Prediction of end-point phosphorus content of molten steel in BOF with machine learning models

Y. Kang [a], M.-M. Ren [a], J.-X. Zhao [a,*], L.-B. Yang [b,*], Z.-K. Zhang [a], Z. Wang [a], G. Cao [a]

[a] School of Metallurgical Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, Shaanxi, China

[b] Central Iron & Steel Research Institute, Beijing 10081, China

*Corresponding author:

Jun-xue Zhao, zhaojunxue1962@126.com

Li-bin Yang, libyangz@126.com

(Received 06 March 2023; Accepted 08 April 2024)

Abstract

The main task of basic oxygen furnace (BOF) steelmaking is dephosphorization, thus the prediction and control of End-point phosphorus content of molten steel is of great significance. Four machine learning regression models (Lasso, Random Forest, Xgboost, and Neural Network) were established to predict end-point phosphorus content of molten steel in BOF according to raw and auxiliary material data, process parameters, and data of production quality. The prediction effect of four models were further compared, and their prediction results were interpreted via model's interpretability and Permutation Importance Method. Results showed that compared with linear regression and Neural Network regression model, two kinds of ensemble tree model had higher prediction accuracy, better stability in small data sets, and lower requirements on data pre-processing. The influencing factors of end-point phosphorus (P) content in BOF were ranked by importance as: Tapping temperature > Turning down times > Steel scrap amount > Operation habits of different work groups > Blowing oxygen amount > Sulfur and Phosphorus content of molten iron > Addition amount of lime, limestone, and light-burned dolomite in slagging agents > Slag-splashing amount.

Keywords

Converter steelmaking; Machine Learning; Ensemble tree model; Model interpretability; Influencing factor ranking; End-point prediction

1 Introduction

Rising cost of raw material and fierce market competition requires iron and steel enterprises to ceaselessly exploit their potential, optimize producing process, and refine operations to realize effective, low-cost and greenized manufacturing[1]. The final purpose of BOF steelmaking, a significant step in steelmaking process, is to obtain molten steel with qualified constitution and temperature while dephosphorization, as the core mission in steelmaking, is of crucial importance to reach greenization and high efficiency. Common research methods tend to evaluate and analyze the manufacturing process via

establishing dephosphorization models. Metallurgical practitioners have proposed various mechanism models and empirical formulas based on metallurgical principles and their own experience[2-5]. However, these models and formulas can neither accurately predict end-point phosphorus (P) content in BOF nor analyze the dephosphorization process in depth because of their distinguished production environment and the multi-dimensional, multi-variable, non-linear, uncertain, and various factor-coupling correlated nature of most steps in BOF steelmaking[6]. A new solution is in urgent need.

Development and application of machine learning inspired metallurgical practitioners to apply new technologies in addressing related issues such as end-point P content prediction in BOF process etc. Li et al. [7] established a model based on Levenberg-Marquardt algorithm of BP neural network to predict the end-point phosphorus (P) content in BOF process, which increased model's rate of convergence when avoiding the problem of local minimum. Wang et al. [8] hybridized weighted K-means clustering algorithm and GMDH (Group Method of Data Handling) polynomial neural network techniques, and built a prediction model of end-point P content in BOF that is more advantaged than BP Neural Network. He et al.[9] adopted principal component analysis to reduce dimension of the factors influencing the prediction, thus proposing a prediction model of end-point P content in BOF based on PCA and BP Neural Network. Sala et al.[10] used Ridge Regression and two kinds of ensemble tree model to form two data-driven prediction models for end-point components and temperature in BOF, one of which contained time series data while the other did not. Results showed the model with time series data had higher prediction accuracy. Zhou et al.[11] set steelmaking process and end-point P content as the constraint condition of BP neural network based on metallurgical rules, and established a prediction model of end-point P content in BOF based monotone-constrained back-propagation (BP) neural network. Li et al.[12] applied Least Squares Twin Support Vector Machines to classify end-point phosphorus partition ratio in BOF steelmaking based on slag components and tapping temperature, which achieved a relatively high prediction accuracy. Li et al.[13] selected 11 basic machine learning models and integrated them with Averaging and Stacking, further improving the prediction effect on steel quality.

In addition to the prediction of the end point phosphorus content of the converter, many scholars have also proposed new methods for the determination of other end point components. Liu et al.[14] proposed a novel method based on accurate and fast multi flame features extraction and general regression neural network (GRNN). Wang et al.[15] combined the characteristics of genetic algorithm and BP neural network completely, and a combined GA-BP neural network model was established. Chang et al.[16] proposed a new multi-channel graph convolutional network to integrate these correlations with the process variables together for a more accurate prediction model. Gu et al.[17] proposed an improved CBR model using time-series data (CBR_TM) to predict the end-point carbon content and temperature in the converter according to the data types of process parameters. Song et al.[18] established an intelligent case-based hybrid converter model to predict the converter endpoint and process operations.

Existing researches concern more about models' prediction accuracy and have achieved their purpose from perspective of data, features, and algorithms etc. Yet neural network models have poor interpretability, which makes metallurgical practitioners unable to conclude metallurgical rules from these models or apply them in improving manufacturing process.

The ceiling of machine learning is determined by data and their features, and can only be approached by models and algorithms. Different data processing strategies have different points, and can deal with different issues. Researchers should pay more attention to algorithm's interpretability and make purpose-oriented choice. This study established two kinds of ensemble tree models (Random Forest [RF] and Xgboost [XGB]), one neural network model [BPNN] and one linear regression model (Lasso) based on integrated and processed raw data from BOF steelmaking. Furthermore, results of four models were compared and interpreted from the perspective of metallurgy.

2 Models Adopted

2.1 Ensemble methods and ensemble tree regression models

Bagging randomly selected n sampling sets from the original data set by Bootstrap sampling, based on which n weak learners can be trained. The final strong learner can be obtained through conducting ensemble strategy on weak learners. RF is a representative algorithm in Bagging, and its random selection on features endows it with better generalization capability[19]. Besides, RF may sort all features by their importance to the predicted target while holds interpretability.

