ISSN (Print) 1226-1750
ISSN (Online) 2233-6656
Journal of Magnetics 31(2), 153-164 (2026) https://doi.org/10.4283/JMAG.2026.31.2.153

Prediction of Magnetic Properties of Nd-Fe-B Permanent Magnets Using Ensemble Machine Learning Models

Chunghee Nam*

Department of Electrical and Electronic Engineering, Hannam University, Daejeon 34430, Republic of Korea

(Received 15 April 2026, Received in final form 4 June 2026, Accepted 4 June 2026)

This study predicts the key magnetic properties of Nd-Fe-B permanent magnets, coercivity (H_{cj}) and remanence (B_r), using a dataset of 635 compositions and processing parameters. Three ensemble models (Random Forest, XGBoost, and LightGBM) were evaluated, with LightGBM demonstrating the highest predictive performance ($R^2 \approx 0.95$ for both properties). Explainable machine learning analysis using SHAP confirmed that composition-based features heavily influence B_r, whereas processing parameters such as sintering time dictate H_{cj}. Model validation against experimental data showed low errors in interpolation regions but limited extrapolation capability. Overall, this study clarifies the complex correlations between composition, processing parameters, and magnetic performance, demonstrating the potential of explainable machine learning for designing next-generation, high-performance permanent magnets.

Keywords : ensemble model, machine learning, Nd-Fe-B permanent magnet

1. Introduction

Permanent magnet materials have long been studied as essential components in drive motors for electric and hybrid vehicles, as well as in various generators including wind turbines [1, 2]. The development of permanent magnet materials has progressed over more than a century, with numerous materials having been introduced. Since the 1980s, NdFeB-based materials have attracted considerable attention due to their high energy density, and various new technologies have been introduced in the areas of alloy element additions and processing development [3–5]. Since both high remanence and high coercivity are simultaneously required to increase energy density, continuous improvement of these two properties has been achieved through compositional modification and microstructural control via processing parameter optimization [3–5]. Demand for motor components considering miniaturization, weight reduction, and cost efficiency is expected to continue growing across applications such as unmanned aerial vehicles, next-generation mobility platforms, maritime transportation, and robotics systems, ensuring that the development of high-performance permanent magnets will remain an active area of research [6].

NdFeB permanent magnets have been developed in two main categories: sintered magnets and hot-deformed magnets. Sintered magnets are more widely adopted across industrial sectors owing to their high energy density and ease of processing into various geometries. Despite their superior energy density, NdFeB-based materials suffer from poor thermal stability, and research aimed at enhancing their thermal stability has attracted increasing attention. Since coercivity decreases sharply at elevated temperatures, the addition of heavy rare-earth elements such as Tb and Dy is the most widely studied approach for coercivity enhancement [5, 7, 8]. In particular, grain boundary diffusion has emerged as an efficient method for improving coercivity with heavy rare-earth elements, given their high raw material cost. However, this approach has limitations including diffusion depth constraints and supply chain risks. Research on substituting Fe with elements such as Zr, Nb, Al, Cu, and Ga is also ongoing to improve remanence and enhance the stability of NdFeB-based permanent magnets [9]. More recently, as rare-earth materials have become strategically important, light rare-earth elements including Ce, Pr, and La are being utilized as part of efforts to reduce rare-earth consumption, and this area is expected to see substantial research activity in future material development.

*Corresponding author: Tel: +82-42-629-7472
e-mail: chnam@hnu.ac.kr

While remanence is closely related to the elemental composition of the material, coercivity is known to be significantly influenced by both processing conditions and microstructural features. In particular, there exists a well-known challenge referred to as Brown's Paradox: while theoretical calculations predict that high coercivity values can be achieved at grain sizes of several hundred nanometers, experimentally observed coercivity values reach only approximately 20% of the theoretical limit [10]. This discrepancy arises from domain wall behavior in actual microstructures, including nucleation phenomena and weak-link effects. To better understand these issues, new material designs and previously unreported microstructural characteristics have been investigated through micromagnetic simulations using synthetic data generated from three-dimensional structural models that closely replicate actual microstructures [10, 11]. Consequently, designing high-performance NdFeB-based permanent magnets requires not only the optimization of diverse compositional elements and their relative fractions, but also the concurrent optimization of processing conditions. However, given the vast compositional space and the large number of possible combinations of discrete experimental parameters, optimization via conventional trial-and-error approaches inevitably incurs significant time and economic costs.

To address these limitations, machine learning-based approaches for data-driven materials design have been applied across various material systems. These approaches are particularly advantageous for materials with complex compositions, and the field has seen especially active development in areas such as the design of high-entropy alloys and ceramics with superior mechanical properties, and bulk metallic glasses with excellent soft magnetic characteristics [12-14]. In the context of NdFeB materials, recent studies have applied machine learning to predict magnetic properties. These studies can be broadly categorized into two types: those that build machine learning models from literature-collected datasets [15], and those that supplement the literature data with additionally processed information [16]. Furthermore, a case study on process optimization through active learning in the fabrication of NdFeB materials via direct extrusion was reported by the NIMS group in Japan [17].

In the present study, we systematically investigated the effects of compositional features and processing parameters on the magnetic properties of NdFeB-based permanent magnets using statistical data analysis based on the publicly available dataset [16]. Optimization of LightGBM (LGBM), one of the ensemble machine learning models, yielded predictive performance superior to that reported in previous studies [16]. Unlike Random Forest (RF)—a bagging-based algorithm in which multiple decision trees are trained independently on bootstrap-sampled subsets of the data and their predictions are averaged to produce the final output—LGBM belongs to the gradient boosting family, wherein each successive tree is trained to minimize the residual errors of the preceding ensemble. Furthermore, while XGBoost (XGB) employs a level-wise tree growth strategy that splits all nodes at the same depth before proceeding, LGBM adopts a leaf-wise growth strategy that preferentially expands the leaf with the greatest loss reduction, enabling faster convergence and higher predictive accuracy for complex, nonlinear datasets. This sequential learning strategy enables the model to progressively correct prediction errors, thereby achieving higher predictive accuracy than RF for data exhibiting complex nonlinear relationships. In addition, SHAP (SHapley Additive exPlanations) analysis was employed as an explainable machine learning method to quantify the contributions of each compositional feature and processing parameters to the model output. Through this analysis, the effects of rare-earth and transition metal elements—beyond the primary constituents of NdFeB—as well as the influences of magnetic field intensity and heat treatment duration, were elucidated. Furthermore, the trained model was applied to predict the magnetic properties of recently reported NdFeB-based permanent magnets, demonstrating the potential of the machine learning framework as a reliable tool for evaluating the performance of novel materials.

2. Methods

2.1. Dataset

The dataset used for predicting the magnetic properties of NdFeB-based permanent magnets by machine learning includes experimentally measured coercivity and remanence values, as well as the corresponding processing parameters, and was adopted from a previously reported dataset [16]. While many literature sources provide compositional information and magnetic properties, processing parameters are frequently not reported in detail. Following the approach of the aforementioned study [16], unspecified processing conditions were replaced with standard values commonly used in the field, resulting in a dataset of 635 data entries.

