metals

MDPI

Article

Domain-Knowledge-Guided Feature Engineering for Small-Sample Machine Learning Prediction of Mechanical Properties in Low-Carbon Hot-Rolled Steel Strips

Saurabh Tiwari [1,*], Hyoju Ahn [1], Jongwon Lee [1] and Nokeun Park [1,2,*]

[1] School of Materials Science and Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea; ahnalysis@yu.ac.kr (H.A.); jongwon1229@gmail.com (J.L.)
[2] Institute of Materials Technology, Yeungnam University, Gyeongsan 38541, Republic of Korea
* Correspondence: saurabht@yu.ac.kr (S.T.); nokeun_park@yu.ac.kr (N.P.)

Abstract

Industrial steel property prediction is often constrained by limited labelled data, reducing the effectiveness of conventional machine learning models. This study investigated whether metallurgy-informed feature engineering enhances predictive performance under small-data conditions. A representative set of 300 samples from an industrial low-carbon hot-rolled steel strip dataset (C: 0.02–0.06 wt%; Mn: 0.17–0.38 wt%) was used to derive five physically meaningful descriptors: carbon equivalent (CE), nitrogen-to-aluminum ratio (N/Al), microalloying efficiency index (MEI), thermal processing parameter (TPP), and solid solution strengthening index (SSSI). These descriptors were combined with the original 17 compositional and processing variables to create a 22-feature dataset. Random Forest (RF) and Extreme Gradient Boosting (XGBoost) models were evaluated on an independent 60-sample test set using 5-fold cross-validation. Feature engineering improved the prediction accuracy, with the greatest gain observed for elongation. For XGBoost, the mean percentage error decreased from 3.23% to 3.05%, whereas the test-set R^2 increased from 0.4935 to 0.5444, representing a 10.3% improvement in the explained variance. For the yield strength, the Random Forest method increased the R^2 from 0.4744 to 0.4861. Permutation importance and partial dependence analyses identified MEI and TPP as the six most influential predictors across all targets, confirming that the engineered descriptors provide complementary metallurgical information. Learning curve analysis showed slightly higher cross-validation R^2 values at intermediate training sizes (n = 125–175), indicating modestly improved sample efficiency. These findings establish domain-informed feature engineering as an interpretable and practical strategy for improving machine learning in data-limited steel manufacturing processes.

Keywords: random forest; extreme gradient boosting; permutation importance; partial dependence plots; physics-informed descriptors; industrial dataset

Check for updates

Academic Editor: Diego Celentano

Received: 30 July 2026
Revised: 18 August 2026
Accepted: 19 August 2026
Published: 21 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Predicting the mechanical properties of hot-rolled steel strips based on their chemical composition and thermomechanical processing parameters remains a challenge in steelmaking [1,2]. Low-carbon hot-rolled steel strips are widely used in automotive body and structural components, general construction, pipe manufacturing, and consumer goods applications, highlighting the industrial value of accurate and interpretable property prediction [1]. Machine learning (ML) has emerged as a powerful approach for this task, with

feedforward neural networks [1], gradient-boosted trees [3], support vector regression [4], and, more recently, interpretable ensemble models combined with explainable-AI methods [5] achieving practically useful prediction errors below 5% for yield strength (YS), ultimate tensile strength (UTS), and elongation (EL) in industrial datasets of moderate size. A critical yet underexplored challenge in this field is the limited dataset sizes. Industrial steel property databases are expensive to generate because each sample requires a full production trial, chemical analysis of numerous alloying elements, and standardized mechanical testing performed as part of the mill's routine industrial quality-control procedure. Consequently, many industrial datasets contain fewer than 300–400 labelled samples [1,6,7]. Under these conditions, high-dimensional feature spaces, particularly those comprising 17 or more raw compositional variables, increase the risk of overfitting and model instability while reducing the performance advantage of more complex learning algorithms [8,9]. Two established strategies can mitigate the small-data problem: (i) regularization and cross-validation during model development and (ii) feature engineering, which reduces the effective dimensionality while preserving physically meaningful information. The latter approach, adopted in this study, offers the additional benefit of generating interpretable composite descriptors with direct metallurgical significance, thereby improving communication between data scientists and process engineers [10,11].

Feature engineering has long been used in empirical physical metallurgy. The carbon equivalent (CE), originally developed to estimate hardenability and weldability, condenses compositional information into a physically meaningful scalar [12]. Similarly, the strengthening indices derived from Pickering's solid-solution model [13] combine the contributions of multiple substitutional alloying elements into a single descriptor, representing their collective strengthening effect. The nitrogen-to-aluminum ratio reflects the tendency for AlN precipitation, which influences austenite grain boundary pinning [14]. Although these descriptors are widely used in metallurgical analyses, they have rarely been explicitly incorporated as input features alongside the original compositional variables in machine-learning models.

This study aims to address this issue. Using 300 representative samples selected from the industrial low-carbon hot-rolled steel database originally described by Tiwari et al. [1], we (i) constructed five physically motivated composite descriptors; (ii) evaluated whether augmenting the original 17-feature set improved the Random Forest (RF) and Extreme Gradient Boosting (XGBoost) predictions of YS, UTS, and EL using an independent 60-sample test set; (iii) assessed sample efficiency through learning curve analysis; and (iv) interpreted the contribution of engineered descriptors using permutation importance and partial dependence plots (PDPs). Three hundred representative samples were deliberately selected rather than using the full 435-sample database to represent the practically common scenario of limited industrial data, making the findings directly relevant to steel plants in the early stages of data collection. Although the industrial database used in this study originates from a previously reported study [1], which employed a feedforward neural network on the complete 435-sample dataset to optimize the network architecture and predict mechanical properties, the present investigation focuses on physics-informed feature engineering and interpretable machine learning under data-limited conditions—topics not addressed in the earlier work. Using 300 representative samples, we investigated whether metallurgically derived composite descriptors could improve predictive accuracy, sample efficiency, and model interpretability. Therefore, the objective is descriptor discovery and validation of physics-informed feature engineering rather than neural network architecture optimization, with the emphasis shifting from algorithm comparison to descriptor design.

2. Materials and Methods

This section describes the industrial dataset, construction of the five physically motivated engineered descriptors, and machine learning models and evaluation procedures used to assess their impact on predictive performance.

2.1. Dataset

The full dataset originates from an industrial–academic collaboration with Tata Steel, Jamshedpur, India, and comprises 435 observations of low-carbon hot-rolled steel strips with 17 input variables (15 alloying elements together with finish rolling temperature (FRT) and coiling temperature (CTT)) and three target mechanical properties (YS, UTS, and EL) [1]. In this study, 300 representative samples were drawn using stratified random selection (random seed = 99) to simulate a realistic, limited-data industrial scenario. This selected dataset preserved the statistical characteristics of the full database, with mean values of 290.3 MPa versus 290 MPa for YS, 357.7 MPa versus 357 MPa for UTS, and 40.0% versus 40% for EL, confirming that the stratified selection did not introduce a systematic bias relative to the complete 435-sample database. The mechanical properties of all samples were obtained by standardized tensile testing of coil samples as part of the mill's industrial quality-control procedure, and the steel grade considered here is representative of low-carbon hot-rolled strips used in automotive, structural, and general engineering applications. The 300 samples were then split into 240 for training and 60 for testing (80/20 ratio) using a second independent fixed random seed (seed = 42) to ensure that the split was exactly reproducible. Table 1 lists the original 17 input features.

Table 1. Original input features (17) and output properties of the 300-sample dataset. The values were reproduced from Tiwari et al. Adapt form Ref. [1].