Boosting, on the other hand, will train a weak learner 1 from the training set based on the original weight, and adjust weight of samples according to weak learner 1's learning error rate to obtain a weak learner 2. Repeat this process till the number of weak learners reach n set in advance. The final strong learner comes by integrating these n weak learners via ensemble strategy. Different from Adaboost, Gradient Boosting Decision Tree (GBDT) fit the residual in every iteration to realize as less sample loss as possible. Xgboost is an improved GBDT, which takes second-order Taylor expansion for better accuracy and excels in its higher operational efficiency, effective processing of missing values, and great generalization ability[20].

2.2 Neural network regression

BP neural network (BPNN) is a kind of multilayer feedforward neural network whose learning process is composed by signal forward-propagation and error back-propagation[21]. Specifically speaking, for neural network models with only one hide layer BPNN, it mainly works in two stages: for the first stage, sample's features are input from the input layer, and the signal will be propagated forward through the hide layer and reach the output layer; for the second stage, the error between network's actual and expected value will be propagated backward from the output layer to the hide layer, then to the input layer. Weight of neurons in all layers are adjusted based on errors. The target can be achieved in weight adjustment of iterative calculation round by round.

2.3 Lasso regression

L1-regularized linear regression is usually called Lasso regression. Compared with common linear regressions, Lasso regression adds a L1-regularized term to loss function. In L1-regularized term, there is a constant coefficient α regulating the mean square error term and the weight of regularized term in loss function. Specific loss function of Lasso regression is as **Eq. (1)**:

$$J(\theta)=\frac{1}{2}n(X\theta-Y)^{T}(X\theta-Y)+\alpha\|\theta\|_{1} \quad (1)$$

Number of samples is expressed as n, constant coefficient as α, L1 norm as $\|\theta\|_1$.

Lasso regression may minimize coefficient for some features, even lowering some with relatively small absolute value to 0. This would improve model's generalization capability[22]. As a linear regression model, Lasso can select different features to simplify influencing factors and improve model's interpretability, which will be of great convenience for metallurgical practitioners' understanding of steelmaking process.

3 Model establishment

3.1 Feature selection and data processing

Data in this paper are derived from manufacturing data of one enterprise's BOF plants. The steelmaking process in BOF is shown in **Figure 1**. **Figure 1a** is the slag-splashing furnace protection of BOF. Slag-splashing operation concerns mainly slag-splashing timing and blowing nitrogen amount. These two factors indirectly reflect the internal outline of BOF and physical and chemical properties of slag at early stage, which is related to dephosphorization. **Figure 1b** refers to the loading stage in steelmaking. The change of scrap ratio will influence bath temperature, slagging rate, slag components, and slag-splashing effect etc. at the same time, influencing dephosphorization directly or indirectly; besides, the components, temperature, and adding amount of molten iron are initial conditions that directly influences dephosphorization process. **Figure 1c** represents the blowing stage, the most important stage of dephosphorization, of steelmaking. It mainly affects the process by oxygen blowing operation and drossing process, in which the former one determines the dynamic condition while the latter one determines the thermodynamic condition by slag's physical and chemical properties. These two factors act and coupling with each other. What is more, turning down times indicates the effect of single/double slag operation and reblowing/un-reblowing on dephosphorization. Staff in different groups have different operation habits and experience, especially in the adjustment of oxygen lance height and the adding moment of various slagging agents. At data level, such differences will lead to fluctuation in the distribution of same factor in different groups.

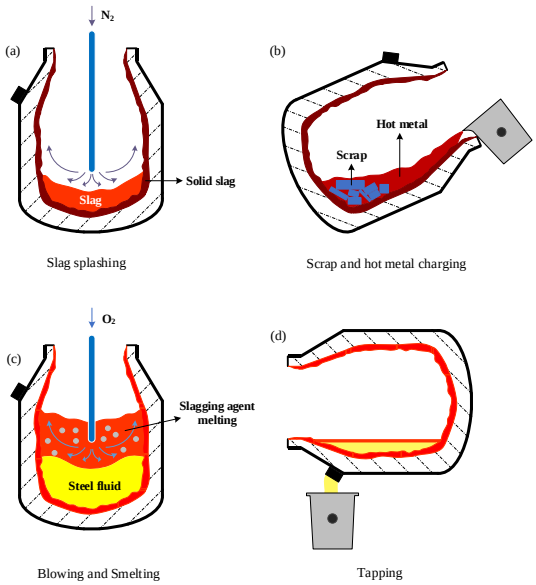

Figure 1 Flow chart of steelmaking process in BOF. **a** Slag splashing; **b** Scrap and hot metal charging; **c** Blowing and smelting; **d** Tapping

Based on these facts, all features in production data of BOF that may influence target prediction were considered in this paper, mainly the addition of raw and auxiliary materials, molten iron condition, and operation coefficients in manufacturing process etc. Details are as follow: Work group, Carbon content in molten iron, Silicon content in molten iron, Manganese content in molten iron, Phosphorus content in molten iron, Sulfur content in molten iron, Temperature of molten iron, Quality of molten iron, Quality of steel scrap, Steel output, Oxygen blowing time, Oxygen blowing amount, Turning down times, Lime, Limestone, Raw dolomite, Iron ore, Light-burned dolomite, Temperature of the first turning-down, Reblowing time, Tapping temperature, Nitrogen blowing amount, Slag-splashing time, Iron mixing times. Except for work groups, which belong to character data, all other process data are numerical data. Within these numerical data, 'Pour furnace' and 'Iron mixing times' are integers (discrete ones) larger than 0, while other data are continuous values. All process data do not include unstructured data such as time series data or image data. Brief notes for all features are listed in **Table 1**.