Figure 1 shows the distribution of property for 635 NdFeB-based permanent magnet materials used in this study. Fig. 1(a) illustrates the relationship between remanence (B_r) and coercivity (H_{cj}), which exhibits a trade-off characteristic typical of NdFeB permanent magnet

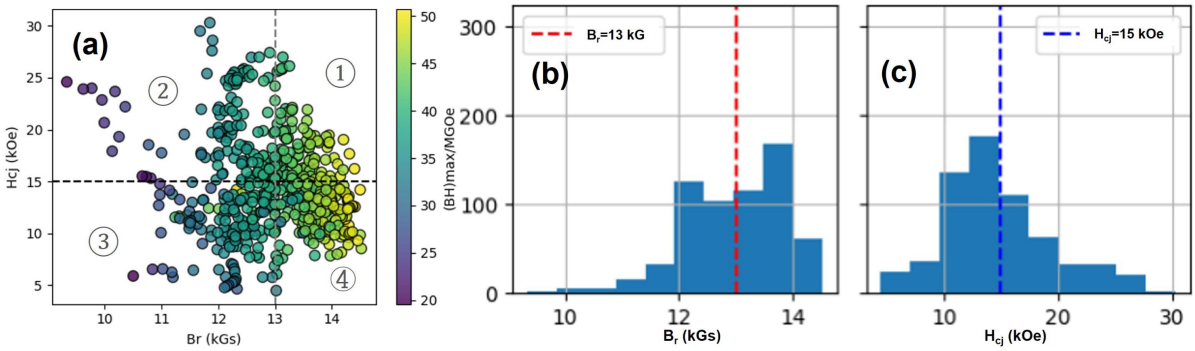

Fig. 1. (Color online) (a) Scatter plot of remanence versus coercivity. Histograms of (b) remanence and (c) coercivity for 635 Nd-Fe-B permanent magnet alloys.

alloys—indicating that simultaneous optimization of both properties through materials and process design is necessary. In Fig. 1(a), threshold values were defined for both coercivity and remanence to partition the data into four distinct regions. Region ① represents materials with both high remanence and high coercivity; Region ② contains materials with high coercivity but remanence below the threshold; Region ③ includes materials where both properties fall below their respective thresholds; and Region ④ encompasses materials with high remanence but coercivity below the threshold. The color mapping by maximum energy product $(BH)_{max}$ in Fig. 1(a) further reinforces the need to optimize both properties simultaneously: materials with higher remanence tend to exhibit higher $(BH)_{max}$, whereas materials with high coercivity do not necessarily achieve high $(BH)_{max}$ values, and thus populate Region ②. This trend is clearly illustrated in Fig. 1(a), where a trade-off relationship is observed between the two properties. Specifically, materials with high coercivity (H_{cj} ≥15 kOe) but lower remanence populate Region ②, whereas those with high remanence (B_r ≥13 kG) but lower coercivity are distributed in Region ④. The color mapping of the maximum energy product $(BH)_{max}$ further demonstrates that strategies employed to increase coercivity often lead to a reduction in remanence, thereby limiting the simultaneous optimization of both properties.

Figures 1(b) and 1(c) present histograms of remanence and coercivity, respectively. The dashed lines in each panel denote the threshold values established as references for relative performance: 15 kOe for coercivity and 13 kG for remanence. The dashed lines in each panel denote the threshold values established as references for relative performance: 15 kOe for coercivity and 13 kG for remanence. These specific thresholds were determined based on the statistical characteristics of the 635-sample dataset. The B_r threshold (13.0 kG) aligns almost exactly with the dataset's median (13.04 kG) and mean (12.97 kG), effectively bisecting the population. Meanwhile, the H_{cj} threshold (15.0 kOe) sits close to the dataset's mean (14.74 kOe), representing the top 38.7% of the samples. This asymmetrical distribution reflects the inherent materials-processing difficulty in achieving high coercivity compared to remanence. The histogram for remanence confirms that a high proportion of the dataset exceeds the threshold, reflecting the extensive research efforts directed toward optimizing composition and elemental ratios. In contrast, the histogram for coercivity in Fig. 1(c) shows a relatively lower proportion of materials exceeding the threshold, consistent with the fact that coercivity is sensitive to not only composition but also processing parameters.

Figure 2 presents the compositional distribution of the dataset, organized by individual elements. The distributions of the primary elements Nd, Fe, and B are shown in Fig. 2(a). The distributions of seven rare-earth elements added to enhance permanent magnet performance are shown in Fig. 2(b), while the distributions of other metallic elements are presented in Fig. 2(c). The category of other elements includes transition metals, post-transition metals, and Mg from the alkali metals. As shown in the inset of Fig. 2(c), cobalt (Co) was used over a relatively broad range up to 12.97 wt%, with a more uniform distribution compared to the other metallic elements. According to previously reported literature, Co has been extensively studied as a substitute for Fe, with research demonstrating its effects on mechanical properties, remanence enhancement, and $(BH)_{max}$ improvement through microstructural control [18, 19]. As shown in Fig. 2(c), Nb and Zr were also used at relatively high proportions among the other metallic elements. Nb has been reported to suppress the reduction in remanence while simultaneously increasing coercivity when used in combination with Dy [20]. This is consistent with findings that Nb exhibits the most pronounced coercivity enhancement

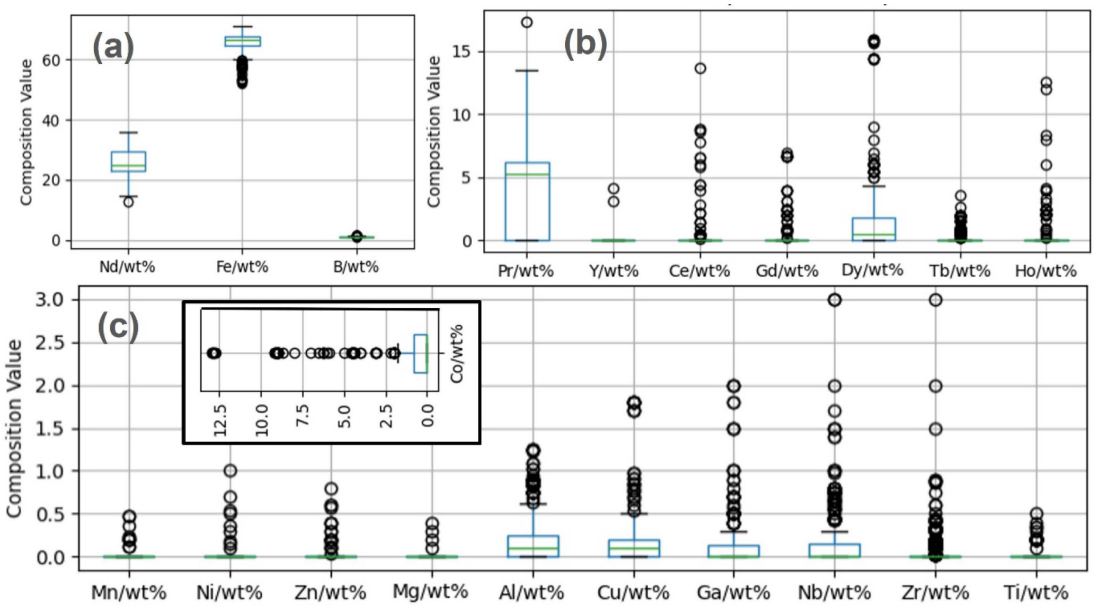

Fig. 2. (Color online) Box plots represent the wt% distribution of each element: (a) major elements Nd, Fe, and B; (b) seven rare-earth metals; and (c) eleven other metallic elements.

among four elements—Al, Cu, Ga, and Nb—added to sintered NdFeB magnets [21]. Additional recent studies have indicated that Nb also plays an important role in stabilizing the Nd-rich phase, improving corrosion resistance, and maintaining magnetic properties through defect reduction [22].