Parameter	Min	Max	Mean	Std
Al (wt%)	0.021	0.065	0.043	0.011
AlS (wt%)	0.020	0.063	0.041	0.010
C (wt%)	0.020	0.060	0.040	0.008
Cr (wt%)	0.014	0.039	0.026	0.006
Cu (wt%)	0.005	0.012	0.008	0.002
Mn (wt%)	0.170	0.380	0.275	0.052
Mo (wt%)	0.001	0.003	0.002	0.001
N (ppm)	26	58	42	8
Nb (wt%)	0.000	0.004	0.002	0.001
Ni (wt%)	0.008	0.018	0.013	0.002
P (wt%)	0.007	0.024	0.015	0.004
S (wt%)	0.003	0.020	0.011	0.004
Si (wt%)	0.004	0.028	0.016	0.006
Ti (wt%)	0.000	0.003	0.001	0.001
V (wt%)	0.001	0.002	0.001	0.0004
FRT (°C)	850.23	901.14	875.68	12.73
CTT (°C)	570	650	610	20.41
YS (MPa)	242	338	290	24.12
UTS (MPa)	317	397	357	20.15
EL (%)	34	46	40	3.02

2.2. Engineered Feature Construction

Five composite features were constructed from the original inputs based on established physical metallurgy relationships. Each descriptor is physically motivated, dimensionally consistent, and computed from features already present in the dataset, requiring no addi-

tional experimental measurements. The equations, physical basis, dataset, and ranges for all engineered descriptors are summarized in Table S1 (Supplementary Materials).

2.2.1. Carbon Equivalent (CE)

The carbon equivalent was computed using the International Institute of Welding (IIW) formula adapted for low-alloy steels: CE = C + Mn/6 + (Cr + Mo + V)/5 + (Ni + Cu)/15 [12]. CE provides a single-valued summary of the combined hardenability contribution of all substitutional strengthening elements, normalized to the equivalent carbon-strengthening effect. For the present low-carbon steel family, the CE content ranged from 0.062 to 0.125 wt% (mean = 0.094 wt%). CE is widely used in welding codes and structural standards because it captures the synergistic composition–hardening relationship that individual elemental features encode separately. The CE was originally developed for weldability and hardenability assessment; here, it is used as an empirical, composition-based descriptor for the machine learning models rather than as a direct predictor of weldability.

2.2.2. Nitrogen-to-Aluminum Ratio (N/Al Ratio)

The molar ratio of dissolved nitrogen to total aluminum was computed as N (ppm)/(Al (wt%) × 10,000), where the denominator converts Al to an equivalent ppm. This ratio governs the tendency of AlN precipitation during hot rolling and subsequent cooling. When N/Al > approximately 0.045 (stoichiometric ratio), excess nitrogen is more likely to remain in the solid solution, where it can act as a solid-solution strengthener. When N/Al < 0.045, aluminum is in excess and tends to bind nitrogen as AlN dispersoids. Aluminum is added to these steels as a deoxidizer and grain refiner and has a strong affinity for nitrogen; under Al-rich conditions, fine AlN particles can precipitate and pin austenite and ferrite grain boundaries (Zener pinning), promoting grain refinement and contributing to strength through the Hall–Petch relationship while reducing the nitrogen available for solid-solution strengthening [14]. Therefore, the N/Al ratio reflects the competition between grain boundary pinning and interstitial strengthening mechanisms, both of which contribute to the YS.

2.2.3. Microalloying Efficiency Index (MEI)

Niobium, titanium, and vanadium are microalloying elements that strengthen through precipitation and recrystallization retardation, but with substantially different efficiencies per unit weight. Based on the published strengthening coefficients for low-carbon ferritic steels [15], the MEI was defined as follows: MEI = 3000 × Nb + 1500 × V + 1000 × Ti (MPa); here, the coefficients represent the approximate YS contributions per wt% within the relevant compositional range. The MEI is an empirical, composition-based index used to combine the microalloying variables into a single physically motivated descriptor for the machine learning models, rather than a direct measurement of their individual strengthening contributions. The MEI ranged from 2.08 to 16.94 MPa across the dataset (mean = 8.64 MPa), capturing the combined microalloying potential as a single descriptor of the MEI.

2.2.4. Thermal Processing Parameter (TPP)

A single thermal severity index was defined to capture the combined effect of the finish rolling temperature and coiling temperature on grain coarsening and strength reduction: TPP = 0.5 × (FRT − 850) + 0.3 × (CTT − 570). The coefficients 0.5 and 0.3 reflect the relative sensitivity of the ferritic grain size to the FRT and CTT, respectively, estimated from published parametric studies on similar steels [1,16–18]. TPP is an empirical index that combines the FRT and CTT using these coefficients. It is intended as a compact descriptor of relative thermal severity for use as a model input, rather than a direct physical measurement

of thermal exposure or grain coarsening severity. Higher TPP values correspond to greater total thermal exposure and are expected to correlate negatively with YS through Hall–Petch grain coarsening. The TPP ranged from 0.38 to 45.75 (mean = 25.5), with a Pearson correlation of −0.348 with the YS.

2.2.5. Solid Solution Strengthening Index (SSSI)

Based on Pickering's empirical strengthening model for low-carbon ferritic steels [13], the SSSI was computed as SSSI = 5500 × N (wt%) + 37 × Mn + 83 × Si + 39 × Cu + 31 × Cr, where N is converted from ppm to wt% (N (wt%) = N (ppm)/1,000,000). The coefficients represent the published solid-solution strengthening contributions (MPa per wt%) for each element in the body-centered cubic iron matrix. The SSSI is an empirical, composition-based index used as a physically motivated descriptor of solid-solution strengthening potential rather than a direct measurement of strengthening. The SSSI ranged from 8.25 to 17.26 MPa (mean = 12.98 MPa).

2.3. Machine Learning Modelling and Evaluation

Two tree-based ensemble algorithms, Random Forest (RF) and Extreme Gradient Boosting (XGBoost), were selected because of their strong performance on structured tabular regression problems, particularly under small-to-medium dataset conditions [8,19]. RF was trained using 300 estimators with max_features = sqrt, whereas XGBoost employed 400 estimators, max_depth = 4, a learning rate of 0.05, subsample = 0.8, and colsample_bytree = 0.8. Separate models were developed for each mechanical property (YS, UTS, and EL) using both the original 17-feature and expanded 22-feature datasets, resulting in 12 different model configurations. Feature scaling was not applied because both algorithms were insensitive to the monotonic feature transformations.

The model performance was evaluated using a fixed 240/60 train–test split, corresponding to the commonly adopted 80/20 ratio for small data sets. The test set performance was quantified using the mean percentage error (MPE), root-mean-square error (RMSE), coefficient of determination (R^2), and mean absolute error (MAE). The generalization capability was assessed using 5-fold cross-validation for all 300 samples. To evaluate the sample efficiency, learning curves were generated by progressively increasing the training size from 15% to 100% of the training set while recording both the training and cross-validation R^2. The interpretability of the model was investigated using permutation importance and partial dependence plots (PDPs). Permutation importance was calculated on the test set using 30 random permutations per feature, with feature importance measured as the mean decrease in R^2. PDPs were generated for the six most influential predictors identified by permutation importance for each target, chosen to keep the interpretability analysis focused and readable, as most of the remaining features in the 22-feature set showed near-zero importance. Using the training dataset, the PDPs provide a visual representation of the marginal effect of each feature on the predicted yield strength while averaging the remaining variables [20].

3. Results

This section reports the test set performance of the original and engineered feature sets, followed by learning curve, permutation-importance, partial-dependence, cross-validation, and correlation analyses to characterize the contribution of the engineered descriptors to model performance and interpretability. Exact numerical results are reported in tables for reproducibility and cross-study comparison, followed by figures showing the corresponding visual results of the study.

3.1. Test-Set Performance of Original and Engineered Feature Sets

Table 2 reports the predictive performance of Random Forest (RF) and Extreme Gradient Boosting (XGBoost) using the original 17-feature and engineered 22-feature datasets for yield strength (YS), ultimate tensile strength (UTS), and elongation (EL). All models achieved mean percentage errors (MPEs) below 3.5%, demonstrating robust predictive accuracy despite the limited dataset size used for training the models. XGBoost with the original feature set provided the highest accuracy for UTS (MPE = 2.15%, R^2 = 0.4051), whereas RF with the engineered feature set provided the best performance for YS (MPE = 2.54%, R^2 = 0.4861). Feature engineering had the greatest impact on elongation prediction. For XGBoost, the MPE rate decreased from 3.23% to 3.05%, whereas the test-set R^2 increased from 0.4935 to 0.5444, corresponding to a 10.3% improvement in explained variance. RF also showed improved EL prediction, with R^2 increasing from 0.4167 to 0.4620. For YS, engineered features produced a modest improvement for RF (R^2: 0.4744 to 0.4861), whereas the performance of XGBoost remained essentially unchanged. In contrast, the UTS predictions showed only minor differences between the original and engineered feature sets, indicating that the original compositional variables captured most of the information governing the tensile strength.