Table 1 List of Influencing Factors and Their Brief Notes

Influencing factors	Brief notes	Influencing factors	Brief notes
Work *group* A	F1	Oxygen blowing amount	F14
Work group B	F2	Turning down times	F15
Work group C	F3	Lime	F16
Carbon content in molten iron	F4	Limestone	F17
Silicon content in molten iron	F5	Raw dolomite	F18
Manganese content in molten iron	F6	Iron ore	F19
Phosphorus content in molten iron	F7	Light-burned dolomite	F20
Sulfur content of molten iron	F8	Temperature of the first turning-down	F21
Temperature of molten iron	F9	Reblowing amount	F22
Molten iron amount	F10	Tapping temperature	F23
Steel scrap amount	F11	Nitrogen blowing amount	F24
Steel output	F12	Slag splashing time	F25
Oxygen blowing time	F13	Iron mixing times	F26

Extreme outliers were manually removed from original data based on feature distribution map and manufacturing experience while other outliers were removed according to 3σ principle. Missing values mainly include unrecorded data and null data. Unrecorded data, such as temperature of molten iron and steel scrap amount, were filled by the average value. Null data, such as Raw dolomite and Addition of Limestone, were filled by 0; or by the default value, which was 1, such as Turning down times and Iron mixing times. Work group A, B, and C were dumb-coded and expressed as [(1,0,0), (0,1,0), (0,0,1)] respectively. **Eq. (2)** was adopted for z-score standardization of data features.

$$y_i = \frac{x_i - \bar{x}}{s} \tag{2}$$

Figure 2 shows correlation efficient among each feature. The correlation between each variable, and between variables and their predicted targets, were analyzed by heat map. Results can provide references for engineers and technicians in process optimization.

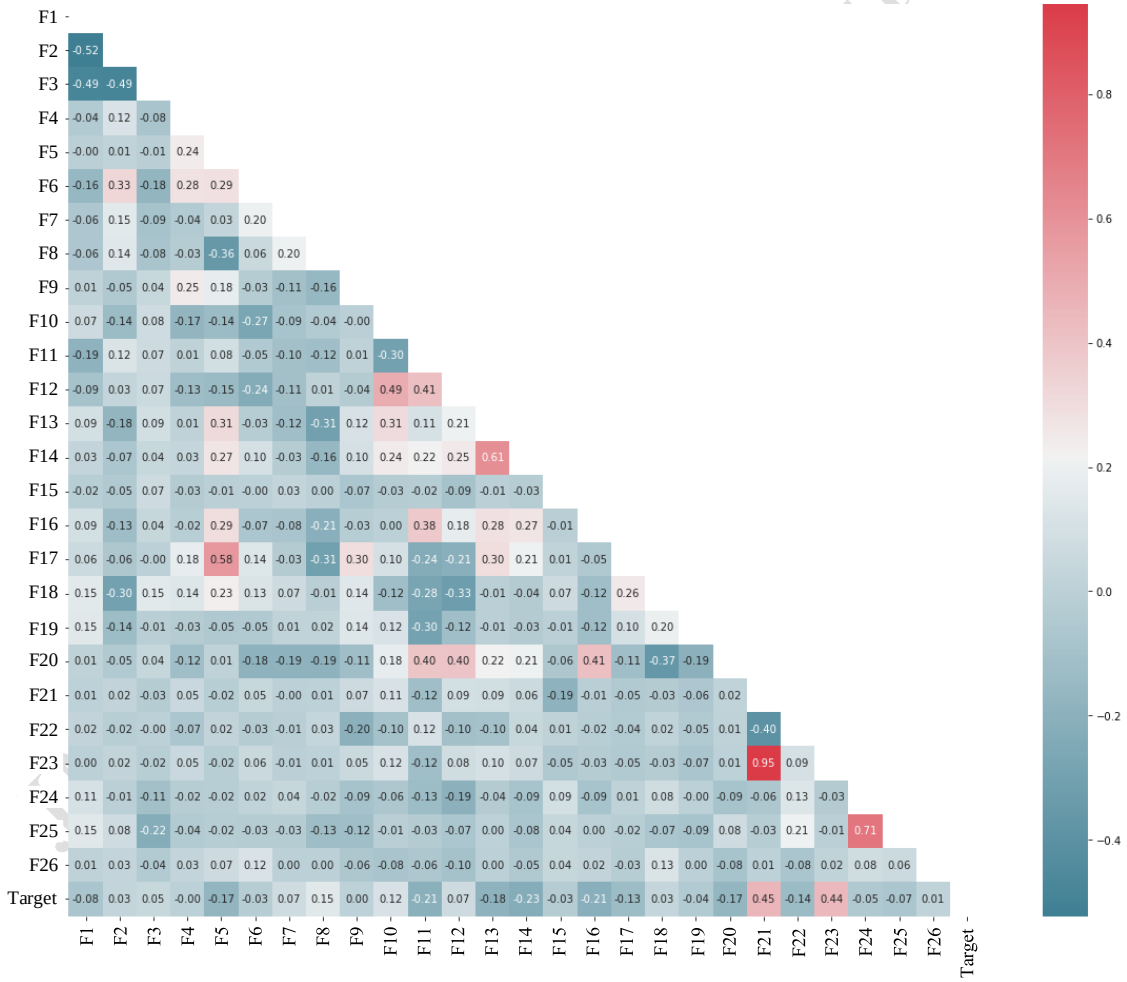

Figure 2 Correlation efficient among feature variables

3.2 Model training

All model hyperparameters were determined using a grid search and cross-validation.

The main hyperparameters of RF are “n_estimators”, “min_samples_split”, “min_samples_leaf”, “max_depth”, “max_features”. “n_estimators” represents the number of trees in the forest. If it is too small, the model would be easy to underfitting. When the number of trees exceeds a certain level, its improvement would be limited. In this paper, the number varied between 20 to 300, with a step size of 10. “min_samples_split” represents the minimum number of samples required to split an internal node. It was set between 2 to 32 with a step size of 2. “min_samples_leaf” restricts the minimum number of samples for a leaf node. It was set between 1 to 16 with a step size of 1. “max_depth” were not limited to pursue a smaller training error. These three parameters were used to avoid overfitting to improve the generalization of the model. “max_features” can be set to “auto”. Only when the number of features is large should we consider limiting the maximum number to control the generation time of the decision tree. All other hyperparameters were set to default.

