Croatian Society for Materials Protection

Metalurgija/Metallurgy

Vol. 65 No. 4 / 2026

Data-driven Prediction and Application of Steel Material Parameters

Yongshuai Xiu [1], Xiaohu Deng [1]*, Yixiao Sun [2], Yuedong Yuan [3], Gang Shen [4], Zunzhong Du [3], Xiaojun Yang [3], and Dongying Ju [4]*

https://doi.org/10.64486/m.65.4.5

[1] School of Mechanical Engineering, Tianjin University of Technology and Education, Tianjin 300222, China; 15506435784@163.com
[2] School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan 114051, China; robinsunyixiao@163.com
[3] Changshu Tiandi Coal Mining Equipment Co., Ltd., Changshu 215500, China; csyyd@126.com
[4] Zhejiang XCC Group Co., Ltd., Xinchang, Zhejiang 312500, China; 260842037@qq.com
* Correspondence: dengxh@tute.edu.cn; dyju.sitec@gmail.com

Type of the Paper: Article
Received: December 8, 2025
Accepted: February 27, 2026

Abstract: Accurate prediction of steel material parameters during heat treatment is essential for reliable finite element analysis (FEA) and process optimisation. Conventional experimental measurements and empirical models are often costly, time-consuming, and difficult to generalise to complex chemistries, microstructures, and temperature ranges. In this work, a data-driven prediction model is established using a comprehensive dataset that integrates simulation and experimental data, covering 18 elemental compositions, three typical microstructures, and a wide temperature range. Six key parameters are predicted simultaneously: thermal conductivity, specific heat capacity, yield stress, coefficient of thermal expansion, Young's modulus, and density. Five machine learning models are evaluated, among which XGBoost shows the best performance for thermal parameters, while Gradient Boosting provides the highest accuracy for mechanical properties. After hyperparameter optimisation with grid search and cross-validation, all models achieve R^2 values above 0.99 and relative prediction errors within 5 %. An integrated Steel Materials Data Management System (S-MDMS) is further developed to combine data storage, visualisation, and online property prediction. The proposed model provides an efficient route for rapid acquisition and application of steel parameters in FEA-based heat treatment design and process optimisation.

Keywords: intelligent forecasting; machine learning; heat treatment; database; steels

1. Introduction

Due to their superior combination of strength, toughness, and manufacturability, steels occupy an irreplaceable position in key industrial sectors such as aerospace, automotive manufacturing, and heavy machinery [1]. The thermal and mechanical parameters of steel are crucial for process design, service performance, and failure analysis, and their accuracy strongly affects the reliability of finite element analysis (FEA) simulations [2].

Traditionally, steel parameters have been obtained mainly through experimental measurements and theoretical calculations. Although experiments are direct and reliable, they are often limited by high cost and long testing cycles, which makes them unsuitable for large-scale studies involving multicomponent chemistries,

complex microstructures, and wide temperature ranges. Classical theoretical and empirical models also struggle to capture highly nonlinear relationships among alloy composition, microstructural evolution, and material properties, which leads to limited predictive accuracy in engineering practice [3,4]. With the growing demand for high-performance and customised steels, conventional approaches can no longer meet the requirements of efficient parameter acquisition and accurate numerical simulation.

In recent years, data-driven methods represented by machine learning (ML) have shown great potential in materials science for modelling complex data relationships and improving prediction accuracy [5-14]. For example, Xiong et al. used 360 steel datasets from the NIMS database and applied five ML algorithms—random forest (RF), least-squares regression, k-nearest neighbours, artificial neural networks, and symbolic regression—to predict fatigue strength, tensile strength, fracture strength, and hardness. They reported that RF and ANN achieved the best performance for different properties and identified tempering temperature, C, Cr, and Mo as key factors through feature selection [15]. Bhandari et al. developed a dataset for thermal conductivity of additively manufactured alloys and showed that XGBoost outperformed other algorithms, enabling rapid screening of new alloys [16]. Similar successes have been reported for predicting thermal conductivity, creep life, and other properties of steels and alloys [17-19].

At the same time, the construction and utilisation of materials databases have progressed rapidly. Zhang et al. designed a dedicated database for heat treatment process simulation and used ML algorithms to treat missing data [20]. Yang et al. pointed out that materials databases have evolved from offline storage to online sharing, and that data-mining techniques are now widely used for property prediction and curve fitting [21]. Properly constructed databases can integrate dispersed experimental and simulation data, provide high-quality training samples for ML models, and enable standardised storage and rapid retrieval of parameters [22].

However, several important challenges remain. Many existing studies focus on predicting a single property (such as thermal conductivity, fatigue strength, or creep life) in isolation, without considering intrinsic correlations among different parameters or the need for multi-property prediction in multi-field coupled simulations. In addition, although some research has introduced databases, systematic data management and integrated application platforms are still lacking. Composition, microstructure, processing, and multi-property data are frequently stored in different formats and locations, making it difficult to maintain data quality and continuously improve models.

To address these issues, this work aims to accurately predict six key parameters of steel—thermal conductivity, specific heat capacity, yield stress, coefficient of thermal expansion, Young's modulus, and density—while constructing a supporting database framework. The main objectives are:

(1) To integrate simulation data from materials computation software with experimental measurements and build a comprehensive dataset that covers 18 elemental compositions, three typical microstructures, and a wide temperature range.
(2) To evaluate five machine learning models—RF, Gradient Boosting Regression (GBR), XGBoost, Lasso regression, and Ridge regression—identify the optimal model for each target parameter, and perform hyperparameter optimisation using GridSearchCV.
(3) To develop an integrated Steel Materials Data Management System (S-MDMS) that supports systematic data management and online prediction of material parameters.

The proposed framework provides a unified platform for efficient acquisition, prediction, and application of key steel parameters. It supplies accurate input data for FEA simulations and supports both research and engineering applications in intelligent heat treatment design.

2. Materials and Methods

2.1. Dataset construction

A data fusion strategy was adopted to construct a high-quality dataset for property prediction. Simulation data from materials computation software were combined with experimental data collected from industrial plants, ensuring both diversity and reliability of the samples. The combined dataset consists of approximately

https://doi.org/10.64486/m.65.4.5

60 % simulation data and 40 % experimental data, with the experimental portion derived from factory production records. This composition was intentionally designed to exploit the extensive coverage of computationally generated scenarios while ensuring that the model remains grounded in physically measured results. Potential systematic biases inherent to simulation data, such as those arising from idealized boundary conditions or modeling assumptions, are alleviated by the inclusion of experimental data. In total, 1290 data points were compiled. The input features include chemical composition, microstructural descriptors, and temperature, while the target variables consist of six key properties: thermal conductivity (TC), specific heat capacity (CP), yield stress (YS), coefficient of thermal expansion (TE), Young's modulus (YM), and density (D).