Table 2. Test-set performance metrics (60 samples) for RF and XGBoost using the original (17 features) and engineered (22 features) datasets.

Model	Features	Target	MPE (%)	RMSE	R^2	MAE
Random Forest	Original (17)	YS	2.573	9.476	0.4744	7.414
Random Forest	Original (17)	UTS	2.199	9.430	0.3396	7.948
Random Forest	Original (17)	EL	3.410	1.666	0.4167	1.377
XGBoost	Original (17)	YS	2.649	9.202	0.5044	7.632
XGBoost	Original (17)	UTS	2.151	8.950	0.4051	7.767
XGBoost	Original (17)	EL	3.233	1.552	0.4935	1.305
Random Forest	Engineered (22)	YS	2.539	9.370	0.4861	7.298
Random Forest	Engineered (22)	UTS	2.186	9.209	0.3701	7.899
Random Forest	Engineered (22)	EL	3.291	1.600	0.4620	1.328
XGBoost	Engineered (22)	YS	2.582	9.238	0.5005	7.428
XGBoost	Engineered (22)	UTS	2.168	9.373	0.3476	7.827
XGBoost	Engineered (22)	EL	3.049	1.472	0.5444	1.227

The best value for each target property is highlighted in bold. For comparison, the source study [1] reported MPE values of 4.44%, 3.54%, and 4.84% for YS, UTS, and EL, respectively, using the complete 435-sample dataset.

In Figure 1, we compare the predicted and measured mechanical properties of the XGBoost model trained using the engineered feature set. Predictions for all three properties were closely distributed around the line of perfect agreement, with most samples lying within the ±5% error band. The relatively tight clustering is consistent with the quantitative performance metrics presented in Table 2. The model evaluation metrics are listed in Table S2 (Supplementary Materials).

In Figure 2, we compare the test set R^2 values obtained using the original and engineered feature sets. The engineered descriptors consistently improved the RF performance across all three properties, with the largest increase observed for EL (+0.045). For XGBoost, the principal benefit was also obtained for EL (+0.051), whereas the changes in YS (−0.004) and UTS (−0.058) were comparatively small. These results indicate that the engineered descriptors provide the greatest benefit for properties governed by multiple interacting strengthening mechanisms, particularly in terms of elongation.

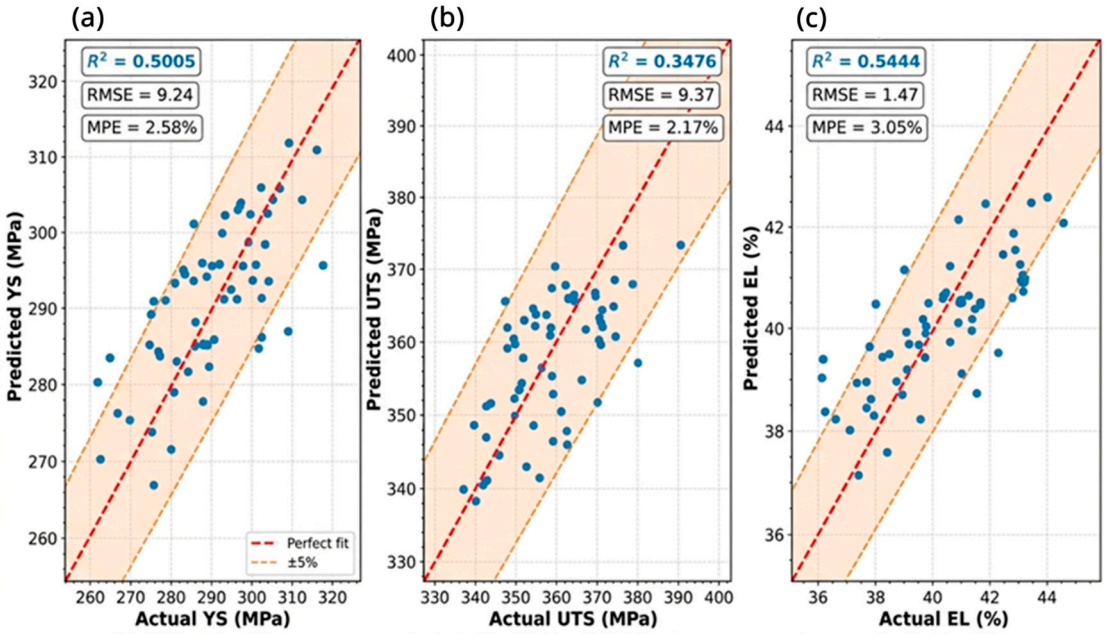


Figure 1. Predicted versus measured mechanical properties for the XGBoost model using the engineered 22-feature dataset on the 60-sample test set: (**a**) yield strength (YS), (**b**) ultimate tensile strength (UTS), and (**c**) elongation (EL). The red dashed line represents perfect agreement, and the orange shaded bands indicate a ±5% prediction error.

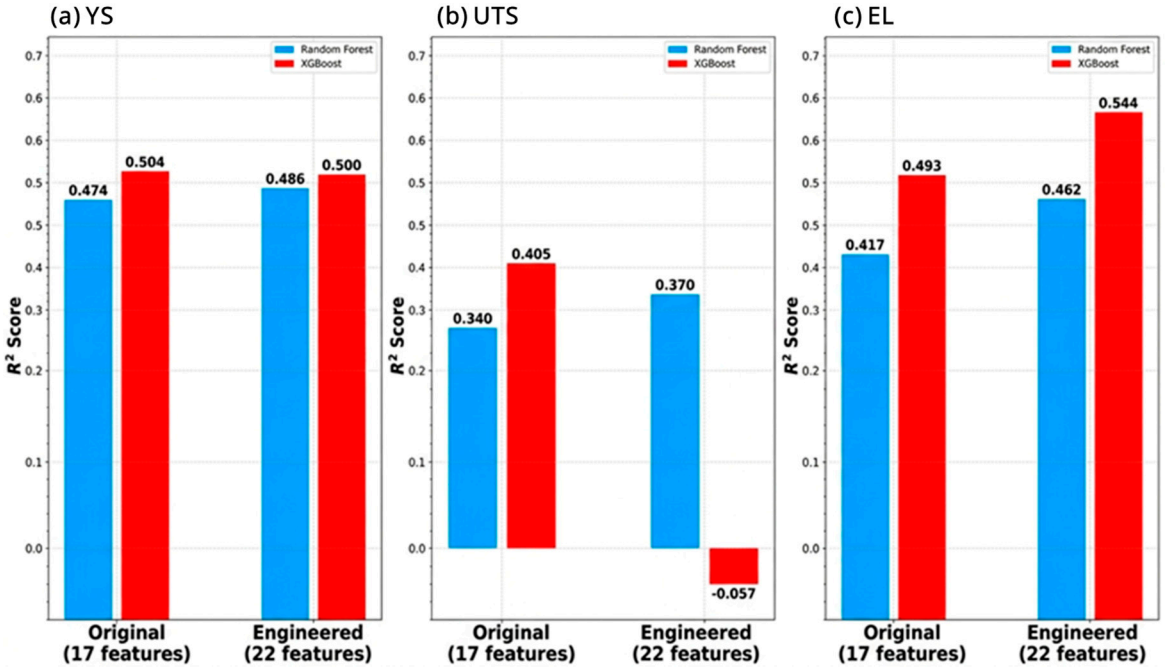


Figure 2. Test-set R^2 comparison between the original (17 features) and engineered (22 features) datasets for RF and XGBoost across (**a**) YS, (**b**) UTS, and (**c**) EL. ΔR^2 values above each bar pair indicate the effect of feature engineering; green indicates improved performance, whereas red indicates reduced performance.

3.2. Learning Curve Analysis

A learning curve analysis was performed to evaluate model generalization and sample efficiency as the training set size increased. In Figure 3, we compare the learning curves of XGBoost trained using the original and engineered feature sets for yield strength (YS), showing the training R^2 and 5-fold cross-validation (CV) R^2 as functions of the training set

size. Both feature sets exhibited the characteristic behavior of small-data learning: training R^2 remained close to 1.0 across all training sizes, reflecting the high fitting capacity of XGBoost, whereas CV R^2 increased rapidly at small training sizes before approaching a stable plateau as more samples were incorporated.