Figure 3 presents the compositional characteristics of the 112 materials belonging to Region ① in Fig. 1(a), which exhibit both high remanence and high coercivity. As shown in Fig. 3(a, b), the combined content of the primary elements Nd, Fe, and B was predominantly distributed in the range of approximately 90–94 wt%, indicating that approximately 6–10 wt% of various additive elements were utilized to optimize both magnetic properties. The addition ratios and distributions of rare-earth and other metallic elements within this subset of 112 materials are shown in Figs. 3(c) and 3(d), respectively. In contrast to the overall dataset (Fig. 2), the proportion of Dy is noticeably reduced in Fig. 3(c) for the rare-earth elements. Although Dy enhances coercivity, its substantial adverse effect on remanence requires additional compensating strategies for materials to qualify for Region ①. Fig. 3(d) reveals marked differences in the proportions of Co, Nb, and Zr for Region ① compared to the overall dataset: Co is present at concentrations below 5 wt%, while Zr and Nb are present at below 2 wt%.

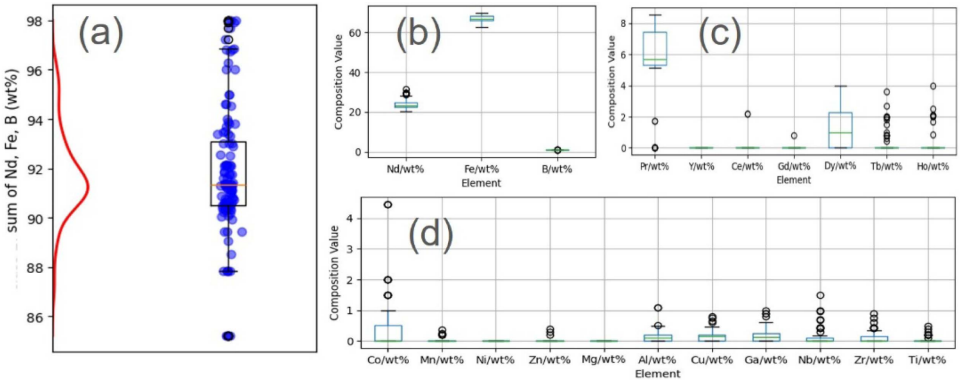

Fig. 3. (Color online) Box plots of the wt% composition from 112 Nd-Fe-B alloy data corresponding to Region ① in Fig. 1(a): (a) distribution of the total content of the major elements Nd, Fe, and B; (b) individual distributions of Nd, Fe, and B; (c) seven rare-earth metals; and (d) eleven other metallic elements.

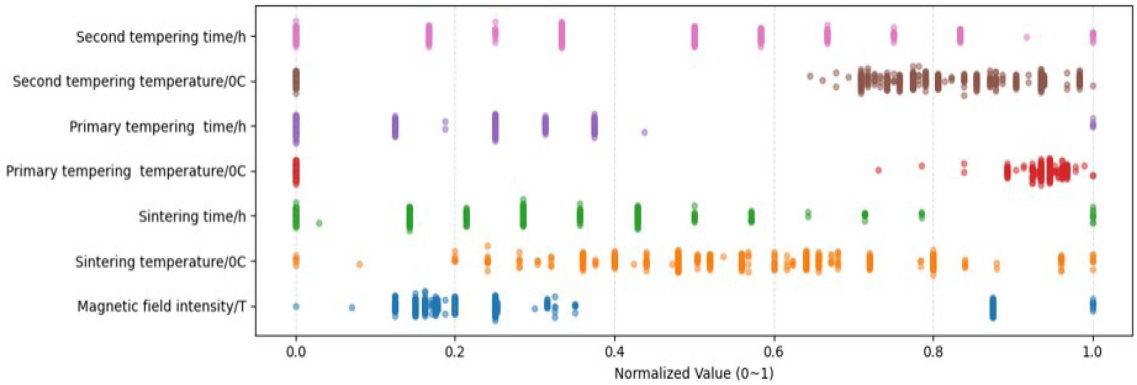

Fig. 4. (Color online) Discrete distributions of seven processing parameters from the entire dataset, normalized to the range [0, 1] using min-max scaling.

Figure 4 presents a visualization of the distribution of seven experimental processing parameters across the entire dataset. Each condition was normalized to the range [0, 1] using min-max scaling and plotted as individual data points along a single axis. The results confirm that the magnetic field strength, sintering temperature, sintering time, and both primary and secondary annealing temperatures and durations were set across a wide range of values. Conditions with a normalized value of zero in Fig. 4 indicate that the corresponding process, such as magnetic field intesnity, was not applied. The distributions demonstrate that the experimental conditions are represented as discrete values spanning a broad parameter space, indicating that the experimental design was configured to encompass a comprehensive range of processing parameters. Fig. 4 provides a meaningful reference for interpreting the influence of individual processing parameters in the explainable machine learning analysis, which is discussed in further detail in the Results and Discussion section.

2.2. Machine Learning Models

To precisely predict the coercivity and remanence of NdFeB-based permanent magnets, this study evaluates and compares three widely adopted ensemble machine learning models: Random Forest (RF), XGBoost (XGB), and LightGBM (LGBM). As outlined in the Introduction, these models utilize fundamentally distinct architectural strategies—ranging from RF's independent tree bagging to the sequential gradient boosting of XGB and LGBM—to capture highly complex and nonlinear relationships. In this work, all three algorithms were trained using the identical dataset of 635 entries, establishing a rigorous benchmark to evaluate how their structural differences affect predictive performance and computational efficiency specifically within the domain of permanent magnet composition and processing parameters [23, 24].

2.3. Machine Learning Model Optimization

To develop regression prediction models for coercivity

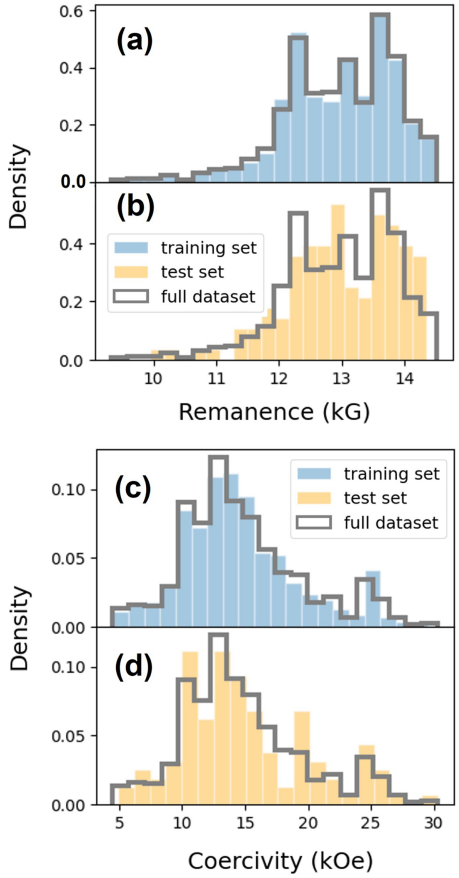

Fig. 5. (Color online) Comparison of data distributions: (a) training data versus the full dataset for remanence, (b) test data versus the full dataset for remanence, (c) training data versus the full dataset for coercivity, and (d) test data versus the full dataset for coercivity.