The following hyperparameters are optimal for RF regression model. The number n of weak learners was set as 250. Minimum samples split was set as 2, and minimum samples leaf nodes as 8.

For XGBoost, the main hyperparameters needed to be adjusted were “n_estimators”, “max_depth”, “min_child_weight”, “gamma”, and “learning_rate”. Other framework parameters, booster and objective, were just set as default parameters in the regression task. The “learning_rate” shrinks the feature weights to make the boosting process more conservative after each boosting step. Smaller “learning_rate” means more iterations of the weak learners to reach the model's optimal results, but it would take longer time. In this paper, we fix a learning rate first, and adjust “n_estimators”. The “learning_rate” was set at 0.01~0.3. “n_estimators” represents the number of trees, and the model would be too complex when it is too large. It was set between 10 to 200 with a step size of 10. “max_depth” controls maximum depth of a tree. Increasing this value will make the model more complex and more likely to overfit. It was set between 2 to 20 with a step size of 2. “min_child_weight” represents minimum sum of instance weight (hessian) needed in a child. The larger it is, the more conservative the algorithm will be. “gamma” is an important hyperparameter used by XGB to prevent overfitting, and it represents the minimum loss reduction required to make further partition on a leaf node of the tree. The larger “gamma” is, the more conservative the algorithm will be. The parameter was gradually increased from 0 to 1. Above hyperparameters are the main objects of adjustment. Only when the model remains overfitting after their adjustment, other hyperparameters would be considered.

The following hyperparameters are optimal for XGB regression model. Subtree number was set as 120, weight reduction factor for each weak learner as 0.08, max depth of tree structure as 6, minimum child weight as 90, and gamma as 0.

BP neural network structured as a hidden layer with 15 neurons. Experimentally, more complex structure tends to lead to model overfitting. The activation function was ReLU function because it is faster to train, avoiding the problem of vanishing gradients and it is suitable for most neural networks, especially MLP and CNN. “max_iter” is the maximum number of iterations. It was set between 10~1000. The solver was “lbfgs”. For small datasets, however, “lbfgs” can converge faster and perform better. If the solver is “lbfgs”, the regressor would not use minibatch and “learning_rate”. “Alpha” stands for strength of the L2 regularization term. It prevents the neural network from overfitting and

was set between 0.00001~0.1, and the search followed the logarithemic scale. The final L2 regularization term α=0.01.

The Lasso regression model only needs to be optimised for the hyperparameter "alpha". It constant that multiplies the L1 term, controlling regularization strength. It ranged from 0.00001 to 10, and the final L1-regularized term α=0.00005.

All four models suit regression issues with multiple features, and their effects were all evaluated by **Eq. (3)**, mean absolute error (MAE), and **Eq. (4)**, mean squared error (MSE).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|predicted_i - actual_i\right| \tag{3}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(predicted_i - actual_i\right)^2 \tag{4}$$

4 Comparison of model's prediction results

Figure 3 compared evaluation indexes MAE and MSE of four models. Two indexes were obtained through simple cross validation. Simple cross-validation here refers to cross-validation performed only once in data split. It is shown in **Figure 3** that MAE and MSE of two ensemble learning models were smaller than that of Lasso and BPNN model. Specifically speaking, Lasso surpassed BPNN while RF slightly outrun XGB. This is because that compared with linear regression models, ensemble models work based on cart decision tree, which endows it with a higher accuracy in regression. Besides, multi-model ensemble methods, Bagging and Boosting, were adopted by two ensemble models, which improved their fitting accuracy and generalization capability at the same time and guaranteed their better performance than single-model BPNN and Lasso. As for Lasso model, it surpassed BPNN since it can reduce features and avoid overfitting issue of linear regression for its addition of L2-regularized term. Compared with the other three models, BPNN's performance may be limited by restricted sample amount, which lowered its fitting accuracy.

It should be noted that the actual end-point P content is 0.030 % on average, and the minimum absolute error of prediction is 0.0046 %, which is 15.3 % off the average. In actual production process, the P content is measured by direct reading spectrometer, which also causes certain errors in the sampling and location of the measurement points. From the stand of converter steelmaking's requirements, such error is acceptable.

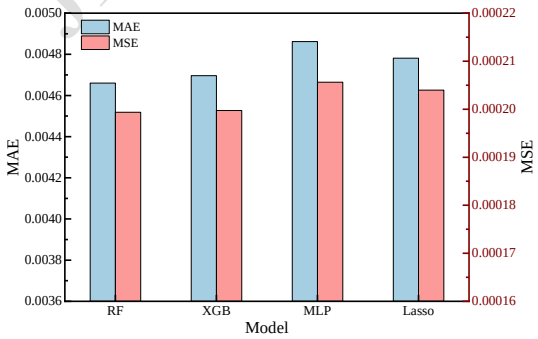

Figure 3 Evaluation indexes for four models

Figure 4 showed the learning curves of four models evaluated by index MAE (score). Cross-validation score comparison results of four learning curves in **Figure 4** were in line with their performance in **Figure 3**, yet they all had lower error than in **Figure 3** that adopted single cross validation. This is because learning curves were drawn with cross-validation conducted via ShuffleSplit function (n_splits=100, test size=0.2), which means every point's score was calculated as the average error of 100 random samplings, thus lowering error fluctuation. Combing results of **Figure 3** and **Figure 4**, accuracy of different models was ranked as RF > XGB > Lasso > BPNN.