The chemical composition is represented by the mass fractions (wt.%) of 18 alloying elements, including Fe, C, Si, Mn, Cr, Ni, Mo, V, Cu, P, S, Al, W, and Ti. Three Boolean variables—SPHSA, SPHSB, and SPHSM—indicate the presence (1) or absence (0) of austenite, bainite, and martensite, respectively. The temperature ranges from 298 K to 1258 K, covering typical processing and service conditions for steel. Basic statistical information is summarised in Table 1, and the distributions of the six target properties are shown in Figure 1. (a) Distribution of thermal conductivity; (b) Distribution of thermal expansion; (c) Distribution of yield stress; (d) Distribution of specific heat capacity; (e) Distribution of density; (f) Distribution of Young's modulus

Table 1. Summary of statistical data of the dataset used

Parameter	Min	Max	Mean	Std
Fe wt.%	94.310	97.7610	96.429462	1.122703
C wt.%	0.110	1.0500	0.338750	0.289181
Si wt.%	0.180	0.3900	0.275750	0.055838
Mn wt.%	0.300	0.8600	0.623250	0.209391
Cr wt.%	0.710	1.6800	1.221000	0.364618
Ni wt.%	0.000	3.3300	0.803725	1.106759
Mo wt.%	0.000	0.2700	0.113025	0.095723
V wt.%	0.000	0.0100	0.001250	0.003325
Cu wt.%	0.000	0.2500	0.041937	0.082766
P wt.%	0.011	0.0300	0.019213	0.007851
S wt.%	0.001	0.0400	0.016450	0.015064
Al wt.%	0.000	0.9100	0.113750	0.302534
W wt.%	0.000	0.0100	0.001250	0.003325
Ti wt.%	0.000	0.0091	0.001187	0.003008
SPHSA	0.000	1.0000	/	/
SPHSB	0.000	1.0000	/	/
SPHSM	0.000	1.0000	/	/
T/K	298	1258	765.25	381.750125

2.2. Data preprocessing

Systematic data preprocessing was conducted to minimise the influence of noise and inconsistent scales on model performance. First, the dataset was checked for missing values and anomalies. Outlier detection was performed using the boxplot method, where the interquartile range (IQR) for each feature served as the reference interval. Samples falling outside the range defined by the IQR criterion were classified as outliers. No missing or anomalous values were found, because the dataset had been carefully curated and manually verified in advance.

Next, feature normalisation was applied to address the large differences in units and numerical scales among the input variables. For example, temperature spans almost 1000 K, whereas the carbon content is confined to a narrow range. Z-score standardisation was used, as defined in Eq. (1):

$$x_{norm} = \frac{x - \mu}{\sigma} \quad (1)$$

https://doi.org/10.64486/m.65.4.5

where x is the original value, μ is the mean, and σ is the standard deviation [23]. After standardisation, all continuous features have zero mean and unit variance, which eliminates unit disparities and balances their contributions during model training. Microstructural features were not normalised, because they are binary indicators.

Finally, the whole dataset was randomly divided into training and test sets using a 9:1 ratio. This sampling strategy ensures that both subsets maintain similar distributions of alloy composition and temperature, and it avoids evaluation bias due to uneven data allocation. The overall data processing and modelling workflow is illustrated in Figure 2.

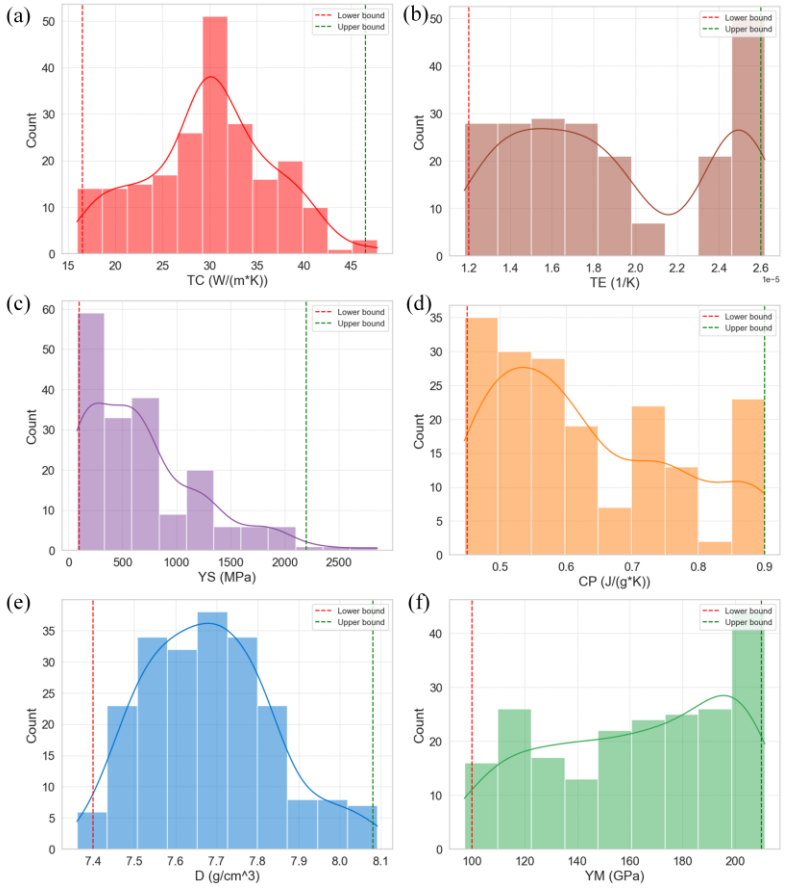

Figure 1. (a) Distribution of thermal conductivity; (b) Distribution of thermal expansion; (c) Distribution of yield stress; (d) Distribution of specific heat capacity; (e) Distribution of density; (f) Distribution of Young's modulus