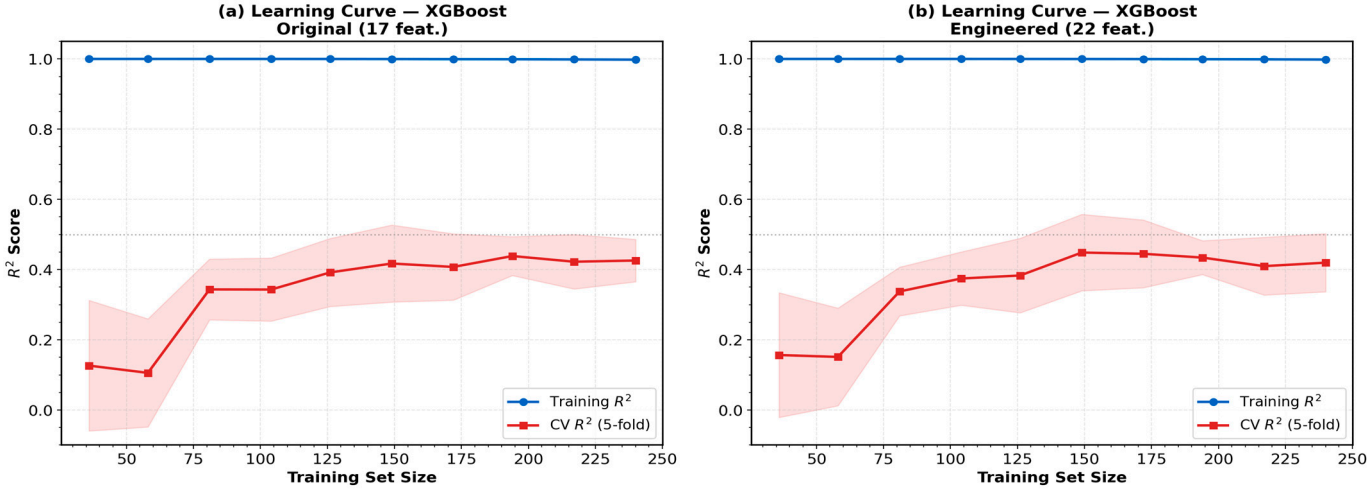


Figure 3. Learning curves for XGBoost prediction of yield strength (YS) using (**a**) the original 17-feature dataset and (**b**) the engineered 22-feature dataset. The solid lines represent the training R^2, the dashed lines represent the 5-fold cross-validation (CV) R^2, and the shaded regions indicate ±1 standard deviation across the CV folds. The engineered feature set achieved a marginally higher CV R^2 at intermediate training sizes (n = 125–175), with both feature sets converging to comparable plateau values beyond n = 175.

For the original 17-feature set (Figure 3a), CV R^2 increased from approximately 0.13 at n = 36 to approximately 0.35–0.40 between n = 75 and n = 125, before stabilizing at approximately 0.42–0.44 from n = 175 onwards. The engineered 22-feature set (Figure 3b) followed a similar trend, beginning at a slightly higher CV R^2 (approximately 0.16 at n = 36), reaching approximately 0.45 at n = 150, and converging to a comparable plateau of approximately 0.42–0.44 beyond n = 175. Both feature representations achieved a similar asymptotic CV R^2, indicating that the engineered descriptors did not substantially alter the predictive limit imposed by available data. However, the engineered feature set consistently attained slightly higher cross-validation R^2 values at intermediate training sizes (n = 125–175), suggesting a modest improvement in sample efficiency within the most practically relevant data-limited regime. The persistent gap between the training R^2 (≈1.0) and CV R^2 (≈0.42–0.44) highlights the importance of cross-validation in assessing generalization performance when training data are limited.

3.3. Permutation Feature Importance

Permutation importance analysis was performed for the XGBoost model using the engineered 22-feature dataset to quantify the contribution of each predictor to the model performance (Figure 4). Feature importance was measured as the mean decrease in R^2 over 30 random permutations, with error bars representing ±1 standard deviation. The results revealed a consistent importance hierarchy across the three target properties, highlighting the target-specific contributions of engineered descriptors. Nitrogen was expected to rank highly because it contributes through direct interstitial solid-solution strengthening and AlN precipitation/grain-boundary pinning (Section 2.2.2). Despite its low concentration (26–58 ppm), nitrogen can have a strong strengthening effect because interstitial solutes interact strongly with the dislocations. For YS (Figure 4a), nitrogen was the dominant

predictor, with a mean decrease in R^2 of approximately 0.35, more than four times that of the second-ranked feature.

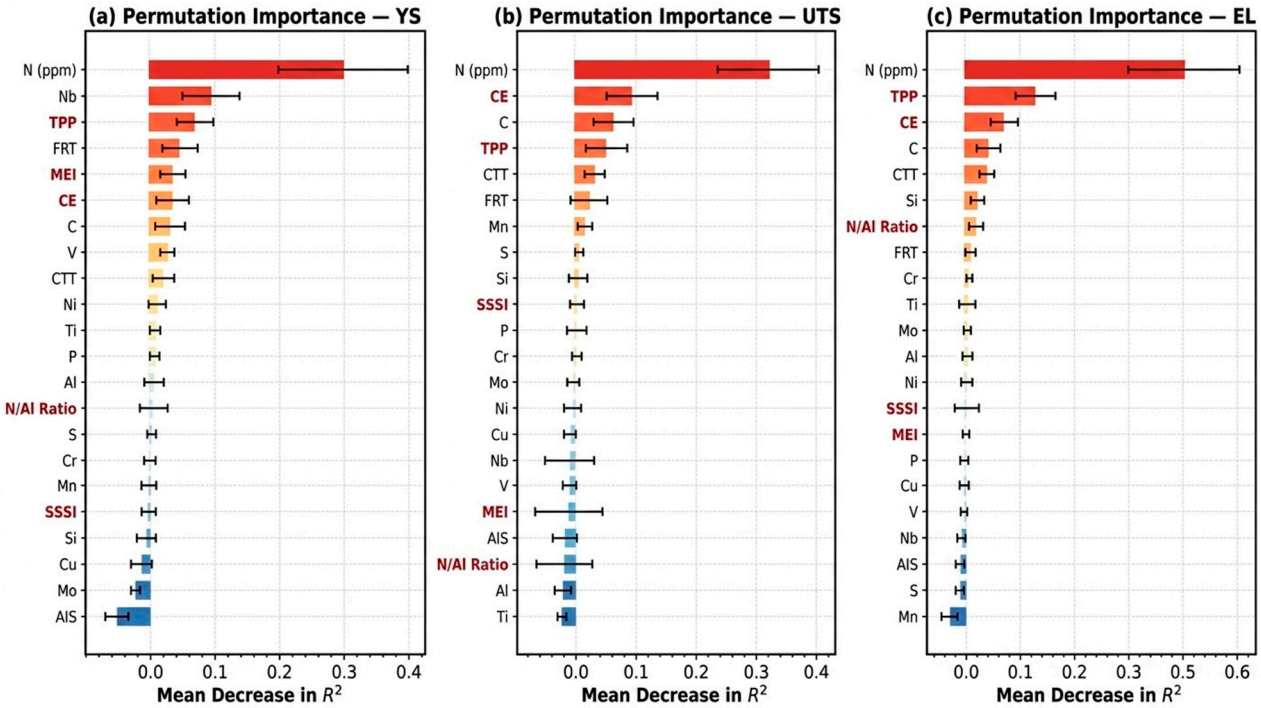


Figure 4. Permutation importance for the XGBoost model using the engineered 22-feature dataset for (**a**) yield strength (YS), (**b**) ultimate tensile strength (UTS), and (**c**) elongation (EL) prediction. Feature importance was expressed as the mean decrease in R^2 over 30 permutations.