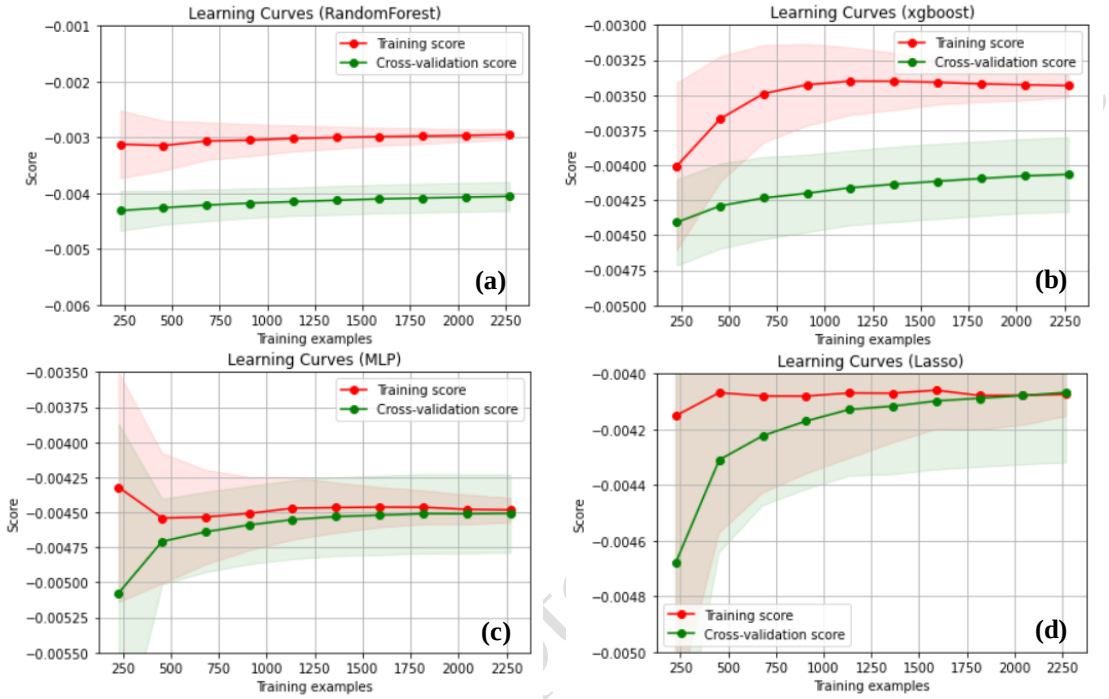

Figure 4 Model learning curves with multi-sampling

Figure 5 showed the learning curves with single sampling. That means all points' scores were obtained by one sampling, i.e. the split n was set as 1 in ShuffleSplit function. It was shown in **Figure 5a** and Figure 5b that the learning curves of ensemble tree models tend to smooth out rapidly, reflecting RF and XGB's stability in processing this issue type. Their performance remained good in small data set. **Figure 5c** and **Figure 5d** indicated that BPNN and Lasso required at least 1,500 items of data. Lasso showed bad stability in relatively small data set and BPNN's learning curve gradually converged when data reach 1,600 items. All these suggested BPNN and Lasso had higher requirements on data amount, and larger error in small data set. RF and XGB, on the other hand, did not need normalization or standardization for original data. XGB algorithm can deal with missing values by itself, making its application easier. Considering results above, two kinds of ensemble tree models had better stability in small data set and lower requirements on data processing than BPNN and linear regression model.

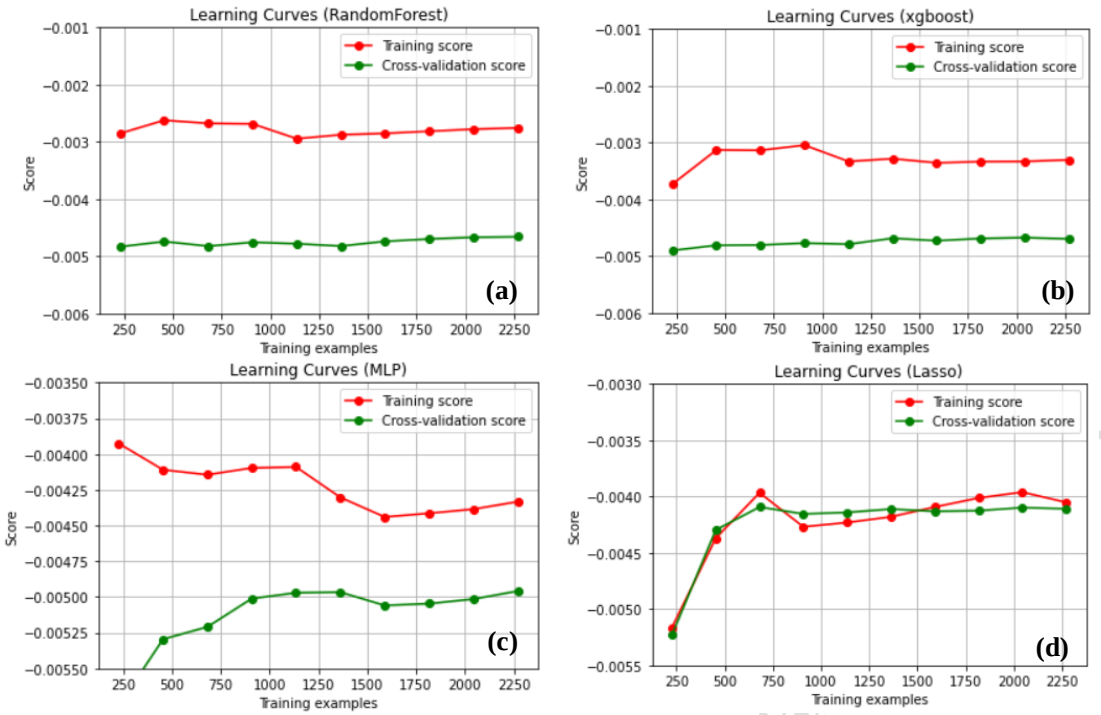

Figure 5 Learning curves with single-sampling

5 Model interpretation and analysis

Metallurgical practitioners mostly optimize the manufacturing process by selecting influencing factors and controlling variables based on their own experience, which requires much time and effort. Yet with machine learning, technicians may realize optimization by integrating evaluation results in multi-models, selecting influencing factors and determining their priorities.