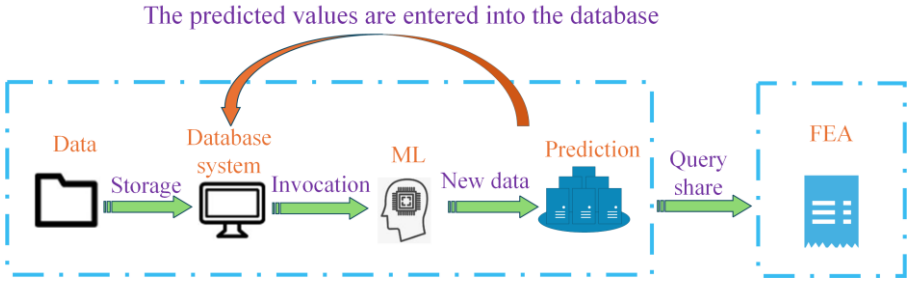

Figure 2. The workflow diagram of constructing parameter prediction and database

https://doi.org/10.64486/m.65.4.5

2.3. Hyperparameter optimisation

To fully exploit the predictive potential of the models, comprehensive hyperparameter optimisation was conducted. Grid search was used to explore key parameters such as learning rate, maximum tree depth, number of estimators, subsample ratio, column sampling ratio, and regularisation parameters, with the search ranges listed in Table 2 [24]. A ten-fold cross-validation strategy was adopted: in each iteration, nine folds were used for training and one fold for validation.

The performance of the optimised models on the training and test sets is presented in Figure 3. Hyperparameter tuning significantly reduces the prediction errors of all target properties. For example, the MAE of XGBoost for TC decreases from 0.447 W/(m·K) to 0.402 W/(m·K), corresponding to a reduction of about 10 %. For YS, the MAE of the GBR model is reduced from 41.10 MPa to 11.23 MPa, and its R^2 value increases to 0.998, which confirms the substantial improvement in reliability.

Figure 4 shows scatter plots of predicted versus measured values for the training and test sets. In all cases, the data points are densely clustered along the unit-slope line without obvious systematic deviation or overfitting. All target parameters reach R^2 values above 0.99, indicating that the models explain more than 99 % of the observed variance. Both RMSE and MAE remain at low levels, which verifies the robustness of the optimised models for practical engineering applications and real-time prediction scenarios.

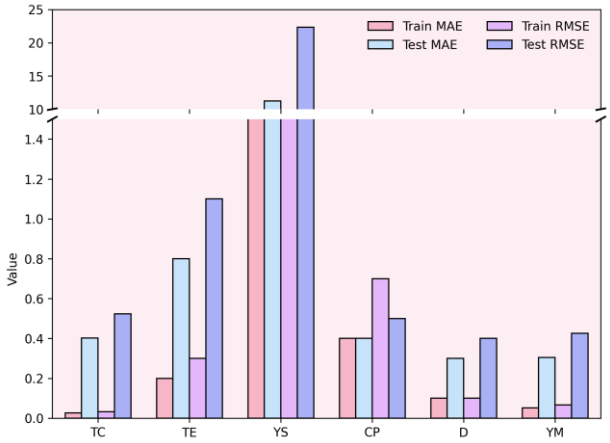

Figure 3. Optimized model performance on training and test sets

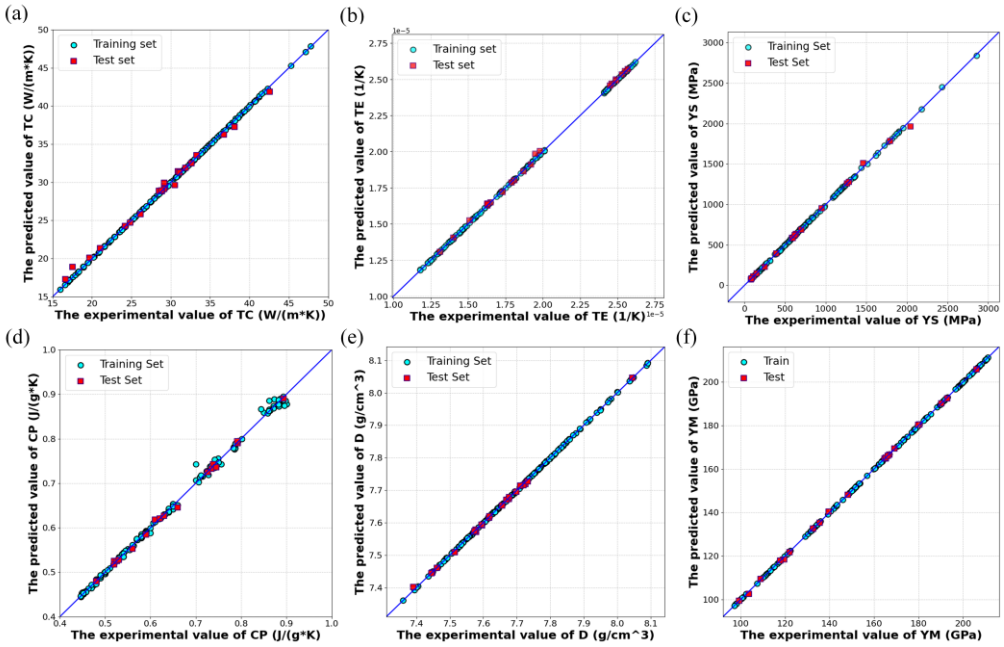

Figure 4. Scatter plot comparing the machine learning predicted and experimental values of six steel material parameters

https://doi.org/10.64486/m.65.4.5

Table 2. The range of grid search values

Number	Parameters	Hyperparameter Value
1	Learning rate	[0.01,0.1,0.5]
2	max_depth	[3,4,5,6,8]
3	n_estimators	[100,500,900]
4	Subsample	[0.5,0.7,1.0]
5	Colsample_bytree	[0.3,0.4,0.6]
6	Gamma	[0, 1, 4]

3. Results

3.1. Model training and evaluation

Five representative regression models were selected to evaluate their predictive performance in multi-target estimation of steel parameters: RF, GBR, XGBoost, Lasso regression, and Ridge regression. These models were chosen because of their proven ability to handle nonlinear relationships, complex feature interactions, and high-dimensional data structures commonly encountered in materials informatics [25[26]. Their basic principles and typical application scenarios are summarised in Table 3.