Nb ranked second, followed by the engineered thermal processing parameter (TPP), finish rolling temperature (FRT), microalloying efficiency index (MEI), and carbon equivalent (CE). The presence of TPP, MEI, and CE among the highest-ranked predictors demonstrates that these engineered descriptors provide complementary information beyond the original variables of the study. In contrast, AlS exhibited a slightly negative permutation importance, indicating that it contributed little independent information owing to its strong collinearity with Al. For UTS (Figure 4b), N again ranked as the most influential predictor (approximately 0.33), followed by CE (approximately 0.09), confirming that the carbon equivalent is the most important engineered descriptor for tensile strength prediction. Carbon (C), TPP, CTT, and FRT also made positive contributions. In contrast, Nb ranks among the least important features, whereas both MEI and the N/Al ratio exhibit slightly negative permutation importance, suggesting that these descriptors provide limited additional information once the dominant compositional and processing variables are considered. For EL (Figure 4c), N remained the dominant predictor, with the highest importance observed for all target properties (approximately 0.50). TPP (approximately 0.10) and CE (approximately 0.08) were the next most influential predictors, followed by C and CTT. Silicon (Si) and the N/Al ratio also contributed positively, whereas the MEI and solid solution strengthening index (SSSI) exhibited negligible or slightly negative importance, indicating limited independent contributions to the elongation prediction. First, nitrogen was consistently the most influential predictor of YS, UTS, and EL, with mean decreases in R^2 of approximately 0.35, 0.33, and 0.50, respectively, highlighting its dominant role in determining the mechanical behavior of this family of steels. Second, CE and TPP rank among the five most influential predictors of multiple target properties, establishing them as the most broadly useful, engineered descriptors. Third, the comparatively large error bars associated with nitrogen, particularly for EL, indicate greater variability in model

sensitivity across permutation repeats, reflecting the heterogeneous distribution of nitrogen within the industrial dataset.

3.4. Partial Dependence Plots

Partial dependence plots (PDPs) were used to examine the marginal effect of the six most influential features on the XGBoost predictions of the yield strength (YS) (Figure 5). Computed from the 240-sample training set, the PDPs complemented the permutation importance analysis by illustrating how changes in individual features influenced the predicted YS while averaging the remaining variables. Nitrogen (Figure 5a) exhibited a stepwise positive relationship, with the predicted YS increasing from approximately 281 MPa at N = 28 ppm to 297.5 MPa at N = 57 ppm, with distinct upward transitions near 35 and 44 ppm. The overall increase of approximately 16 MPa represented the largest PDP range among all features, consistent with nitrogen being the most influential predictor identified in the permutation importance analysis. The staircase pattern likely reflects discrete compositional clusters within the industrial dataset and is consistent with the combined effects of interstitial solid-solution strengthening and AlN grain boundary pinning [14]. Niobium (Figure 5b) exhibited a threshold-type response. The predicted YS remained nearly constant below approximately 0.0015 wt% Nb, increased rapidly between 0.0015 and 0.0025 wt%, and reached a plateau above 0.0025 wt%, corresponding to an overall strengthening increment of approximately 7 MPa. This behavior is consistent with the increasing contribution of Nb microalloying through precipitation strengthening and recrystallization retardation when sufficient Nb is available [15].

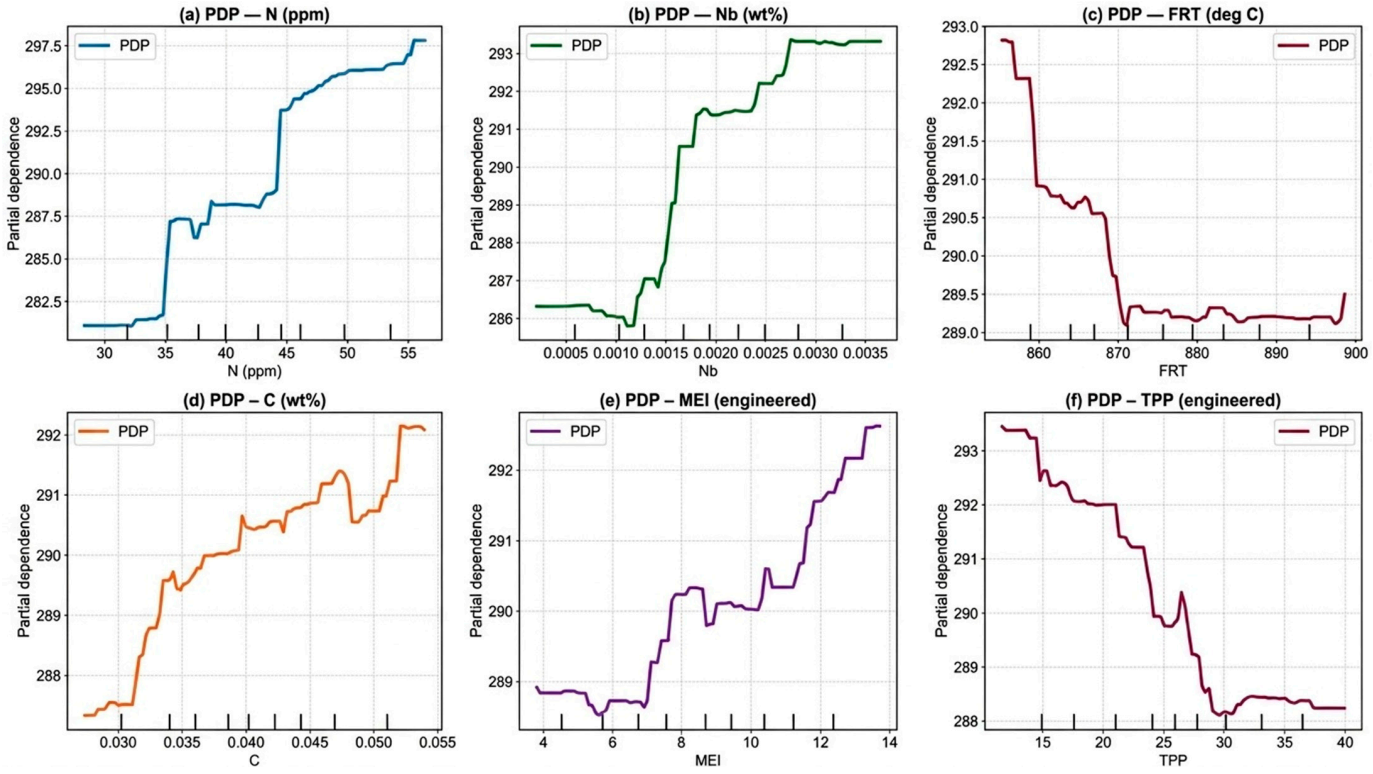


Figure 5. Partial dependence plots (PDPs) for the six most influential features affecting yield strength (YS): (**a**) nitrogen (N), (**b**) niobium (Nb), (**c**) finish rolling temperature (FRT), (**d**) carbon (C), (**e**) microalloying efficiency index (MEI), and (**f**) thermal processing parameter (TPP). PDPs show the marginal effect of each feature on the predicted YS while averaging the distribution of the other features. The engineered descriptors (MEI and TPP) are indicated in purple.

The PDP for the finish rolling temperature (FRT) (Figure 5c) revealed a negative relationship with the YS. The predicted YS decreased from approximately 292.7 MPa to 290.5 MPa between 855 °C and 870 °C, after which the curve became nearly constant at approximately 289–290 MPa. This trend suggests that the greatest sensitivity of YS to the finish rolling temperature occurs within the lower portion of the processing window. Carbon (Figure 5d) exhibited a gradual positive relationship, increasing from approximately 287 MPa at 0.027 wt% to approximately 292 MPa at 0.055 wt%. The nearly monotonic trend, interrupted only by minor local fluctuations, is consistent with the progressive solid-solution strengthening throughout the investigated low-carbon composition range. The engineered descriptors provide additional insights into the combined effects of the composition and processing. The microalloying efficiency index (MEI) (Figure 5e) remained nearly constant below MEI ≈ 8, increased steadily when MEI = 8–13, and approached a plateau above 13, corresponding to an overall strengthening effect of approximately 4 MPa. Similarly, the thermal processing parameter (TPP) (Figure 5f) exhibited a negative stepwise relationship, decreasing from approximately 293–294 MPa at low TPP values to approximately 288–289 MPa at higher values, representing an overall reduction of approximately 5–6 MPa. The staircase behavior of both engineered descriptors likely reflects discrete combinations of composition and processing conditions within the industrial dataset rather than continuous physical transitions.