Figure 6 compared the importance of all influencing factors. RF and XGB measured their importance (parameter as gain) via the extent of gain brought by split of different features based on the decision tree. Larger gain reflects greater importance. BPNN measured them by the Permutation Importance, which means to observe the change of index after permuting one certain feature. Larger change of index refers to higher importance[23]. For better comparison, importance calculation results of RF, XGB, and BPNN model were unified into percentage and drawn as a histogram.

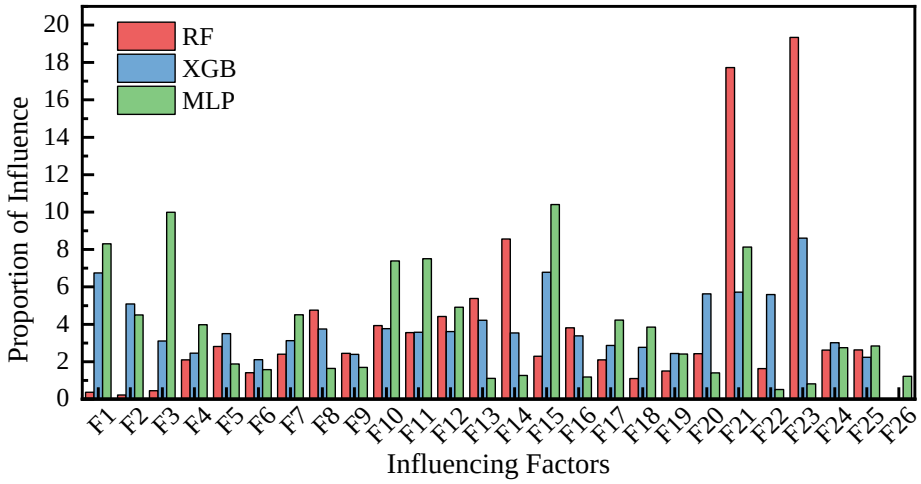

Figure 6 Comparison of different influencing factor's importance

On one hand, three models have different mathematics theories and basis in importance calculation. On the other hand, each model evaluates via relatively single index that is distinguished from each other. These led to different importance sorting of all features, and results of one model cannot represent the actual situation in manufacture practice. Therefore, the average value of RF, XGB, and MLP was taken as the comprehensive and final evaluation on feature's importance. Coefficient in linear regression was adopted to judge whether the feature's influence was positive or negative. What should be mentioned is that positive influence here means higher coefficient value relating to higher end-point P content, but it is actually a negative influence in metallurgical practice. **Figure 7** showed comprehensive importance evaluation results. The comprehensive evaluation score was shown in left vertical coordinate, and the coefficient of linear regression model drawn in point and line chart whose values can be checked by right vertical coordinate.

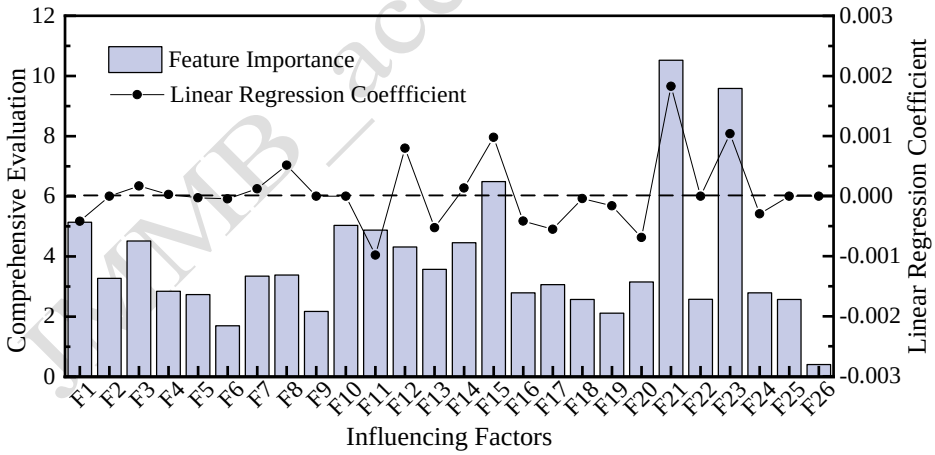

Figure 7 Comprehensive evaluation on feature's importance

All influencing factors were divided into 11 groups based on BOF operations. As shown in **Figure 7**, the importance of all groups were ranked as: Tapping temperature in BOF (F21, F23) > Turning down times (F15) > Material addition and Steel output (F10, F11, F12) > Work groups (F1, F2, F3) > Oxygen blowing operation (F13, F14) > Molten iron components (F4~F8) > Slagging agents (F16~F20) > Slag-splashing operation (F24,

F25) > Reblowing amount (F22) > Molten iron temperature (F9) > Iron mixing times (F26).

