lysov V E, Antipov D V, et al. Calculation and statistical experiment on the Monte Carlo method when assessing the stability of the technical characteristics of the automobile generator set in mass production. In: Proceedings of the 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), Saint Petersburg and Moscow, Russia, 2019: 565–568.
[23] Zhou J, Ji X, Liao D, et al. Research and application of enterprise resource planning system for foundry enterprises. China Foundry, 2013, 10(1): 7–17.
[24] Lee D K. Data transformation: A focus on the interpretation. KJA, 2020, 73(6): 503–8.
[25] Singh D. Investigating the impact of data normalization on classification performance. Applied Soft Computing, 2020, 97(Pta2): 105524.
[26] Yuan Y, Wu L, Zhang X. Gini-impurity index analysis. IEEE Transactions on Information Forensics and Security, 2021, 16: 3154–3169.
[27] Takaya S, Marc R, Guy B. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE, 2015, 10(3): 1–21.
[28] Shen Z Q, Zheng H L, Li T T, et al. The Influence of the residual Mg content in the ductile cast iron on the formation law of spherodial graphite. Advanced Materials Research, 2011, 415–417: 907–914.