Their clear monotonic trends and substantial PDP ranges demonstrate that these engineered features capture meaningful metallurgical information complementary to the original variables. Notably, TPP ranked higher than CTT in the permutation importance analysis (Figure 4a), indicating that the combined thermal descriptor provides a more informative representation of the processing effects than either temperature parameter alone. Overall, the PDPs are consistent with the permutation-importance results and demonstrate that the engineered descriptors are not only among the most influential predictors but also exhibit physically interpretable relationships with yield strength. In particular, the positive influence of MEI and the negative influence of TPP support their effectiveness as domain-informed composite features for improving model interpretability.

3.5. Cross-Validation Robustness

Five-fold cross-validation was used to evaluate the robustness and generalization of the machine learning models for the three target properties (Figure 6). The box plots in Figure 6 show the distribution of the cross-validation R^2 values for Random Forest (RF) and XGBoost using the original and engineered feature sets. For YS (Figure 6a), RF with the engineered feature set achieved the highest median cross-validation R^2 (0.47), followed by RF with the original features (0.43), XGBoost with the engineered features (0.39), and XGBoost with the original features (0.36). The consistently higher median performance of RF suggests that the bagging-based ensemble provides more stable generalization than gradient boosting for this relatively small dataset. For UTS (Figure 6b), RF with engineered features again produced the highest median cross-validation R^2 (0.40). XGBoost trained on the original feature set exhibited the widest interquartile range, with several folds producing negative R^2 values, indicating greater sensitivity to the composition of the training data.

Although the engineered feature set did not improve the median XGBoost performance for the UTS, it reduced the variability across the folds, suggesting enhanced model stability. For EL (Figure 6c), RF with engineered features achieved the highest median cross-validation R^2 (0.49) and the narrowest distribution among all model configurations, indicating the most consistent predictive performance for the ductility. The improvement in both the median performance and distribution width provides further evidence that

the engineered descriptors enhance the robustness of the model for elongation prediction. Overall, RF combined with the engineered 22-feature dataset consistently achieved the highest or joint-highest median cross-validation R^2 values for the three mechanical properties. These results indicate that the engineered descriptors improve the model stability and generalization, with the greatest benefit observed for the Random Forest under limited data conditions.

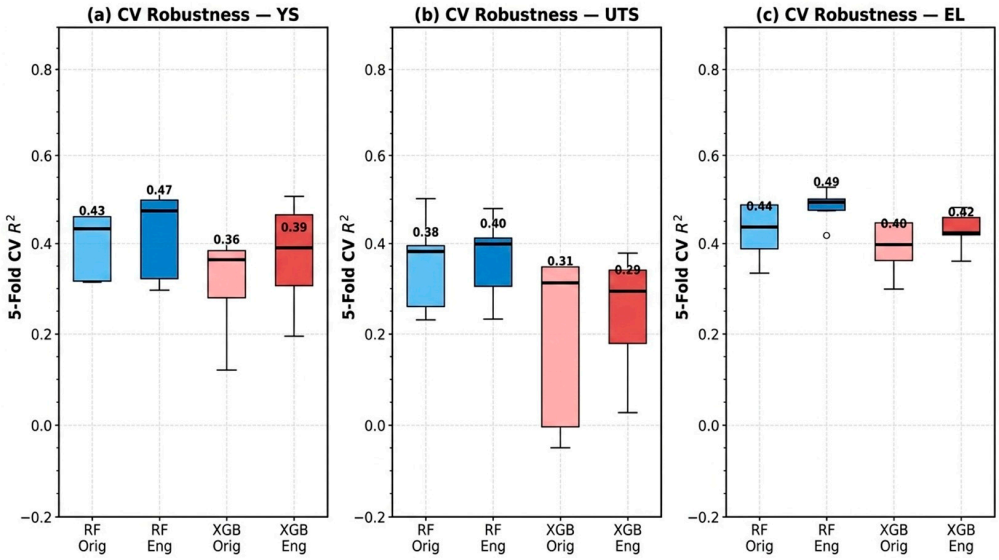


Figure 6. Five-fold cross-validation R^2 distributions for Random Forest (RF) and XGBoost using the original (17 features) and engineered (22 features) datasets for (**a**) yield strength (YS), (**b**) ultimate tensile strength (UTS), and (**c**) elongation (EL). RF-Orig = Random Forest with the original feature set; RF-Eng = Random Forest with the engineered feature set; XGB-Orig = XGBoost with the original feature set; XGB-Eng = XGBoost with the engineered feature set. The median R^2 values are indicated above each box.

3.6. Correlation of Engineered Features with Mechanical Properties

The linear relationships between the engineered descriptors and mechanical properties are shown in Figure 7. Pearson correlation analysis provides an initial assessment of the predictive relevance of each engineered feature before considering the nonlinear relationships captured by machine learning models. The carbon equivalent (CE) (Figure 7a) exhibited a modest positive correlation with YS (r = 0.241), indicating that an increase in the combined hardenability generally contributes to a higher strength, although the considerable scatter suggests that the relationship is not entirely linear. The microalloying efficiency index (MEI) (Figure 7b) showed the strongest positive correlation with YS (r = 0.340), consistent with the contributions of Nb, Ti, and V to precipitation strengthening. In contrast, the thermal processing parameter (TPP) (Figure 7c) displayed the strongest overall correlation with YS (r = –0.348), indicating that an increase in thermal exposure during processing is associated with a reduced yield strength.

The N/Al ratio (Figure 7d) also showed a moderate positive correlation with YS (r = 0.305), reflecting the combined influence of dissolved nitrogen and AlN precipitation. For the remaining properties, the solid solution strengthening index (SSSI) exhibited a modest positive correlation with UTS (r = 0.223), supporting its role as a secondary strengthening descriptor (Figure 7e). In contrast, MEI showed only a weak linear correlation with elongation (r = 0.058) (Figure 7f), indicating that its influence on ductility is limited when assessed using a linear relationship. Overall, the correlation analysis indicated that the engineered descriptors captured meaningful trends with the mechanical properties while remaining only moderately correlated with the target. This observation is consistent with

the improved performance of the nonlinear machine learning models, which can exploit interactions among these descriptors beyond the information contained in linear correlations.