and remanence of the 635 collected Nd-Fe-B permanent magnet alloy compositions and processing parameters, the entire dataset was partitioned into training and test sets at an 8:2 ratio, and predictive performance was evaluated based on results on the test set. To facilitate comparison with the overall data distribution, the distributions of both target properties—coercivity and remanence—in the training and test sets are presented in Fig. 5. Fig. 5 displays these distributions as histograms relative to the full dataset: Figs. 5(a) and 5(b) correspond to remanence, while Figs. 5(c) and 5(d) correspond to coercivity. Both the training and test sets exhibit a broadly uniform distribution, which helps prevent biased predictive performance arising from overrepresentation of specific value ranges. The evaluation metrics used were the R^2 score and root-mean-squared error (RMSE), both of which are commonly employed in regression prediction models. During training, 10-fold cross-validation was applied to minimize data dependency while comparing model performance. Hyperparameter tuning for each model was performed using grid search, with the optimized parameters summarized in Table 1.

Table 1. Optimized hyperparameters of the machine learning models.

LGBM	RF	XGB
colsample_bytree (1.0)	max_depth (11)	colsample_bytree (0.8)
learning_rate (0.2)	max_features (15)	gamma (0)
max_depth (11)	min_samples_leaf (1)	learning_rate (0.1)
n_estimators (200)	min_samples_split (2)	max_depth (5)
reg_alpha (0)	n_estimators (300)	n_estimators (200)
reg_lambda (10)		reg_alpha (0)
subsample (0.8)		reg_lambda (1)
		subsample (0.8)

3. Results and Discussion

t-distributed Stochastic Neighbor Embedding (t-SNE) is a nonlinear dimensionality reduction algorithm widely used to project high-dimensional materials data into a lower-dimensional space, enabling effective visualization of latent clusters and patterns within the data. This technique is particularly useful for representing the complex, nonlinear relationships among alloy compositions and processing parameters as two-dimensional scatter plots, thereby facilitating intuitive exploration and identification of regions associated with superior magnetic properties such as coercivity and remanence. Fig. 6 presents the results of t-SNE dimensionality reduction applied to the entire dataset, projected onto a two-dimensional space. Figs. 6(a) and 6(b) are scatter plots in which each data point represents an individual material, color-mapped according to the magnitude of remanence and coercivity, respectively. The two figures together reconfirm the trade-off relationship between coercivity and remanence: as shown in Fig. 6(a), two regions with low remanence (indicated by arrows) correspond to materials with high coercivity, while in Fig. 6(b), two regions with low coercivity (indicated by arrows) correspond to materials with high remanence. Statistical analysis of the dataset confirms the competing relationship between these two properties, underscoring the importance of a thorough understanding of composition-based knowledge of materials data for the effective application of machine learning to enhance the energy density of NdFeB-based permanent magnets.

Figure 7 presents radar plots of the rare-earth metal and other metallic element compositions (in wt%) for the four material groups classified according to the B_r–H_{cj} plot in Fig. 1(a). Group ① contains materials exhibiting both

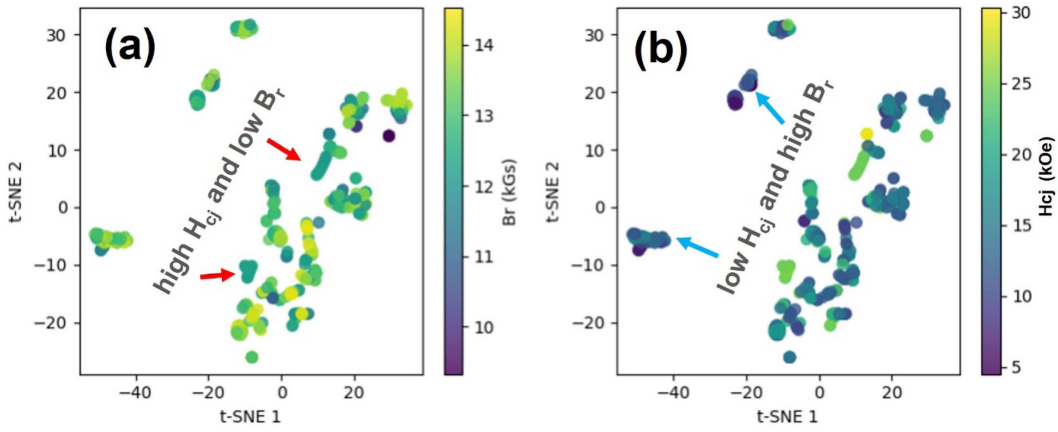

Fig. 6. (Color online) t-SNE scatter plots of the input dataset after dimensionality reduction: (a) colored by remanence and (b) colored by coercivity.

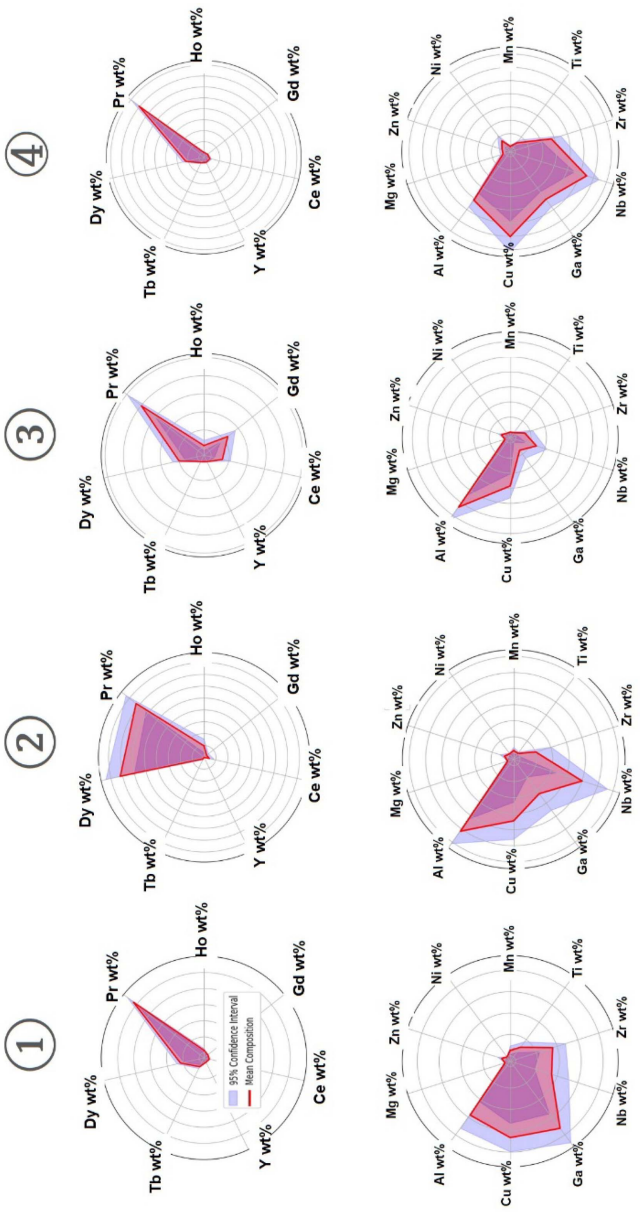

Fig. 7. (Color online) For the four regions in Fig. 1(a): the upper panels show the wt% distributions of rare-earth elements, while the lower panels present radar plots of the wt% distributions of ten other metallic elements excluding Co.