Model performance was quantitatively assessed using mean absolute error (MAE), root mean square error (RMSE), and the coefficient of determination (R^2), as defined in Eqs. (2)–(4) [27]. MAE measures the average magnitude of prediction errors, RMSE emphasises large deviations, and R^2 describes the proportion of variance in the target variable that is explained by the model.

$$E_{MAE} = \frac{1}{n}\sum_{i=1}^{n}| y_i - y_i | \tag{2}$$

$$E_{EMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i)^2} \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

The comparison of model performance is shown in Figure 5. For TC, XGBoost achieves the highest accuracy, with test-set MAE and RMSE reduced by 22 % and 20 %, respectively, compared with the second-best algorithm. This confirms its strong ability to capture the nonlinear mapping of thermal transport behaviour. XGBoost also provides the best prediction performance for CP. For YS, TE, YM, and D, GBR delivers the lowest errors on the test set, demonstrating accurate representation of plastic deformation onset, thermal expansion behaviour, elastic stiffness, and density. For YM, GBR reduces the prediction error by nearly half compared with RF, and for D, it exhibits highly consistent errors between training and test sets, indicating good generalisation.

3.2. Model robustness stress test

To further verify the reliability of the proposed model under variations in data distribution and to enhance confidence in its industrial applicability, systematic robustness stress testing was conducted. By progressively reducing the training set ratio from 90 % to 70 % and 50 %, the degradation of model performance with decreasing training data volume was quantitatively evaluated. In parallel, based on the complete dataset of 1290 samples, training subsets of different sizes were generated through subsampling, and learning curves were constructed to analyze the convergence behavior of model performance as the available data increased.

https://doi.org/10.64486/m.65.4.5

The stress-test results are summarized in Table 4. When the training ratio was reduced from 9:1 to 7:3, the average performance degradation across all six target parameters, as measured by the coefficient of determination (R^2), remained below 2 %, indicating that the model exhibits strong robustness to variations in training data volume. The learning curves for CP, shown in Figure 6, reveal that model performance improves rapidly with increasing training sample size and stabilizes once the training data reach approximately 40 % of the total dataset. This behavior suggests that the current dataset size is sufficient to support effective model learning, and that further data expansion leads only to marginal performance gains. Although a slight performance degradation was observed for parameters with higher data heterogeneity, such as TE, when the training data were reduced, the absolute prediction errors remained within acceptable limits for engineering applications.

Overall, the robustness stress testing demonstrates that the developed model is insensitive to variations in training set size and exhibits excellent generalization capability, indicating that its high predictive performance is not a result of overfitting.

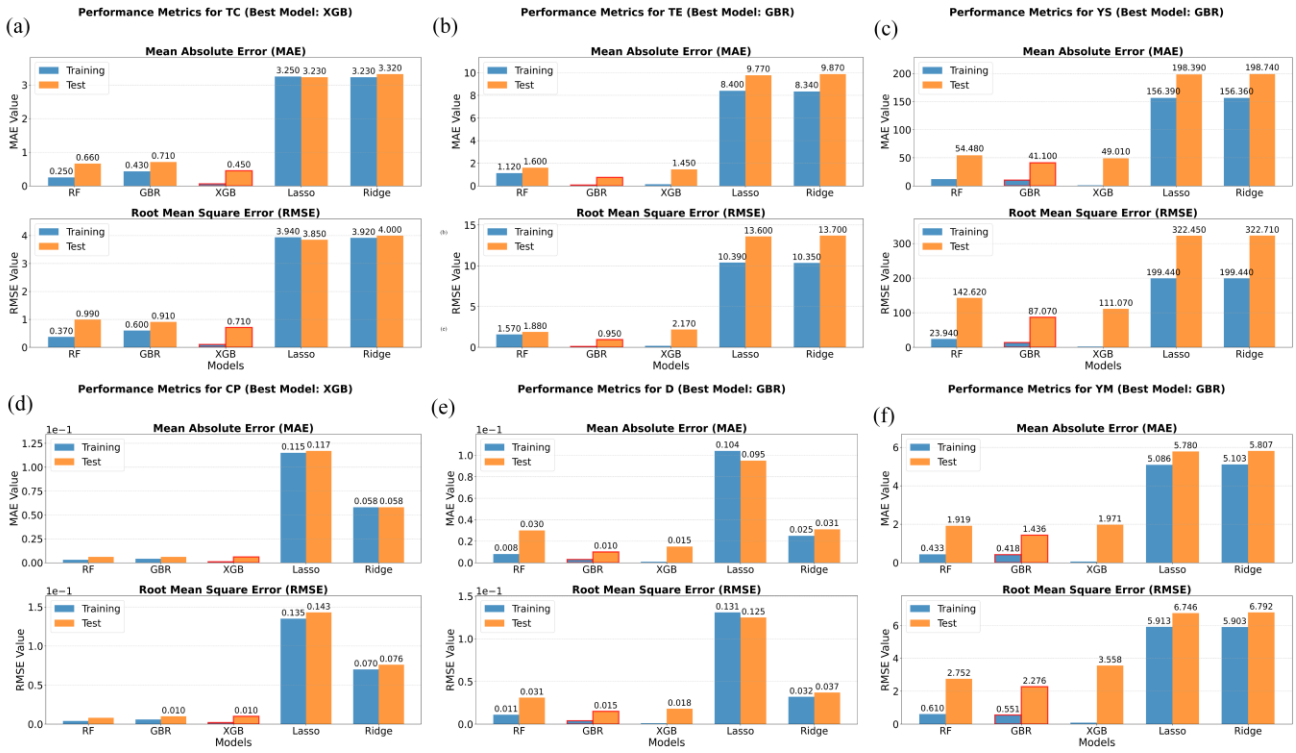

Figure 5. Comparison with the Initial Model. (a) Comparison of TC; (b) Comparison of TE; (c) Comparison of YS; (d) Comparison of CP;(e) Comparison of D; (f) Comparison of YM

3.3. Feature importance analysis by SHAP

To further elucidate the decision-making mechanisms of the models and identify the dominant factors governing each target property, SHAP (SHapley Additive exPlanations) analysis was employed [28]. SHAP is based on cooperative game theory and quantifies the contribution of each input feature to the model output, enabling interpretation of nonlinear relationships and feature interactions.