Based on the sorting of operation processes and coefficient of linear regression model showed in **Figure 7**, it can be concluded that the first group, including F21 (First turning down temperature) and F23 (Tapping temperature), contains the most significant influencing factors. These two temperatures differ when the composition and temperature of molten steel are not qualified. For qualified ones, First turning down temperature will be the Tapping temperature. But for those who do not, further operation will be conducted for adjustment, and the temperature detected then is the Tapping temperature. Temperature plays an important role in dephosphorization. The higher the temperature is, the worse dephosphorization will be[24]. The coefficient of both F21 (First turning down temperature) and F23 (Tapping temperature) in the linear regression model were the maximum positive values, suggesting higher Tapping temperature is related to larger end-point P content, which is in line with manufacturing experience and dephosphorization mechanism. F15 (Turning down times) of the second group also had relatively greater importance. F15 larger than 1 indicated the possibility of reblowing or second slag formation blowing (double-slag process). Turning down times was positively related to end-point P content. Combing **Figure 2**'s correlation analysis, Turning down times was most strongly correlated with work group 3, which suggested their poorer stability than other work groups. In the third group, F10 (Molten iron amount), F11 (Steel scrap amount), and F12 (Steel output), linear regression coefficient indicated the addition of steel scrap amount was negatively related to end-point P content while Steel output is in positive relation. Since Lasso model can reduce variables, it lowered Molten iron amount's coefficient to 0 (Molten iron amount and Steel output have collinear relation), which means staff only need to optimize Steel scrap amount. The fourth group, work group also played a certain role. To the end-point P content, the influence of work group A was negative, of group B was insignificant, of group C was positive. This suggested work group C should learn more experience from group A. As for the fifth group, Oxygen blowing operation, containing F13 (Oxygen blowing time) and F14 (Oxygen blowing amount), influenced the predicted object to a certain extent. Longer F13 was related to higher oxidation in molten pool, thus lower end-point P content. F14 is in positive influence for end-point P content, which may because excessive F14 will lead to more rapid temperature rising in molten pool, inhibiting dephosphorization or even trigger rephosphorization[25]. This indicated staff should to pay more attention to the control of oxygen blowing amount. The sixth group included molten iron components F4~F9. In molten iron, S and P had positive effect on end-point P content while others did not have significant influence, suggesting technicians can improve dephosphorization effect by controlling molten iron components. The seventh group discussed the influence of slagging agent addition, among which the most significant ones were lime, limestone, and light-burned dolomite. All these were negatively related to end-point P content. For the eight group, including F24 (Nitrogen blowing amount) and F25 (Slag splashing time), it was showed that higher F25 may raise the oxidation of molten pool at primary stage, and assist dephosphorization in early stage. Both importance score and linear regression model coefficient showed that F9 (Molten iron temperature), F22 (Reblowing amount), and F26 (Iron mixing times) contained in last three groups had relatively small influence on end-point P content.

However, in Lasso model, the features of coefficients close to 0 may exist in two cases. One is caused by the existence of L1 regular term. This can be solved by reducing the hyperparameters "Alpha". When the coefficient deviates from 0, we can clarify the

extent and direction of the influence, but the model's generalization capability may also be impaired at this time. Another situation is that the feature itself has little influence on end-point P in the linear regression model. But that only means this feature is unimportant just in Lasso model. The judgement of their influencing extent needs to be coupled with the importance ranking given by the other three algorithms.

In conclusion, removing factors with relatively small influence and those cannot be improved in every group, factors needed to be concerned are ranked as: Tapping temperature in BOF > Turning down times > Steel scrap amount > Group's operation habit > Oxygen blowing amount > S and P content in molten iron > addition of lime, limestone, light-burned dolomite in slagging agents > Slag splashing amount. Among all these, the former 6 ones had positive influence while the latter two had negative one.

6 Conclusions

1. Influencing factors of end-point P content mainly includes raw and auxiliary material addition, condition of molten iron, coefficients of process operation, and work group's operation habits. Four machine learning models were established to predict the end-point P content in BOF. The accuracy of four models were ranked as: RF > XGB > Lasso > BPNN. Besides, compared with BPNN and linear regression model (Lasso), two kinds of ensemble tree models had smaller error and better stability in small data set, and lower requirements on data processing.
2. Through comprehensive consideration on evaluation results of RF, XGB, BNPP, and Lasso model, influencing factors of end-point P content in BOF were ranked by importance as: Tapping temperature in BOF > Turning down times > Steel scrap amount > Work group's operation habits > Oxygen blowing amount > S and P content in molten iron > addition of lime, limestone, light-burned dolomite in slagging agents > Slag splashing amount. Among all these, the former six were of positive influence while the latter two were of negative influence. The optimization of manufacturing process can be guided by interpreting prediction results of these models.
3. BOF steelmaking is a complicated process. There are still bunches of key processing data that cannot be taken into regression calculation directly, including voiceprint data in sonar slag-reducing technique, mouth flame image data of converter, and time series data such as oxygen lance position, lime loading model, bottom blown pattern, and change of gas composition etc. These data contain important information of changes in the whole metallurgy process. An effective processing or conversion method for them is in urgent need. Bringing these key data into regression analysis can help engineers and technicians further understand and analyze steelmaking process.

Acknowledgements

This work was supported by the National Natural Science Foundation of China [grant numbers 51674185, 51674186].

CRediT authorship contribution statement

Yi Kang: Conceptualization, Methodology, Programming, Validation, Formal analysis. **Meng-meng Ren:** Investigation, Writing - review & editing. **Junxue Zhao:** Supervision, Funding acquisition. **Li-bin Yang:** Resources, Investigation. **Zhen-kai**

Zhang: Investigation, Formal analysis. **Ze Wang:** Investigation, Writing - review & editing. **Geng Cao:** Data preprocessing, Software.

Data Availability Statement

Research data are not shared.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] K. He, L. Wang, A review of energy use and energy-efficient technologies for the iron and steel industry, Renewable and Sustainable Energy Reviews, 70 (2017) 1022-1039. https://doi.org/10.1016/j.rser.2016.12.007

[2] G. W. Healy, New look at phosphorus distribution, J Iron Steel Inst, 208(7) (1970) 664-668.