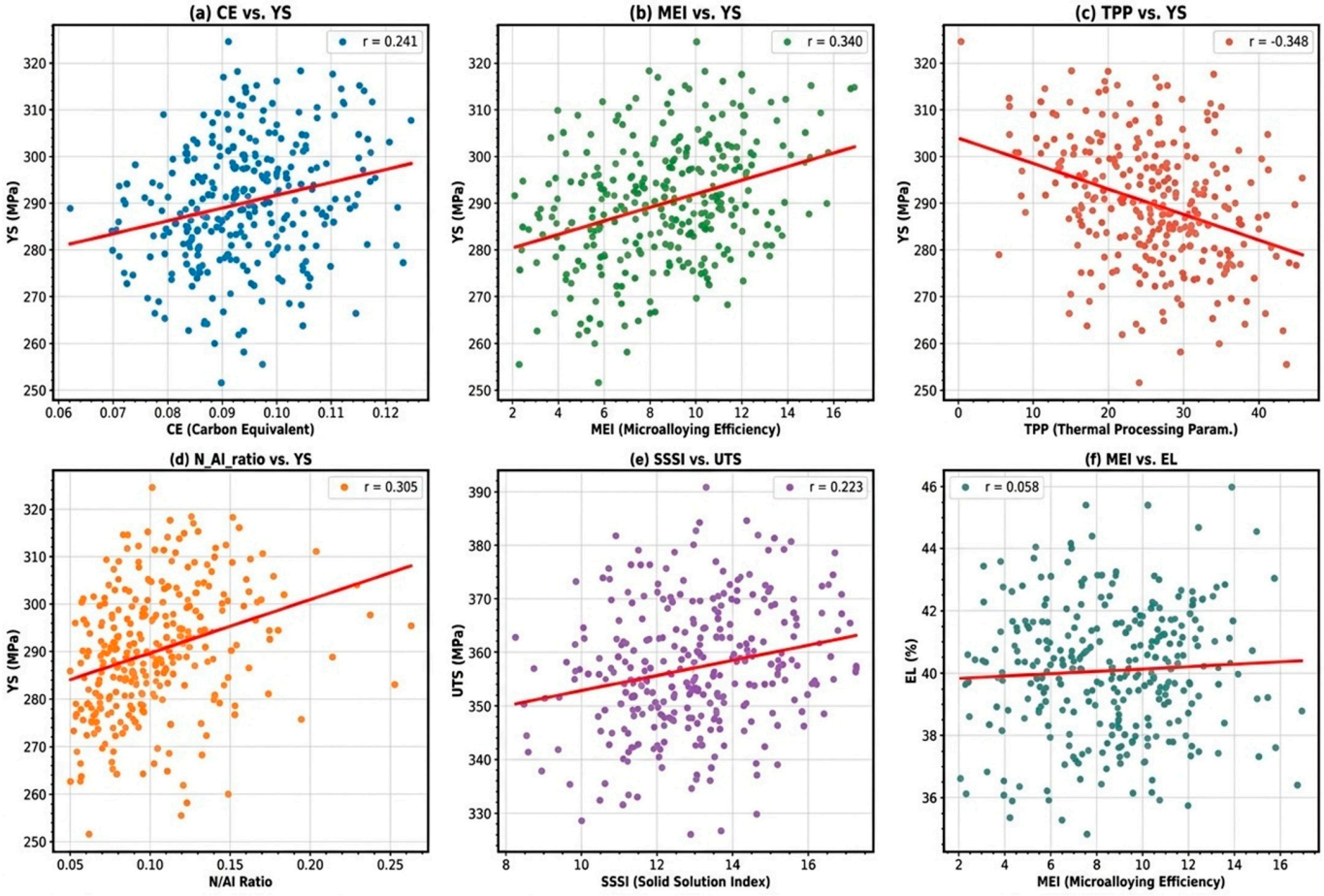


Figure 7. Scatter plots showing the relationships between engineered descriptors and their most strongly correlated mechanical properties: (**a**) carbon equivalent (CE) versus yield strength (YS), (**b**) microalloying efficiency index (MEI) versus YS, (**c**) thermal processing parameter (TPP) versus YS, (**d**) N/Al ratio versus YS, (**e**) solid solution strengthening index (SSSI) versus ultimate tensile strength (UTS), and (**f**) MEI versus elongation (EL). The red lines represent linear regression fits, and the Pearson correlation coefficients (*r*) are shown in each panel of the figure.

4. Discussion

4.1. Value of Feature Engineering on Small Datasets

The results demonstrate that domain-knowledge-driven feature engineering provides consistent, albeit modest, improvements in model performance when the training data are limited to 300 samples. The most pronounced gains occurred for elongation prediction, where the XGBoost test-set R^2 increased by 10.3% (from 0.4935 to 0.5444) after adding the five engineered descriptors. This improvement is physically meaningful because elongation is governed by a complex interplay of strengthening mechanisms, microstructural homogeneity, and work-hardening behavior, which is less directly represented by individual elemental concentrations than by physically derived composite descriptors such as the microalloying efficiency index (MEI) and the thermal processing parameter (TPP). By combining the individual contributions of Nb, V, and Ti into a single metallurgically weighted descriptor, the MEI enables the model to recognize precipitation strengthening effects more effectively under limited data conditions. The learning curve analysis provides supporting evidence for improved sample efficiency in data-limited regimes. While both feature sets converge to similar plateau cross-validation R^2 values (approximately 0.42–0.44) at the

largest training sizes, the engineered feature set achieves a slightly higher cross-validation R^2 of approximately 0.45 at intermediate training sizes (n = 125–175), compared with approximately 0.43 for the original feature set in the same range. Although the improvement is modest, it is practically relevant for steel plants operating with limited labelled datasets, where generating additional experimental data is often expensive and time-consuming. Notably, R^2 varies more across model–feature configurations than the MPE (Table 2), which remains between 2.5% and 3.5% in all cases. With only 60 test samples and a relatively narrow target range (Table 1), R^2 can be more sensitive to sampling variations than MPE. Therefore, we consider MPE as the primary indicator of practical predictive accuracy, while R^2 and the 5-fold cross-validation results (Figure 6) provide complementary information on model robustness and generalization. This interpretation is consistent with the learning curve analysis (Section 3.2), where the cross-validation R^2 plateaus at 0.42–0.44, while the MPE remains low and stable. For reference, our earlier feedforward neural network study using the complete 435-sample dataset [1] reported MPE values of 4.44%, 3.54%, and 4.84% for YS, UTS, and EL, respectively, whereas the tree-based models developed here achieved lower MPE values (2.5–3.3%) despite using a smaller training set. Although YS and UTS are physically related in this steel family, feature engineering improved the YS prediction more than the UTS prediction (Random Forest: R^2 0.4744 → 0.4861 for YS versus 0.3396 → 0.3701 for UTS). This difference may be related to the additional influence of the work-hardening behavior on UTS, which is not directly represented by the composition- and processing-based descriptors used here.

4.2. Physical Interpretation of Engineered Feature Importance

The appearance of the microalloying efficiency index (MEI) and thermal processing parameter (TPP) among the six most important predictors for all three target properties confirmed that these engineered descriptors encoded complementary metallurgical information beyond the original elemental and processing variables of the alloys. The MEI combines the precipitation and recrystallization retardation contributions of Nb, V, and Ti using physically grounded weighting coefficients (3000, 1500, and 1000 MPa per wt%, respectively), based on published strengthening data for low-carbon ferritic steels [15]. The PDP for MEI shows a threshold-type positive response: the predicted YS remains near 289 MPa for MEI < 8, then increases to approximately 292–293 MPa between MEI = 8 and 13, before plateauing above 13. This threshold is consistent with the combined microalloying level at which NbC precipitation becomes sufficiently dense to produce measurable strengthening, in agreement with the published precipitation-strengthening estimates of 10–20 MPa per 0.001 wt% Nb at the optimum precipitate size for this composition range [15]. The TPP combines the effects of the finish rolling temperature (FRT) and coiling temperature (CTT) into a single descriptor using physically motivated weighting (FRT coefficient = 0.5; CTT coefficient = 0.3), reflecting the greater sensitivity of the ferrite grain size to the finish rolling temperature within the investigated processing window [16]. The PDP for TPP exhibited a stepwise negative response, with the steepest reduction in the predicted YS occurring between TPP values of 20 and 30, which is consistent with accelerated ferrite grain coarsening as both thermal processing parameters increased toward their upper bounds. The higher permutation importance of TPP relative to CTT further suggests that combining the two processing variables provides a more informative representation of the thermal history than either parameter. The stepwise negative PDP response for TPP (Figure 5f) is consistent with increased ferrite grain coarsening at higher combined thermal exposures, which can reduce the yield strength through the Hall–Petch relationship.

4.3. Practical Implications, Limitations, and Future Perspectives

The findings of this study have several practical implications for developing machine learning models for industrial steel property prediction under limited data conditions. First, the five engineered descriptors (CE, N/Al ratio, MEI, TPP, and SSSI) can be incorporated into the original compositional and processing variables without requiring additional experimental measurements or incurring significant computational costs. Their inclusion consistently improved the predictive performance while preserving model interpretability. Second, the results indicate that Random Forest combined with the engineered feature set provides the most robust prediction of yield strength, whereas XGBoost with engineered features achieves the greatest improvement in elongation prediction. These observations suggest that algorithm selection should consider the target property rather than relying on a single model for all prediction tasks. Third, when fewer than 200 labelled samples are available, the engineered feature set is particularly valuable for training the model. The learning curve analysis indicates that at training sizes of n = 125–175, XGBoost with engineered features achieves cross-validation R^2 values approximately 0.02–0.05 higher than those obtained using the original feature set, representing a modest but consistent improvement in the most practically relevant data-limited regime. Furthermore, permutation importance analysis is a useful diagnostic tool for assessing whether the engineered descriptors capture meaningful metallurgical information from new datasets.