The SHAP analysis results are summarized in Figure 7. For TC, Fe content and temperature exhibit the widest SHAP value ranges and the highest data density, indicating their dominant influence. This observation is consistent with the well-established dependence of heat conduction in steels on lattice structure and electron–phonon interactions, both of which are strongly affected by alloy composition and thermal state. For TE, temperature is the primary contributing factor, reflecting its direct role in atomic vibration and lattice expansion, while C and other alloying elements provide secondary contributions by modifying the crystal lattice and bonding characteristics. Regarding YS, C and temperature show the largest SHAP value ranges, highlighting their critical roles in governing plastic deformation behavior. This aligns with the known strengthening effect of

https://doi.org/10.64486/m.65.4.5

carbon through solid solution and microstructural refinement, as well as the temperature sensitivity of dislocation motion. Alloying elements such as Cr and Mo exhibit secondary but non-negligible effects, consistent with their roles in precipitation strengthening and solid-solution hardening. For CP, temperature remains the most influential feature, which is expected given the intrinsic temperature dependence of heat capacity in solids. The SHAP distributions of Si and C further suggest their influence on vibrational modes and bonding environments, thereby affecting energy storage behavior. In the case of D, temperature, C, Si, and Ni all contribute noticeably, reflecting the combined effects of thermal expansion and compositional differences in atomic mass and lattice packing. For YM, Fe shows a strong and concentrated SHAP distribution, confirming its primary role as the matrix element controlling elastic stiffness in steels. Additional contributions from Si, Cr, and C are observed within specific SHAP intervals, indicating their influence on interatomic bonding strength and elastic response through alloying-induced lattice modifications.

Overall, although the dominant features vary among individual target properties, temperature, Fe, C, and Si consistently emerge as key determinants across multiple predictions. By linking the SHAP-derived feature importance with established metallurgical and thermophysical principles, these results enhance the interpretability of the machine learning models and demonstrate that the learned relationships are physically meaningful, thereby supporting their applicability in alloy design and process optimization.

Table 3. The specific principles and applicability of each model

Model	Algorithm principle
Random Forest (RF)	An ensemble learning framework that constructs multiple decision trees and aggregates their predictions (Bootstrap Aggregating, or Bagging). Each tree is trained on a random subset of features, reducing variance and mitigating overfitting.
Gradient Boosting Regression (GBR)	A sequential ensemble method that iteratively trains weak learners (typically decision trees) to fit residual errors, optimizing the loss function via gradient descent.
eXtreme Gradient Boosting (XGB)	An optimized Gradient Boosting algorithm incorporating L_1/L_2 *regularization terms* to control model complexity. It supports parallel processing and handles missing values efficiently. Enhanced accuracy is achieved through a second-order Taylor expansion of the loss function.
Lasso Regression (Lasso)	A linear model employing L_1 *regularization* (penalty term: $\|\beta\|$), which induces sparsity in feature weights to perform automatic feature selection.
Ridge Regression (Ridge)	A linear model utilizing L_2 *regularization* (penalty term: β^2) to address multicollinearity, shrinking coefficients without inducing sparsity.

Table 4. The R^2 scores under different proportions

Parameters	R^2 (9:1)	R^2 (8:2)	R^2 (7:3)
TC	0.9929	0.9871	0.9837
TE	0.9972	0.9812	0.9707
YS	0.9986	0.9912	0.9841
CP	0.9978	0.9936	0.9842
D	0.9981	0.9903	0.9793
YM	0.9928	0.9876	0.9810