high coercivity and high remanence; Group ② includes materials with coercivity above the threshold but remanence below it; Group ③ encompasses materials where both properties fall below their respective thresholds; and Group ④ consists of materials with remanence above the threshold but coercivity below it. The light rare-earth element Pr was used at similar proportions across all four groups, reflecting the high demand for research on light rare-earth additions as substitutes for heavy rare-earth elements to reduce their overall consumption. In Group ②, a substantially higher proportion of the heavy rare-earth element Dy was used for coercivity enhancement. However, the Dy-induced reduction in remanence remains a challenge that must be addressed. The lower panels of Figure 7 show the elemental compositions of other metallic additions for the four groups. In Group ①, Al, Cu, and Ga were used in balanced proportions, resulting in superior performance in both coercivity and remanence. As previously reported, remanence is important for energy output in permanent magnet design, while coercivity is critical for thermal stability; optimizing both properties simultaneously is therefore essential. For such trade-off relationships, a recent study has reported the application of heuristic optimization algorithms to the design of ReFeB (Re = Pr, Nd, La, Ce) permanent magnets when sufficient data are available [25].

To analyze the trends of individual elements in influencing coercivity and remanence in NdFeB-based permanent magnets, the relative usage frequency, average remanence, and average coercivity for all 21 elemental features were visualized and are presented in Fig. 8. In Fig. 8, the bar chart represents the frequency of each element appearing in the alloy compositions within the dataset; the blue line (circle symbols) indicates the average remanence of alloys containing that element, and the red line (square symbols) represents their average coercivity. Excluding the primary elements Nd, Fe, and B, Dy exhibits a high frequency of occurrence and a pronounced association with high coercivity, consistent with the extensive literature demonstrating that Dy, as a heavy rare-earth element, enhances coercivity through grain boundary diffusion. In contrast, Tb and Ho appear with lower frequencies but show a tendency to contribute to high coercivity. This is consistent with reported findings that the combined addition of Tb and Ho stabilizes the Nd-rich phase, thereby enhancing coercivity and simultaneously improving corrosion resistance [26]. Mg, while used at a low frequency, shows a notable contribution to remanence enhancement, consistent with reports that the addition of MgO at grain boundaries leads to increased remanence [27]. These findings confirm the presence of meaningful compositional trends with respect to magnetic properties and validate the utility of composition-based features as inputs for machine learning-based prediction of permanent magnet performance. The following sections therefore describe the construction of predictive models based on the given dataset and the application of explainable machine learning methods to guide novel material design.

21 elemental features in Nd-Fe-B alloys, highlighting their respective influences on magnetic properties.

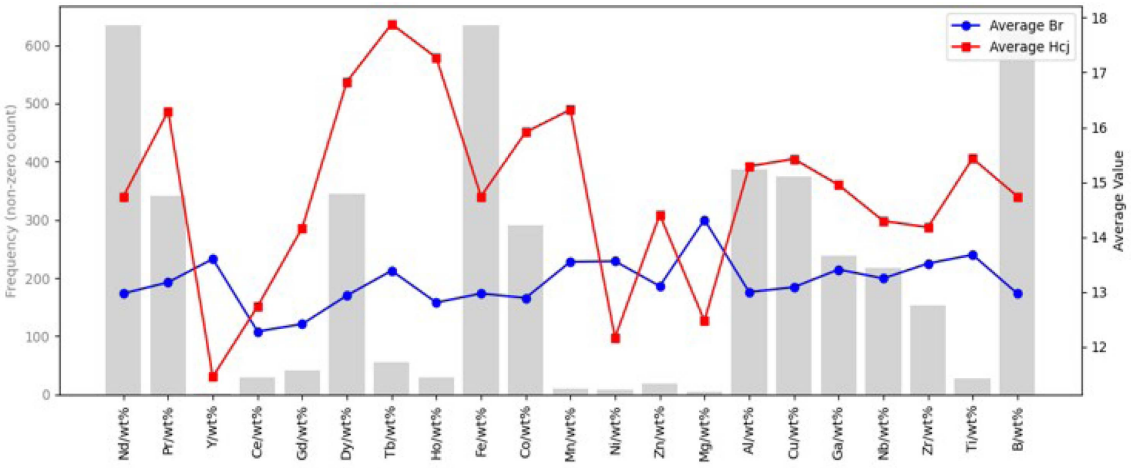

Fig. 8. (Color online) Visualization of the relative frequency, average remanence (blue line with circle symbols), and average coercivity (red line with square symbols) for 21 elemental features in Nd-Fe-B alloys, highlighting their respective influences on magnetic properties.

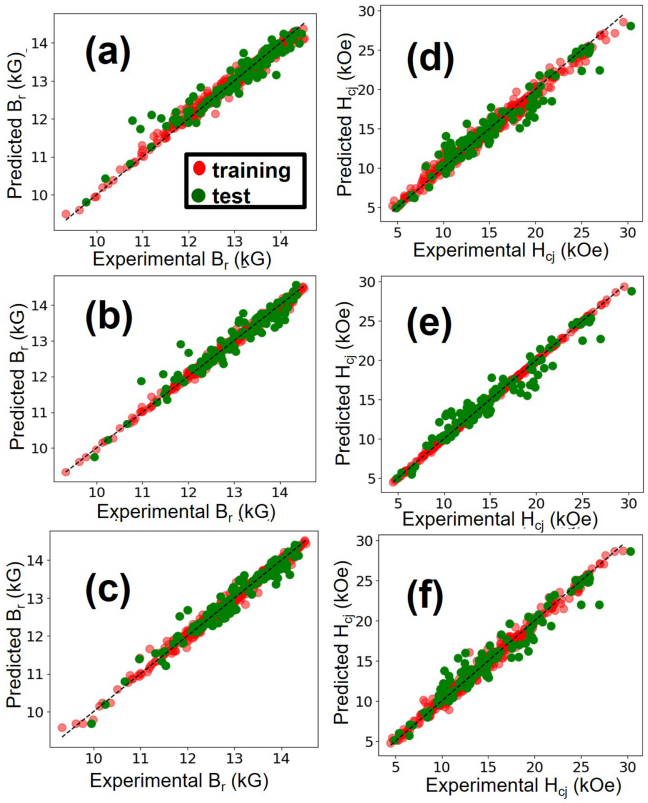

Fig. 9. (Color online) Parity plots of experimental versus predicted values for the optimized machine learning models: (a–c) results of RF, XGB, and LGBM for remanence prediction; (d–f) results of RF, XGB, and LGBM for coercivity prediction. The diagonal dashed line indicates perfect agreement between experimental and predicted values.