Although the proposed descriptors were developed from established physical metallurgy principles, their weighting coefficients were selected for low-carbon hot-rolled steels and may require recalibration for alloys with substantially different compositions, processing routes, or strengthening mechanisms. Similarly, this study considered two widely used tree-based ensemble algorithms. Future studies could evaluate the transferability of the proposed descriptors to other steel grades, larger industrial datasets, and alternative machine learning approaches that incorporate predictive uncertainty, such as Gaussian process regression and Bayesian learning methods. Further experimental characterization, including strain-hardening and microstructural data, could help clarify the differences between the YS and UTS predictions and support the recalibration of the engineered descriptor coefficients for other steel grades. Such studies would help establish the broader applicability of physics-informed feature engineering in various metallurgical systems.

5. Conclusions

This study demonstrated that domain-knowledge-driven feature engineering can improve the machine learning prediction of the mechanical properties of low-carbon hot-rolled steel under limited data conditions. Five physically motivated composite descriptors were incorporated into tree-based machine learning models, resulting in improved predictive performance while preserving model interpretability. The main conclusions are as follows.

(1) Five physically meaningful composite descriptors, namely, carbon equivalent (CE), nitrogen-to-aluminum ratio (N/Al), microalloying efficiency index (MEI), thermal processing parameter (TPP), and solid-solution strengthening index (SSSI), were derived from the original 17 compositional and processing variables. These descriptors incorporate established metallurgical knowledge without requiring additional experimental measurements.

(2) The engineered 22-feature dataset produced the greatest improvement in the elongation prediction. For XGBoost, the mean percentage error decreased from 3.23% to 3.05%, whereas the test-set R^2 increased from 0.4935 to 0.5444, representing a 10.3% increase in explained variance. For the yield strength, Random Forest achieved the largest improvement (R^2: 0.4744 to 0.4861), whereas only marginal changes were observed for the ultimate tensile strength.

(3) Learning curve analysis showed that the engineered feature set achieved a slightly higher cross-validation R^2 at intermediate training sizes (n = 125–175), suggesting a modest improvement in sample efficiency in the most data-limited regime. Both feature sets converged to comparable plateau cross-validation R^2 values at larger training sizes.

(4) Permutation importance analysis identified MEI and TPP as two of the six most influential predictors across all three mechanical properties, confirming that these engineered descriptors provide metallurgical information complementary to that provided by the original variables. Partial dependence plots showed physically consistent responses, including a threshold-type increase in the predicted yield strength above MEI $\approx$ 8, consistent with the onset of precipitation strengthening, and a stepwise decrease with increasing TPP, reflecting the influence of the combined thermal processing on the coarsening of ferrite grains.

(5) All models achieved mean percentage errors below 3.5% for yield strength, ultimate tensile strength, and elongation, demonstrating that accurate and interpretable machine learning models can be developed for industrial steel property prediction using relatively small datasets when physics-informed feature engineering is applied.

Future work should investigate the adaptive calibration of the engineered descriptor coefficients for different steel grades, evaluate their transferability to larger industrial datasets and alternative alloy systems, and explore their integration with uncertainty-aware and physics-informed machine-learning frameworks.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/met16080933/s1. Table S1. Definitions, equations, physical basis, dataset ranges, and Pearson correlation coefficients for engineered descriptors; Table S2. Test-set performance metrics (MPE, RMSE, R^2, and MAE) for all model–feature–target combinations.

Author Contributions: Conceptualization, S.T.; methodology, H.A. and S.T.; software, J.L. and S.T.; validation, J.L. and H.A.; formal analysis and investigation, H.A., S.T. and N.P.; resources, N.P. and S.T.; data curation, H.A., J.L. and S.T.; writing—original draft preparation, S.T.; writing—review and editing, S.T. and N.P.; visualization, J.L. and N.P.; supervision, S.T. and N.P.; project administration, N.P.; funding acquisition, N.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Gyeongbuk ANCHOR system through the Gyeongbuk ANCHOR CENTER, funded by the Ministry of Education (MOE) and the Gyeongsangbuk-do, Republic of Korea (2026-ANCHOR-15-115).

Data Availability Statement: The original contributions presented in this study are included in the article and Supplementary Material. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tiwari, S.; Ahn, H.; Reddy, M.H.; Park, N.; Reddy, N.G.S. Mechanical Property Prediction of Industrial Low-Carbon Hot-Rolled Steels Using Artificial Neural Networks. *Materials* **2025**, *18*, 2966. [CrossRef] [PubMed]
2. Fang, W.; Huang, J.; Peng, T.; Long, Y.; Yin, F. Machine Learning-Based Performance Predictions for Steels Considering Manufacturing Process Parameters: A Review. *J. Iron Steel Res. Int.* **2024**, *31*, 1555–1581. [CrossRef]
3. Wang, X.; Li, X.; Yuan, H.; Zhou, N.; Wang, H.; Zhang, W.; Ji, Y. Prediction and Analysis of Mechanical Properties of Hot-Rolled Strip Steel Based on an Interpretable Machine Learning. *Mater. Today Commun.* **2024**, *40*, 109997. [CrossRef]
4. Smola, A.J.; Schölkopf, B. A Tutorial on Support Vector Regression. *Stat. Comput.* **2004**, *14*, 199–222. [CrossRef]
5. Wang, S.; Li, L.; Zhang, Y. Interpretable Prediction of Mechanical Properties in Hot Strip Rolling by Combining Machine Learning with Shapley Additive Explanations. *Processes* **2026**, *14*, 1547. [CrossRef]

6. Wu, H.; Zhang, B.; Li, Z. Small Sample-Oriented Prediction Method of Mechanical Properties for Hot Rolled Strip Steel Based on Model Independent Element Learning. *IEEE Access* **2024**, *12*, 197300–197311. [CrossRef]
7. Ishtiaq, M.; Tiwari, S.; Nagamani, M.; Kang, S.-G.; Reddy, N.G.S. Data-Driven ANN-Based Predictive Modeling of Mechanical Properties of 5Cr-0.5Mo Steel: Impact of Composition and Service Temperature. *Crystals* **2025**, *15*, 213. [CrossRef]
8. Grinsztajn, L.; Oyallon, E.; Varoquaux, G. Why Do Tree-Based Models Still Outperform Deep Learning on Typical Tabular Data? In Proceedings of the 36th International Conference on Neural Information Processing Systems, New Orleans, LA, USA, 28 November–9 December 2022.
9. Valentim, C.A.; David, S.A. From Data to Physics: Physics-Informed Machine Learning Frameworks in Interdisciplinary Applications. *Dynamics* **2026**, *6*, 16. [CrossRef]
10. Bhadeshia, H.K.D.H. Neural Networks and Information in Materials Science. *Stat. Anal. Data Min. ASA Data Sci. J.* **2009**, *1*, 296–305. [CrossRef]
11. Mukherjee, M.; Singh, S.B. Artificial Neural Network: Some Applications in Physical Metallurgy of Steels. *Mater. Manuf. Processes* **2009**, *24*, 198–208. [CrossRef]
12. Yurioka, N.; Suzuki, H. Determination of Necessary Preheating Temperature in Steel Welding. In *Engineering, Materials Science*; The National Academies of Sciences, Engineering, and Medicine: Washington, DC, USA, 1983.
13. Pickering, F.B. *Physical Metallurgy and the Design of Steels*; Illustrated, Reprint; Applied Science Publishers: Barking, UK, 1978.
14. Gladman, T. *The Physical Metallurgy of Microalloyed Steels*; Institute of Materials: London, UK; Taylor & Francis Group: Abingdon, UK, 2002.
15. Deardo, A.J. Niobium in Modern Steels. *Int. Mater. Rev.* **2003**, *48*, 371–402. [CrossRef]
16. Xue, J.; Zhao, Z.; Bin, C.; Liu, X.; Wu, H.; Li, H.; Xiong, W. Effects of Rolling and Coiling Temperature on the Microstructure and Mechanical Properties of Hot-Rolled High Strength Complex Phase Steel. *Mater. Res. Express* **2019**, *6*, 0965c8. [CrossRef]
17. Mandal, A.; Ghosh, A.; Chakrabarti, D.; Davis, C. Effect of Coiling Temperature on Impact Toughness of Hot Rolled Ultra-High-Strength Multiphase Steel Strips. *Mater. Sci. Eng. A* **2021**, *824*, 141796. [CrossRef]
18. Sellars, C.M.; Whiteman, J.A. Recrystallization and Grain Growth in Hot Rolling. *Met. Sci.* **1979**, *13*, 187–194. [CrossRef]
19. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
20. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.