https://doi.org/10.64486/m.65.4.5

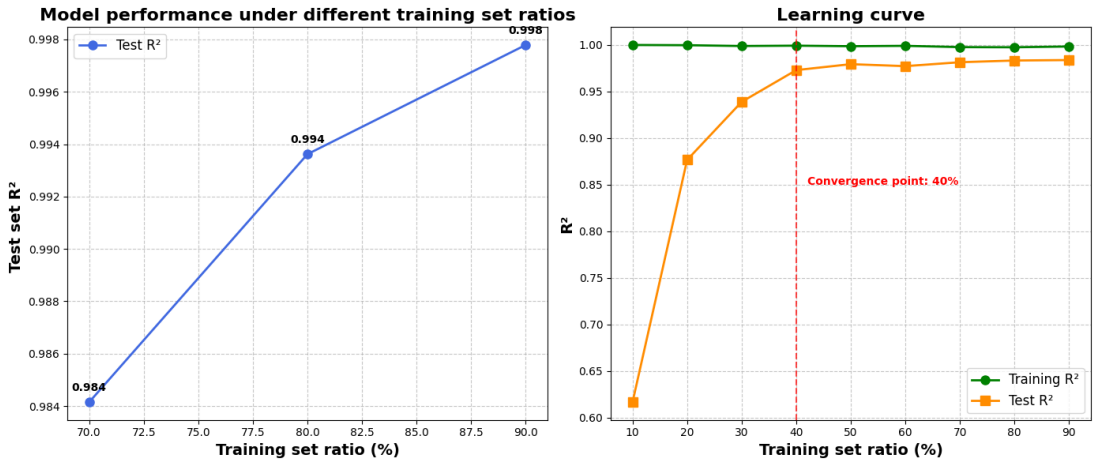

Figure 6. The learning curve of CP

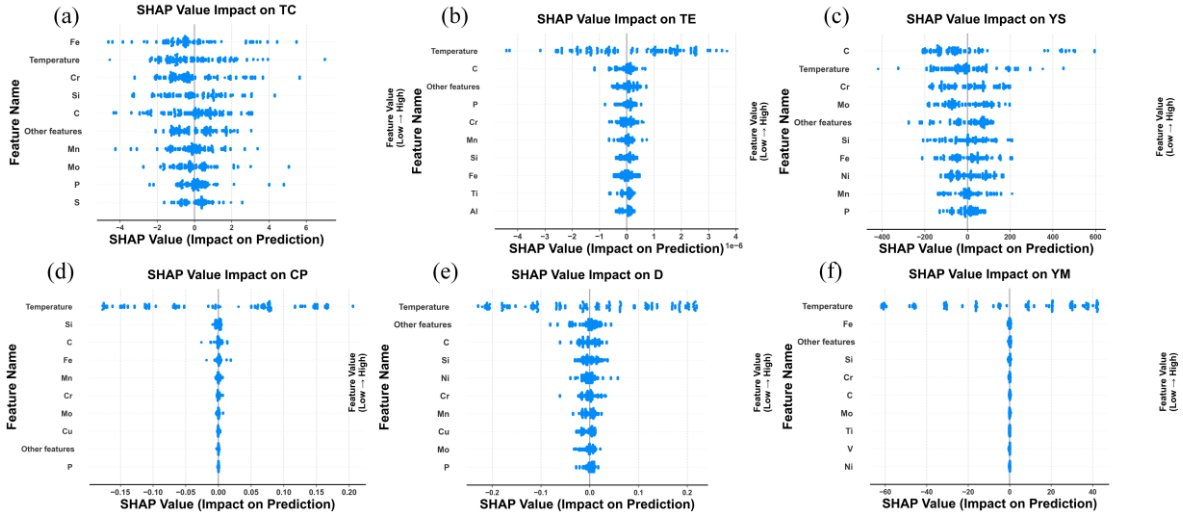

Figure 7. SHAP analysis results

3.4. Independent verification

To further evaluate the generalisation ability of the models, an independent validation set consisting of 20 newly collected samples for each target parameter—completely unseen during training and hyperparameter optimisation—was used [29]. Table 5 lists representative prediction results compared with experimental values.

The relative deviations between predicted and experimental data for all six parameters are controlled within 4.64 %. The maximum error for TC is less than 0.05 %, and the errors for TE and YS are within 1.90 % and 2.72 %, respectively. Although CP shows slightly larger discrepancies for a few cases, its overall relative error remains below 5 %. The prediction errors for D and YM do not exceed 0.3 % and 1.5 %, respectively. These results confirm that the optimised models have strong generalisation capability and high predictive accuracy for unseen data.

4. Database development and application

To support the deployment and application of the prediction models, a dedicated steel material database management system (S-MDMS) was designed and implemented. The system adopts a front-end/back-end decoupled architecture that integrates data management, model application, and result visualisation into a unified platform.

https://doi.org/10.64486/m.65.4.5

As illustrated in Figure 8, the front-end is developed using the Vue.js framework with the Element UI component library, providing an intuitive and responsive user interface. Data visualisation is implemented through ECharts, enabling interactive exploration of parameter trends and distributions. The back-end is built on the Spring Boot framework and provides RESTful API services. MyBatis is used as the persistence layer, and MySQL serves as the core database for data storage and retrieval. The main functional modules include user management, steel material management, and material parameter management.

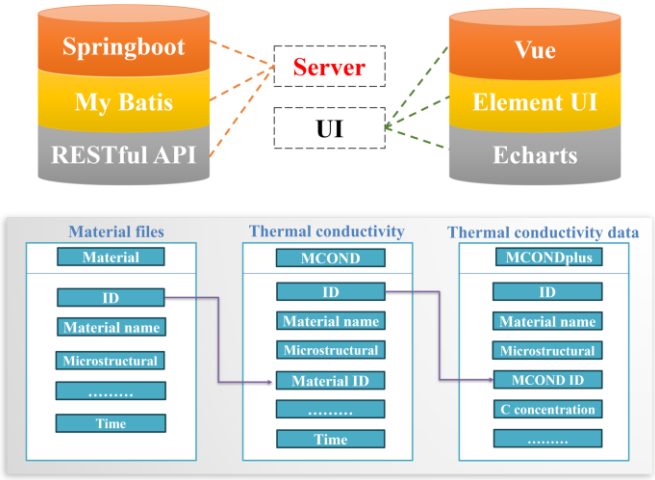

Figure 8. Database construction technology and thermal conductivity E-R diagram

The database archives well-structured datasets for typical steel grades, including chemical composition, thermophysical and mechanical properties, and phase transformation characteristics. Users can perform multi-dimensional queries, add or delete records in batches, and export datasets. The optimised ML models for the six target properties are integrated into the system, enabling online prediction based on user-defined compositions, microstructures, and temperatures. The predicted results are automatically stored in the database and can be directly used as input for FEA simulations. An example interface for thermal conductivity is shown in Figure 9.