[3] H. Suito, R. Inoue, and M. Takada, Phosphorus distribution between liquid iron and MgO saturated slags of the system CaO-MgO-FeOx-SiO2, Tetsu-to-Hagane, 67(16) (1981) 2645-2654. https://doi.org/10.2355/tetsutohagane1955.67.16_2645

[4] A. Sobandi, H. G. Katayama, T. Momono, Activity of Phosphorus Oxide in CaO–MnO–SiO2–PO2. 5 (–MgO, FetO) Slags, ISIJ international, 38(8) (1998) 781-788. https://doi.org/10.2355/isijinternational.38.781

[5] W. Yang, J. Yang, Y. Shi, Z. Yang, F. Gao, R. Zhang, G. Ye, Effect of basicity on dephosphorization of hot metal with a low basicity slag at 1653 K, Ironmaking & Steelmaking, 48(1) (2021) 69-77. https://doi.org/10.1080/03019233.2020.1731256

[6] X. F. Jiang, X. Wang, Study on logistic information acquisition technology in steelmaking practice based on RFID, In 2008 7th World Congress on Intelligent Control and Automation, June, Chongqing, China, 2008, p. 7946-7950. IEEE. https://doi.org/10.1109/WCICA.2008.4594171

[7] C. R. Li, H. W. Zhao, X. Xie, Q. Yin, Prediction of end-point phosphorus content for BOF based on LM BP neural network, Iron and Steel/ Gangtie, 46(4) (2011) 23-25. http://www.chinamet.cn/Jweb_gt/EN/abstract/abstract145429.shtml

[8] H. B. Wang, A. J. Xu, L. X. Ai, N. Y. Tian, Prediction of endpoint phosphorus content of molten steel in BOF using weighted K-means and GMDH neural network, Journal of Iron and Steel Research International, 19(1) (2012) 11-16. https://doi.org/10.1016/S1006-706X(12)60040-5

[9] F. He, L. Zhang, Prediction model of end-point phosphorus content in BOF steelmaking process based on PCA and BP neural network, Journal of Process Control, 66 (2018) 51-58. https://doi.org/10.1016/j.jprocont.2018.03.005

[10] D. A. Sala, A. Jalalvand, A. Van Yperen-De Deyne, E. Mannens, Multivariate time series for data-driven endpoint prediction in the basic oxygen furnace, In 2018 17th IEEE international conference on machine learning and applications (ICMLA), December, Orlando, FL, USA, 2018, pp. 1419-1426. IEEE. https://doi.org/10.1109/ICMLA.2018.00231

[11] K. X. Zhou, W. H. Lin, J. K. Sun, J. S. Zhang, D. Z. Zhang, X. M. Feng, Q. Liu, Prediction model of end-point phosphorus content for BOF based on monotone-constrained BP neural network, Journal of Iron and Steel Research International, 29(5) (2022) 751-760. https://doi.org/10.1007/s42243-021-00655-6

[12] H. Li, S. Barui, S. Mukherjee, K. Chattopadhyay, Least Squares Twin Support Vector Machines to Classify End-Point Phosphorus Content in BOF Steelmaking, Metals, 12(2) (2022) 268. https://doi.org/10.3390/met12020268

[13] F. Li, J. Wu, F. Dong, J. Lin, G. Sun, H. Chen, J. Shen, Ensemble machine learning systems for the estimation of steel quality control, In 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 2245-2252. https://doi.org/10.1109/BigData.2018.8622583

[14] H. Liu, B. Wang, X. Xiong, Basic oxygen furnace steelmaking end-point prediction based on computer vision and general regression neural network, Optik, 125(18) (2014) 5241-5248. https://doi.org/10.1016/j.ijleo.2014.05.004

[15] Z. Wang, J. Chang, Q. P. Ju, F. M. Xie, B. Wang, H. W. Li, B. Wang, X. C. Lu, G. Q. Fu, Q. Liu, Prediction model of end-point manganese content for BOF steelmaking process, ISIJ international, 52(9) (2012) 1585-1590. https://doi.org/10.2355/isijinternational.52.1585
[16] S. Chang, C. Zhao, Y. Li, M. Zhou, C. Fu, H. Qiao, Multi-channel graph convolutional network based end-point element composition prediction of converter steelmaking, IFAC-PapersOnLine, 54(3) (2021) 152-157. https://doi.org/10.1016/j.ifacol.2021.08.234
[17] M. Q. Gu, A. J. Xu, F. Yuan, X. M. He, Z. F. Cui, An improved CBR model using time-series data for predicting the end-point of a converter, ISIJ International, 61(10) (2021) 2564-2570. https://doi.org/10.2355/isijinternational.ISIJINT-2020-687
[18] S. Song, J. Li, W. Yan, Intelligent Case-based Hybrid Model for Process and Endpoint Prediction of Converter via Data Mining Technique, ISIJ International, 62(8) (2022) 1639-1648. https://doi.org/10.2355/isijinternational.ISIJINT-2021-598
[19] T. K. Ho, Random decision forests, In Proceedings of 3rd international conference on document analysis and recognition, Montreal, QC, Canada, 1995, pp. 278-282 vol.1. https://doi.org/10.1109/ICDAR.1995.598994
[20] T. Q. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, August, New York, NY, USA, 2016, pp. 785–794. https://doi.org/10.1145/2939672.2939785
[21] R. Hecht-Nielsen, III.3 - Theory of the Backpropagation Neural Network (E. Editor), Neural networks for perception(Academic Press), 1992, p. 65-93. https://doi.org/10.1016/B978-0-12-741252-8.50010-8
[22] J. Ranstam, J. A. Cook, LASSO regression, Journal of British Surgery, 105(10) (2018) 1348-1348. https://doi.org/10.1002/bjs.10895
[23] A. Altmann, L. Toloşi, O. Sander, T. Lengauer, Permutation importance: a corrected feature importance measure, Bioinformatics, 26(10) (2010) 1340-1347. https://doi.org/10.1093/bioinformatics/btq134
[24] W. Yang, J. Yang, Y. Shi, Z. Yang, F. Gao, R. Zhang, G. Ye, Effect of Temperature on Dephosphorization of Hot Metal in Double‐Slag Converter Steelmaking Process by High‐Temperature Laboratorial Experiments, Steel Research International, 92(3) (2021) 2000438. https://doi.org/10.1002/srin.202000438
[25] M. Wang, W. Yang, Dephosphorization in the early stage of converter steelmaking, Ironmaking & Steelmaking, 47(10) (2020) 1127-1134. https://doi.org/10.1080/03019233.2019.1673546

Appendix

List of Illustrations, Figures, and Tables Captions