Figure 9 presents the regression performance plots (parity plots) of the three optimized machine learning models for coercivity and remanence prediction. In each plot, the x-axis represents the experimentally measured values and the y-axis represents the model-predicted values. Figs. 9(a)–9(c) show the results for remanence prediction using the RF, XGB, and LGBM models, respectively, while Figs. 9(d)–9(f) show the corresponding results for coercivity prediction. The predictive performance metrics are summarized in Table 2. The diagonal dashed line in each plot represents perfect agreement between experimental and predicted values; data points located closer to this line indicate superior predictive accuracy. For remanence, the XGB model achieved an R^2 score close to unity on the training data, whereas the R^2 score on the test data was 0.923, suggesting a tendency toward overfitting in which the model is overly tuned to the training set at the expense of generalizability. Overall, LGBM achieved the highest predictive performance for remanence prediction, with an R^2 score of 0.949 and an RMSE of 0.202 kG. For coercivity prediction, XGB again exhibited signs of overfitting, while LGBM demonstrated the best predictive performance, with an R^2 score of 0.949 and an RMSE of 1.173 kOe.

Table 2. Performance metrics (R^2 and RMSE) of the machine learning models.

	R^2 Score		RMSE	
	B_r	H_{cj}	B_r (kG)	H_{cj} (kOe)
LGBM	0.949	0.949	0.202	1.173
RF	0.923	0.937	0.256	1.300
XGB	0.928	0.944	0.240	1.225

Since machine learning is often regarded as a black-box

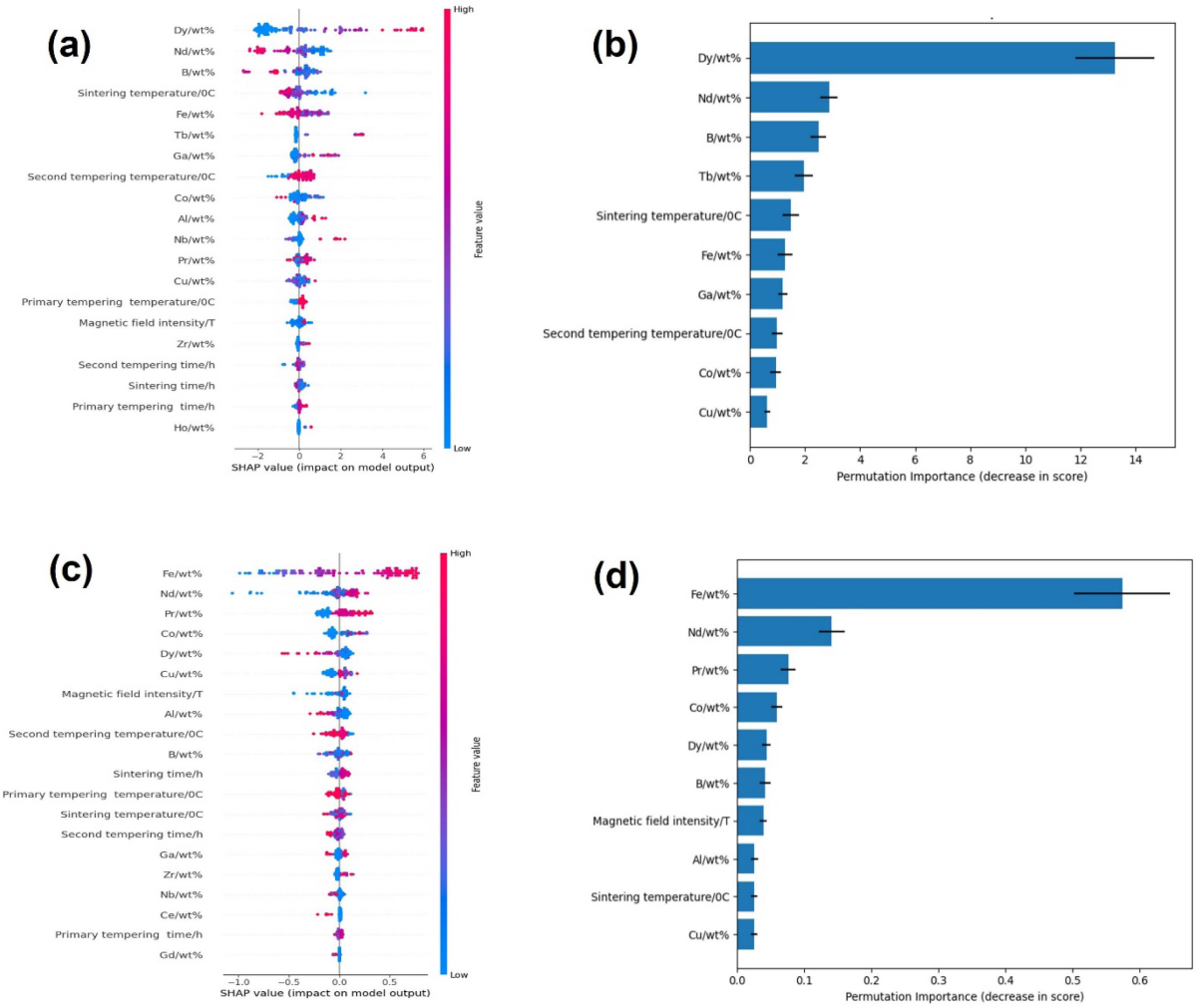

Fig. 10. (Color online) Explainable machine learning analysis of the optimized LGBM model: (a) SHAP values for coercivity, identifying Dy and sintering time as the most influential features; (b) permutation feature importance for coercivity, confirming Dy as the dominant factor; (c) SHAP values for remanence, highlighting Pr and Co as key contributors with Dy showing a negative effect; and (d) permutation feature importance for remanence, consistent with the SHAP analysis, identifying Pr, Co, and Dy as the major influencing factors beyond the primary elements Nd, Fe, and B.

system, the development of explainable AI approaches has emerged as a critical strategy for materials design based on machine learning. In the present study, the LGBM model—which demonstrated the highest predictive performance among the tree-based ensemble methods—was selected, and SHAP-based explainable machine learning analysis was applied to assess the efficiency of composition- and process-guided materials design. SHAP analysis was conducted on the test dataset to determine which input features most significantly influence the model output. Separate SHAP analyses were performed for the optimized LGBM models predicting coercivity and remanence, and the results are presented in Fig. 10. The effects of rare-earth elements, other metallic elements, and processing parameters were examined separately, excluding the primary constituents Nd, Fe, and B. In the SHAP plots, the x-axis represents the magnitude of each feature's contribution to the model output (i.e., coercivity or remanence): positive values indicate a positive relationship (increasing the output), while negative values indicate an inverse relationship (decreasing the output). Features are ranked in descending order of importance from top to bottom, and the color of each data point reflects the magnitude of the corresponding feature value, ranging from blue (low) to red (high), as indicated by the color bar in Fig. 10.

As shown in Fig. 10(a), Dy was identified as the most influential feature for coercivity prediction, consistent with well-established findings that increasing Dy substitution for Nd in NdFeB permanent magnets leads to

enhanced coercivity. Among the processing parameters, sintering time was found to be an important factor influencing coercivity. The SHAP analysis reveals an inverse relationship between sintering time and coercivity: prolonged sintering promotes grain growth, which shifts the dominant demagnetization mechanism toward domain wall motion-driven magnetization reversal, resulting in a reduction in coercivity [28]. Tb, another heavy rare-earth element, exhibits a similar effect to Dy; however, since its frequency of occurrence in the present dataset is relatively low, the generality of its influence cannot be definitively established. To cross-validate the SHAP analysis results, permutation feature importance was also evaluated. In this approach, each feature is individually selected, its values are randomly shuffled, and the resulting change in predictive performance is used to assess its importance: features that are more important will lead to a larger performance drop upon shuffling, while unimportant features will exhibit little change. Fig. 10(b) presents the permutation feature importance for coercivity in descending order, confirming that the Dy composition ratio is the most important feature. The overall ranking is consistent with the SHAP results, supporting the validity of the SHAP analysis.