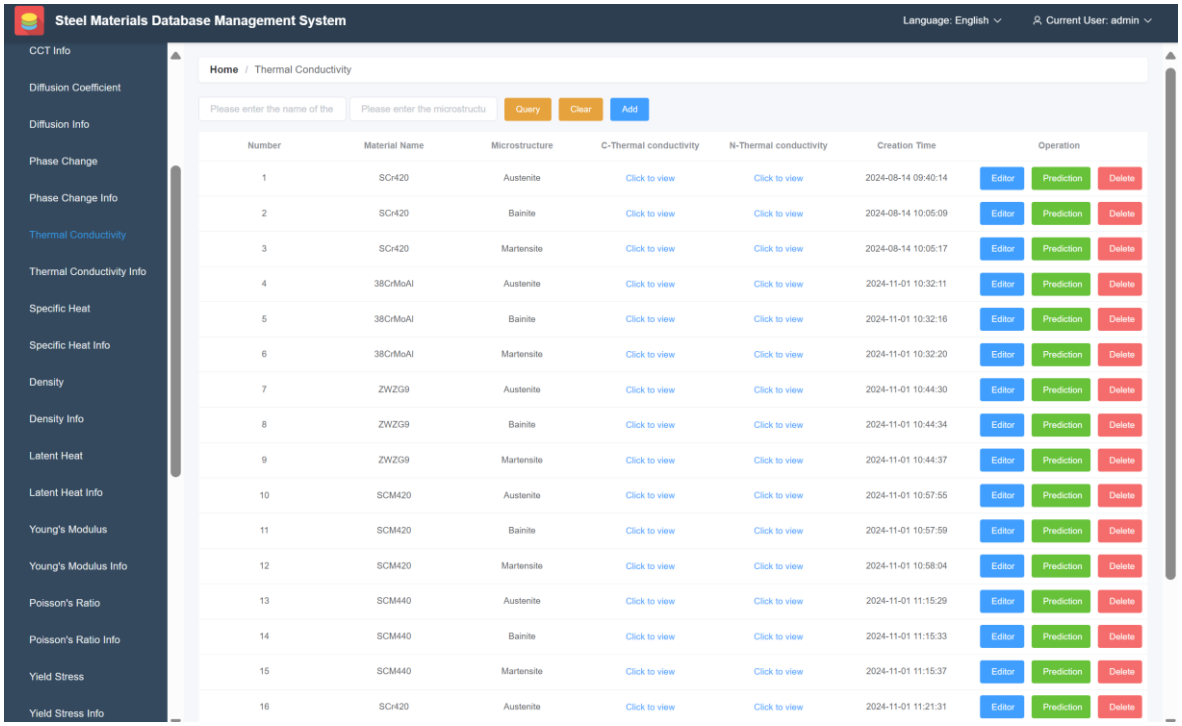

Figure 9. The thermal conductivity interface in the database system

https://doi.org/10.64486/m.65.4.5

Overall, the S-MDMS provides standardised management of steel material data and efficient retrieval and application of parameters, forming a robust infrastructure for digital steel design and intelligent heat treatment optimisation. In addition, the system was designed with a strong emphasis on scalability and continuous model updating. Dedicated interfaces for model management and version control are reserved, enabling online incremental incorporation of new experimental data and automated retraining of the models. Role-based access control and operation log tracing are also implemented to ensure the security and traceability of research data. These features provide a unified supporting platform for subsequent extension to a wider range of material systems and the integration of advanced algorithms such as reinforcement learning and Bayesian optimization.

Table 5. Comparison of experimental values and predicted values of different parameters

Parameter	Experimental value	Predicted value	Error %
TC/(W/mK)	18.91	18.9149	0.025
	20.26	20.2662	0.030
	21.57	21.5602	0.045
	22.85	22.8519	0.008
	24.12	24.1184	0.007
TE/(1/K)	2.47E-05	2.4722E-05	0.09
	2.48E-05	2.4758E-05	0.17
	1.29E-05	1.2660E-05	1.90
	1.51E-05	1.4907E-05	1.30
	1.86E-05	1.8913E-05	1.70
YS/MPa	124.591	123.1835	1.14
	167.930	164.5794	2.04
	192.886	191.4838	0.73
	809.845	788.3683	2.72
	820.455	825.7999	0.65
CP/(J/(g*K))	0.5211	0.4985	4.53
	0.5300	0.5279	0.39
	0.5610	0.5361	4.64
	0.6840	0.6560	4.27
	0.7825	0.7859	0.01
D/(g/cm^3)	7.4210	7.4039	0.23
	7.5787	7.5676	0.15
	7.6253	7.6149	0.14
	7.7287	7.7071	0.28
	8.0692	8.0848	0.19
YM/GPa	163.4464	164.5710	0.69
	167.6519	165.8493	1.09
	186.8945	184.1480	1.49
	191.5521	189.6902	0.98
	205.8271	204.9733	0.42

https://doi.org/10.64486/m.65.4.5

5. Conclusions

In this study, a data-driven framework for predicting steel material parameters and an integrated database management system were developed. The main conclusions are:

(1) A comprehensive dataset was constructed by combining simulation data and experimental measurements, covering 18 elemental compositions, three typical microstructures, and a wide temperature range. Six key parameters—TC, CP, YS, TE, YM, and D—were selected as target variables. Five machine learning models were evaluated. XGBoost exhibited the best performance for thermal parameters, while GBR achieved the highest accuracy for mechanical properties. After grid-search-based hyperparameter optimisation and ten-fold cross-validation, all models achieved R^2 values above 0.99 and relative errors within 5 %.

(2) SHAP analysis revealed the key features controlling each property and showed that temperature, Fe, C, and Si are consistently important across multiple targets. This improves the interpretability of the models and provides guidance for alloy design and process control. An independent validation set confirmed the strong generalisation ability of the optimised models, with all relative errors maintained within 4.64 %.

(3) A steel material database management system (S-MDMS) was implemented to combine data storage, visualisation, and online prediction. The system provides an efficient tool for rapid acquisition and application of steel parameters in FEA-based heat treatment design and process optimisation.

Although the current framework shows high accuracy and practical applicability, several limitations remain. The binary encoding of microstructures cannot fully describe microstructural heterogeneity, and the ML models do not explicitly incorporate physical mechanisms. Future work will focus on integrating more detailed microstructural descriptors and coupling the data-driven models with mechanistic or physics-informed approaches to further improve interpretability and predictive reliability.

Acknowledgments: This work is supported by the Key project of China Coal Technology Engineering Group (2025-TD-CXY002).

References

[1] D. Zhang, X. Zhao, L. He, J. Ren, L. Zhang, and Y. Zhou, "Calibration and verification of dynamic mechanical properties of high-strength armored steel based on Johnson-Cook Constitutive Model," *Acta Armamentarii*, no. 8, pp. 0-43, 2022. (in Chinese). [Online]. Available: https://pubs.cstam.org.cn/article/id/6670e16aca140b49b448b189

[2] L. Li, C. Zhou, L. Liu, Q. Wang, J. Li, and J. Xing, "Research on the effect of thermo-physical parameters on HRB4-00 on numerical simulation of solidification and heat transfer in continuous billet casting," *Continuous Casting*, no. 5, pp. 14-20, 2019, https://doi.org/10.13228/j.boyuan.issn1005-4006.20190027

[3] W. Han, H. Bao, and X Ruan, "Perspective: Predicting and optimizing thermal transport properties with machine learning methods," *Energy and AI*, vol. 8, pp. 100153, 2022, https://doi.org/10.1016/j.egyai.2022.100153

[4] J. Schmidt, M.R.G. Marques, S. Botti, and M.A.L. Marques, "Recent advances and applications of machine learning in solid-state materials science," *Npj Computational Materials*, vol. 5, no. 1, pp.83, 2019, https://doi.org/10.1038/s41524-019-0221-0

[5] Z. Wang, Z. Sun, H. Yin, X. Liu, J. Wang, H. Zhao, C.Pang, T. Wu, S. Li, and X. Yu, "Data-driven materials innovation and applications," *Advanced Materials*, vol. 34, no.36, pp. 2104113, 2022, https://doi.org/10.1002/adma.202104113

[6] E. O. Pyzer-Knapp, M. Manica, P. Staar, L. Morin, P. Ruch, T. Laino, J. R. Smith, and A. Curioni, "Foundation models for materials discovery–current state and future directions," *Npj Computational Materials*, vol. 11, no.1, pp. 61, 2025, https://doi.org/10.1038/s41524-025-01538-0