Figure 10(c) presents the SHAP analysis results for remanence prediction. Among the rare-earth elements, Pr was identified as the most influential contributor. The analysis confirms a positive effect of increasing Pr content (as a substitute for Nd) on remanence, consistent with the extensive literature on Pr as a representative light rare-earth alternative to heavy rare-earth elements. The transition metal Co content was also identified as an important feature for remanence, in agreement with recent studies reporting that the combined addition of Pr and Co promotes the stability of remanence [29, 30]. In contrast, Dy—while positively influencing coercivity—exhibits a negative effect on remanence, consistent with the well-documented antiferromagnetic coupling between Dy and transition metals that leads to reduced remanence [31]. Among the processing parameters, the alignment magnetic field intensity was identified as the most influential parameter, with higher field strengths during the compaction stage showing a positive effect on remanence by enhancing grain orientation. Permutation feature importance for remanence is shown in Fig. 10(d). Excluding the primary elements Nd, Fe, and B, Pr, Co, and Dy were identified as the most important features, in agreement with the SHAP analysis.

Finally, the optimized LGBM model was applied to predict coercivity and remanence for NdFeB-based alloy compositions not included in the 635 alloys dataset, and the predictions were compared against experimentally measured values from recent literature [32]. The external validation dataset comprised 24 material entries: 5 materials (unseen 1) with compositional distributions similar to those of the training and test data, corresponding to interpolation predictions, and 19 materials (unseen 2) with compositions outside the distribution of the training data, corresponding to extrapolation predictions. The compositional distributions of the interpolation and extrapolation datasets are provided in Fig. 11(a).

For the 5 interpolation materials (unseen 1), the prediction errors for both coercivity and remanence are presented in Fig. 11(b). Remanence predictions exhibited

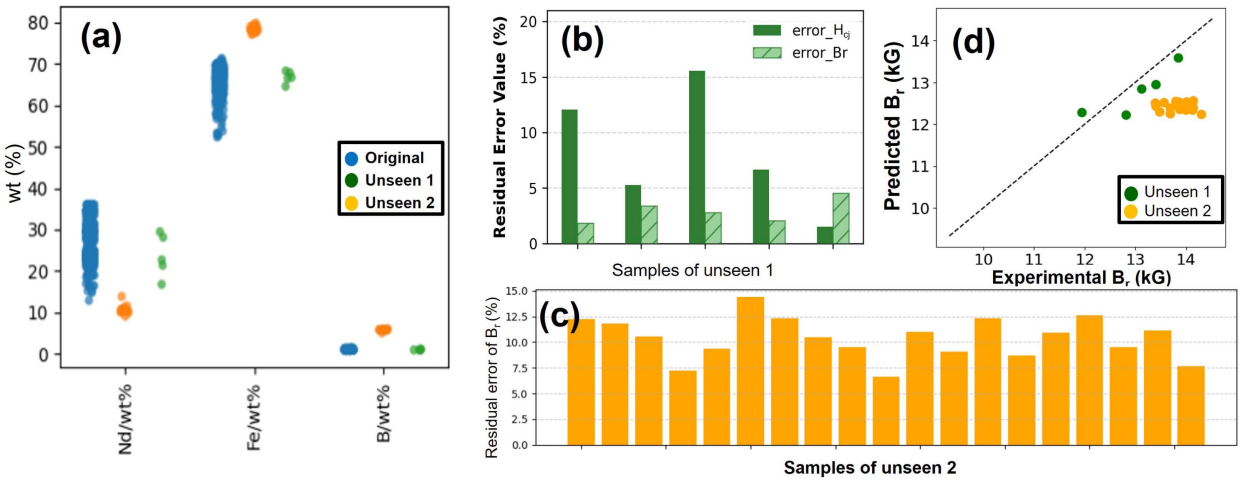

Fig. 11. (Color online) Validation of the optimized LGBM model using 24 newly collected Nd-Fe-B alloy compositions: (a) compositional distributions of interpolation prediction data (unseen 1) and extrapolation prediction data (unseen 2); (b) prediction errors of coercivity and remanence for unseen 1 alloys; (c) comparison of predicted versus experimental remanence values for unseen 2 alloys; and (d) B_r prediction performance comparison between the two unseen datasets.

low errors within 5%, whereas coercivity errors ranged from 1.5% to 15.5%. This larger variability in coercivity prediction can be attributed to the relatively limited number of process-related features (7 features) compared to the 21 composition-based features used as model inputs, combined with the known strong dependence of coercivity on processing parameters. For the 19 extrapolation materials (unseen 2), only remanence predictions were evaluated, as experimental coercivity values were not reported in the referenced literature. The predicted versus experimental remanence values are shown in Fig. 11(c). Prediction errors for remanence ranged from 6.6% to 14.3%, which are notably larger than those observed for the interpolation dataset. This indicates that the trained model encounters difficulty in predicting the performance of new materials whose compositions lie outside the training data distribution. Fig. 11(d) provides a comprehensive comparison of the remanence prediction performance for both external validation datasets, clearly demonstrating that the model performs well for compositions within the training distribution, while performance degrades for out-of-distribution compositions. The larger prediction errors for extrapolation compositions underscore the importance of domain knowledge and data coverage. Since 21 composition-based features were used as model inputs, the model's remanence predictions are primarily driven by compositional information, and predictive performance decreases for compositions not represented in the training data. Continuously expanding the NdFeB dataset to improve predictive performance, as well as designing and optimizing novel compositions through active learning and generative AI approaches, remain important directions for future work.

4. Summary

In this study, a systematic statistical analysis and a comparative evaluation of three ensemble machine learning models were conducted based on a comprehensive dataset of compositions and processing parameters, with the aim of precisely predicting the coercivity and remanence of Nd-Fe-B permanent magnet materials. Data analysis revealed a clear trade-off between these two key properties of Nd-Fe-B materials; specifically, coercivity was found to be strongly influenced by processing parameters, whereas remanence was closely correlated with composition-based features. Among the models evaluated, LightGBM (LGBM) demonstrated the highest predictive performance, achieving an R^2 value of approximately 0.95 for both coercivity and remanence. Furthermore, SHAP analysis, an explainable machine learning approach, reconfirmed that heavy rare-earth elements such as Dy and Tb play critical roles in coercivity enhancement, while elements such as Pr and Co are important contributors to remanence improvement. Additionally, processing parameters including sintering time and alignment magnetic field intensity were shown to be closely linked to microstructural control and to exert significant influences on the predicted magnetic properties.

Model validation using 24 external data points sourced from recent literature demonstrated low prediction errors for interpolation compositions similar to the training distribution, while relatively larger errors were observed for extrapolation compositions. These results emphasize the importance of both training data coverage and composition-based domain knowledge, suggesting that data augmentation, active learning, and generative AI-based material discovery are essential for improving model reliability and generalizability. An inherent limitation of the current framework is the absence of explicit, quantitative microstructural descriptors, such as grain size distribution, grain boundary phase continuity, and oxygen content, which directly dictate coercivity mechanisms. While processing parameters serve as indirect indicators of microstructural evolution in the current model, incorporating high-fidelity microstructural descriptors in future datasets will be vital to significantly improving prediction reliability, particularly for highly structure-sensitive properties like coercivity.

Overall, this study systematically elucidates the correlations between compositional and processing parameters and the magnetic properties of Nd-Fe-B permanent magnets, presenting a robust framework for prediction and explainability based on machine learning. This approach provides an effective data-driven strategy for the design of next-generation, high-performance permanent magnet materials. In particular, the combined use of rare-earth element reduction and substitution, processing parameter optimization, and explainable AI analysis can serve as a key methodology for accelerating the development of next-generation permanent magnets with high energy density and improved thermal stability.

Authorship contribution statement

Chunhee Nam: Conceptualization, Funding acquisition, Methodology, Investigation, Data curation, Software, Validation, Visualization, Writing – original draft.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea

government (MSIT)(RS-2025-25464308).

Data Availability. The data supporting this study's findings are available from the corresponding author upon reasonable request.

Statements and Declarations

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] D. Brown, B.-M. Ma, and Z. Chen, J. Magn. Magn. Mater. **248**, 432 (2002).
[2] Q. Lu, Y. Shao, Y. Yin, H. Chen, H. Xu, W. Liu, M. Liu, C. Zhong, X. Yu, J. Chen, Y. Liu, X. Yi, and M. Y, Sustain. Mater. Technol. **36**, e00615 (2023).
[3] C. Zhao, Q. Jiang, S. Ur Rehman, X. Li, Y. Chen, J. Song, and Z. Zhong, J. Rare Earth. **40**, 1894 (2022).
[4] S.-M. Lee, G. Kim, K.-S. Lee, S. Kim, T.-H. Kim, S.-H. Lee, D.-H. Kim, W. Lee, and J.-G. Lee, Acta Mater. **285**, 120660 (2025).
[5] T.-H. Kim, T.T. Sasaki, T. Ohkubo, Y. Takada, A. Kato, Y. Kaneko, and K. Hono, Acta Mater. **172**, 139 (2019).
[6] Y. Wang, H. Qin, N. Liu, Q.-N. Hu, H.-P. Cong, and S.-H. Yu, Adv. Mater. **37**, 2505193 (2025).
[7] J. Yan, Q. Wu, H. Ge, and Z. Xi, J. Liang, Mater. Chem. Phys. **315**, 128901 (2024).
[8] M. Ji, Z. Wang, W. Liu, Y. Li, H. Chen, H. Wu, R. Du, D. Zhang, M. Yue, X. Yi, Y. Liu, and S. Zha, J. Alloys Compd. **970**, 172598 (2024).
[9] Y. G. Ma, R. S. Li, Z. Yang, M. Matsumoto, A. Morisako, and S. Takei, Mater. Sci. Eng. B **117**, 287 (2005).
[10] A. Bolyachkin, E. Dengina, N. Kulesh, X. Tang, H. S.-A., T. Ohkubo, and K. Hono, npj Comput. Mater. **10**, 34 (2024).
[11] A. Kovacs, J. Fischbacher, H. Oezelt, A. Kornell, Q. Ali, M. Gusenbauer, M. Yano, N. Sakuma, A. Kinoshita, T. Shoji, A. Kato, Y. Hong, S. Grenier, T. Devillers, N. M. Dempsey, T. Fukushima, H. Akai, N. Kawashima, T. Miyake, and T. Schrefl, Front. Mater. **9**, 2022 (2023).
[12] J. He, Z. Li, J. Lin, P. Zhao, H. Zhang, F. Zhang, L. Wang, and X. Cheng, Mater. Des. **246,** 113326 (2024).
[13] C. Nam, Mater. Chem. Phys. **315**, 129076 (2024).
[14] M. Kim, J. Kim, H. Kim, and J. Kim, Int. J. Refract. Met. Hard Mater. **122**, 106738 (2024).
[15] Z. Qiao, S. Dong, Q. Li, X. Lu, R. Chen, S. Guo, A. Yan, and W. Li, J. Alloys Compd. **963**, 171250 (2023).
[16] Q. Zhou, X. Bai, X. Ma, Z. Zhou, P. Lin, X. Xing, Y. Zhang, S. Jiang, and Y. Zhang, J. Alloys Compd. **1022**, 179810 (2025).
[17] G. Lambard, T. T. Sasaki, K. Sodeyama, T. Ohkubo, and K. Hono, Scr. Mater. **209**, 114341 (2022).
[18] A. Melsheimer, M. Seeger, and H. Kronmüller, J. Magn. Magn. Mater. **202**, 458 (1999).
[19] Z. Hu, D. Ma, H. Qu, H. Zhang, C. Luo, and H. Wang, J. Alloys Compd. **763**, 273 (2018).
[20] L. Q. Yu, Y. H. Wen, and M. Yan, J. Magn. Magn. Mater. **283**, 353 (2004).
[21] S. Pandian, V. Chandrasekaran, G. Markandeyulu, K. J. L. Iyer, and K. V. S. Rama Rao, J. Appl. Phys. **92**, 6082 (2002).
[22] K. Zhang, Z. Yue, Y. Ma, J. He, X. Li, W. Gong, and Y. Huang, J. Alloys Compd. **969**, 172391 (2023).
[23] B. Sun, Y.-C. Liang, Y. Zhou, J.-X. Xie, M.-Q. Wang, and G.-P. Chen, Mater. Today Commun. **37**, 107385 (2023).
[24] Y. Wang, S. Li, S. Li, and M. Chen, Mater. Today Commun. **38**, 108294 (2024).
[25] Z. Wang, S. Zhang, J. Wang, M. Zhang, Y. Chen, B. Li, T. Zhao, M. Zhu, F. Hu, B. Shen, and W. Li, Acta Mater. **292**, 121031 (2025).
[26] Y. Cao, Y. Liu, P. Zhang, G. Xu, J. Liu, J. Chen, X. Yi, and Y. Wu, J. Rare Earth. **39**, 1409 (2021).
[27] W. Mo, L. Zhang, A. Shan, L. Cao, J. Wu, and M. Komuro, J. Alloys Compd. **461**, 351 (2008).
[28] F. Viala, F. Joly, E. Nevalainen, M. Sagawa, K. Hiraga, and K.T. Park, J. Magn. Magn. Mater. **242-245**, 1329 (2002).
[29] C. Jin, R. Chen, W. Yin, X. Tang, Z. Wang, J. Ju, D. Lee, and A. Yan, J. Alloys Compd. **670**, 72 (2016).
[30] Y. Wu, K. P. Skokov, L. Schäfer, F. Maccari, H. Xu, X. Wang, C. Jiang, and O. Gutfleisch, Acta Mater. **263**, 119517 (2024).
[31] S. K. Haider, M.-C. Kang, J. Hong, Y. S. Kang, C.-W. Yang, and D. Kim, Sci. Rep. **11**, 6347 (2021).
[32] Z. Wang, S. Chang, X. Bao, Y. Zheng, J. Li, and X. Gao, J. Alloys Compd. **1008**, 176727 (2024).