[7] D. Jha, V. Gupta, L. Ward, Z. Yang, C. Wolverton, I. Foster, and A. Agrawal, "Enabling deeper learning on big data for materials informatics applications," *Scientific Reports*, vol. 11, no.1, pp. 4244, 2021, https://doi.org/10.1038/s41598-021-83193-1

https://doi.org/10.64486/m.65.4.5

[8] A. Agrawal, and A. Choudhary, "Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science," *APL Materials*, vol. 4, no.5, 2016, https://doi.org/10.1063/1.4946894

[9] S. M. Moosavi, and B. Smit, "The role of machine learning in the understanding and design of materials," *Journal of the American Chemical Society*, vol. 142, no. 48, pp. 20273-20287, 2020, https://doi.org/10.1021/jacs.0c09105

[10] D. Nath, D. R. Neog, and S. S. Gautam, "Application of machine learning and deep learning in finite element analysis: a comprehensive review," *Archives of Computational Methods in Engineering*, vol. 31, no. 5, pp. 2945-2984, 2024, https://doi.org/10.1007/s11831-024-10063-0

[11] Y. Liu, W. Zheng, H. Ai, H. Zhou, L. Feng, and L. Cheng, "Application of machine learning in predicting the thermal conductivity of single-filler polymer composites," *Materials Today Communications*, vol. 39, pp. 109116, 2024, https://doi.org/10.1016/j.mtcomm.2024.109116

[12] W. Q. Shen, Y. J. Cao, and Z. B. Liu, "Prediction of plastic yield surface for porous materials by a machine learning approach," *Materials Today Communications*, vol. 25, pp. 101477, 2020, https://doi.org/10.1016/j.mtcomm.2020.101477

[13] K. Yang, G. Yisimayili, and J. Yue, "Prediction of mechanical properties of zinc alloy based on machine learning algorithm." *Metalurgija*, vol. 65, no. 1, pp. 37-44, 2026, https://doi.org/10.64486/m.65.1.4

[14] Li, Z. X., Z. P. Liu, and J. T. Yu. "Prediction of carbon black/carbon nanotube reinforced polydimethylsiloxane properties based on BP neural network." *Metalurgija*, vol.64, no. 1-2, pp. 72-74, 2025, https://hrcak.srce.hr/319864

[15] J. Xiong, T. Zhang, and S. Shi, "Machine learning of mechanical properties of steels," *Science China Technological Sciences*, vol. 63, no. 7, pp. 1247-1255, 2020, https://doi.org/10.1007/s11431-020-1599-5

[16] U. Bhandari, Y. Chen, H. Ding, S. Emanet, P. R. Gradl, and S. Guo, "Machine-learning-based thermal conductivity prediction for additively manufactured alloys," *Journal of Manufacturing and Materials Processing*, vol. 7, no. 5, pp. 160, 2023, https://doi.org/10.3390/jmmp7050160

[17] C. Niu, S. Li, Y. Dan, Z. Cao, X. Li, and J. Hu, "Thermal conductivity prediction based on machine learning," *New Chemical Materials*, vol. 48, no. 3, pp. 134-137, 2020, (in Chinese). [Online]. Available: https://doi.org/10.19817/j.cnki.issn1006-3536.2020.03.031

[18] J. Wang, Y. Fa, Y. Tian, and X Yu, "A machine-learning approach to predict creep properties of Cr–Mo steel with time-temperature parameters," *Journal of Materials Research and Technology*, vol. 13, pp. 635-650, 2021, https://doi.org/10.1016/j.jmrt.2021.04.079

[19] J. F. Durodola, "Machine learning for design, phase transformation and mechanical properties of alloys," P*rogress in Materials Science*, vol. 123, pp. 100797, 2022, https://doi.org/10.1016/j.pmatsci.2021.100797

[20] L. Zhang, Z. Wang, J. Zhao, K. An, J. Xu, and J. Gu, "Design and implementation of materials database for heat treatment process simulation," *Heat Treatment of Metals*, vol. 48, no. 9, pp. 247-252, 2023, https://doi.org/10.13251/j.issn.0254-6051.2023.09.042

[21] L. Yang, H. Su, F. Cai, X. Luo, and L. Duan, "Material database and application status of data mining technology," *Materials China*, vol. 38, no. 7, pp. 672-681+650, 2019, (in Chinese).

[22] H. Liu, H. Zhang, P. Wei, C. Jia, and Y. Li, "Design and development of high performance gear transmission database software," *Computer Integrated Manufacturing Systems*, vol. 29, no. 8, pp. 2513-2523, 2023, https://doi.org/10.13196/j.cims.2023.08.001

[23] D. Leni, "Prediction modeling of low alloy steel based on chemical composition and heat treatment using artificial neural network," *Jurnal Polimesin*, vol. 21, no. 5, pp. 530-537, 2023, http://e-jurnal.pnl.ac.id/polimesin

[24] F. Jirasek, and H. Hasse, "Perspective: machine learning of thermophysical properties," *Fluid Phase Equilibria*, vol. 549, no. 7-8, pp. 11320, 2021, https://doi.org/10.1016/j.fluid.2021.113206

[25] R. Wu, L. Zeng, J. Fan, Z. Peng, and Y. Zhao, "Composition, heat treatment, microstructure and loading condition based machine learning prediction of creep life of superalloys," *Mechanics of Materials*, vol. 187, pp. 104819, 2023, https://doi.org/10.1016/j.mechmat.2023.104819

[26] M. Li, L. Dai, and Y. Hu, "Machine learning for harnessing thermal energy: From materials discovery to system optimization," *ACS Energy Letters*, vol. 7, no. 10, pp. 3204-3226, 2022, https://doi.org/10.1021/acsenergylett.2c01836

[27] D. Zhu, B. Wang, H. Zhao, S. Wu, F. Li, S. Huang, H. Wu, S. Wang, C. Zhang, J. Gao, and X. Mao, "Enhanced hardenability prediction in 20CrMo special steel via XGBoo-st model," *Journal of Iron and Steel Research International*, vol. 32, no. 4, pp. 1023-1033, 2025, https://doi.org/10.1007/s42243-025-01461-0

[28] S. M. LaValle, M. S. Branicky, and S. R. Lindemann, "On the relationship between classical grid search and probabilistic roadmaps," *The International Journal of Robotics Research*, vol. 23, no. 7-8, pp. 673-692, 2004, https://doi.org/10.1177/0278364904045481

[29] S. M. Lundberg, and S. I. Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems*, vol. 30, 2017, https://doi.org/10.48550/arXiv.1705.07874