Machine Learning with Applications 20 (2025) 100634

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

Machine learning-driven predictive modeling of mechanical properties in diverse steels

Movaffaq Kateb [a,*], Sahar Safarian [b]

[a] Structural Chemistry Division, Department of Chemistry-Ångström Laboratory, Uppsala University, Lägerhyddsvägen 1, P.O. Box 538, 75121, Uppsala, Sweden
[b] IVL Swedish Environmental Research Institute, Aschebergsgatan 44, 41133, Gothenburg, Sweden

ARTICLE INFO

Keywords:
Machine learning
Mechanical properties
Small dataset
Steel compositions
Multiple grades

ABSTRACT

This study explores the application of machine learning (ML) in steel design using a small dataset of various steel grades that include 13 key elements and three critical mechanical properties. Random forest (RF) models were systematically evaluated for their robustness and effectiveness in predicting the stress-strain of steel properties. Moreover, other alternative approaches, such as support vector machines, extreme gradient boosting machines, and artificial neural networks, were also evaluated to ensure that the predictions made by the RF model are as accurate as possible. To assess the bias-variance trade-off, 1-seed and random 100-seeds with 80/20 train/test split, and leave-one-out cross-validation for all datasets were conducted. The results demonstrated that the RF models are accurate and reliable, achieving low bias and variance while delivering predictions comparable to, and in some cases better than, those obtained in studies with larger datasets. The analysis revealed a trade-off between strength and ductility, with elongation negatively correlated with yield strength and ultimate tensile strength. This study highlights the feasibility of using small, realistic datasets to develop effective ML models for predicting mechanical properties in steel design. The methodology can also be readily extended to investigate processing-property relationships in other systems, offering a versatile approach for advancing materials science through data-driven techniques.

1. Introduction

Steel variants are the most widely used materials in modern engineering and industrial applications due to their tunability for various demands. Steel probably has more classification than any other engineering material, e.g. based on the carbon content/alloying element, production method, finishing process, phases, mechanical strength, heat treatment, and quality (Alliance Steel, n.d.; Metal Supermarkets, n.d.). Thanks to its combined cost-strength effectiveness, steel is undeniably an important element of many infrastructures such as buildings, railroads, bridges, pipelines, and power plants (Quaranta & Davies, 2022; Safarian, 2023). The oil industry and chemical plants relied on steel for their transport, processing, and storage units while other sectors like shipbuilding, automotive, machinery/tools, appliance, and food packing consume a variety of steels.

While blacksmiths of the pre-industrial age lacked the knowledge of alloying, they intentionally manipulated carbon content and hammered out slag to achieve desired mechanical properties. However, they sometimes accidentally introduced chromium, manganese, nickel, silicon, phosphorus, or sulfur impurities through the ore or charcoal. Nowadays, steels are *deliberately* made with varying amounts of alloying elements and the specific composition greatly influences their properties. Alongside alloy content, one can affect microstructure through several treatments and achieve a new set of properties. Materials engineers efficiently manipulate the physical properties of a single composition through grain-size (Kateb, Hajihoseini, et al., 2018; Kateb, Jacobsen, et al., 2018; Kateb & Ingvarsson, 2017), phases (Kateb et al., 2019; Kateb & Ingvarsson, 2021) and texture (Hajihoseini et al., 2018, 2019). For instance, many high-carbon or high-alloy steels (Diehl Tool Steel, n.d.) annealed to improve their machinability and hardened afterward by a heat treatment to gain their initial properties. However, the choice of treatment depends on the existing elements, meaning these two are strongly correlated. The latter has been verified tanks to data analysis by machine learning (ML) (Guo et al., 2019; Guo & Sha, 2004). Thus, the presence of certain elements by design indicates target mechanical properties.

Uncovering hidden correlations by ML can assist optimization problems, especially in high-dimensional parameter space. For instance,

* Corresponding author.
E-mail addresses: movaffaq.kateb@kemi.uu.se (M. Kateb), Sahar.Safarianbana@ivl.se (S. Safarian).

https://doi.org/10.1016/j.mlwa.2025.100634
Received 10 February 2025; Accepted 27 February 2025
Available online 28 February 2025

Table 1
Statical values of alloying elements in steel samples. All values are in weight percent unit.

Element	C	Mn	Si	Cr	Ni	Mo	V	N	Nb	Co	W	Al	Ti
Min	0	0.01	0.01	0.01	0.01	0.02	0	0	0	0.01	0	0.01	0
Max	0.43	3	4.75	17.5	21	9.67	4.32	0.15	2.5	20.1	9.18	1.8	2.5
Mean	0.0964	0.1462	0.2212	8.0438	8.1840	2.7660	0.1837	0.0055	0.0354	7.008	0.1612	0.2391	0.31089
Std Dev	0.1088	0.396	0.579	5.417	6.326	1.829	0.451	0.018	0.161	6.244	0.918	0.339	0.5557

Diao et al. (2022) highlighted the strength-ductility trade-off and studied the correlation between 13 descriptors and several mechanical properties for 97 carbon steels. They chose Cr, Ni, and C as well as normalizing, quenching, and tempering temperatures to train a model for yield strength (YS) and elongation (El). Similarly, Wang et al. (2020) considered highly correlated features in 60 low activation ferritic-martensitic (LAFM) steel, and utilized random forest (RF) to model YS and El based on the key features, tempering time and temperature along with C, Cr, W, Ti, and B content, to design optimum strength-ductility. Xiong et al. (2020) studied mechanical properties of 360 low carbon and low alloy steels. They found that tempering temperature and C, Cr, and Mn, respectively, are the most influential features on the mechanical properties. They also considered several approaches and observed the best performance using RF. Unlike others, Xie et al. (2021) studied more than 11000 datasets, obtained during the hot-rolling of several steel grades. The latter, including several steel grades, had both technological and scientific interests. They compared the performance of the deep neural network (DNN) against support vector machines (SVM), *k*-nearest neighbors (*k*NN), and RF. Using the 5-fold cross-validation with 80/20 splits for train/test, they observed low accuracy for SVM due to its limitation for handling high-dimensional datasets i.e. 27 descriptors in this case. However, they noticed a slight difference between *k*NN, RF, and DNN all being acceptable. The latter is in agreement with the general belief that a large dataset can produce meaningful results. On the other hand, there exist reports on small datasets even using SVM but they reached higher accuracy because of introducing intermediate parameters called physical metallurgy-guided approach (Shen et al., 2019). This is an indication of the fact small datasets can be useful and produce meaningful results provided that they are utilized in narrow space parameters while larger datasets may provide a larger window and remain accurate.

Drawing inspiration from previous works on the ML-assisted analysis of steel, this study aims to develop an ML model capable of accurately predicting the mechanical properties of a few steel grades focusing on small datasets. This essentially means we are considering a larger space parameter window than earlier studies that were focused on a single grade. We also took advantage of different grades as small datasets for single grades are prone to biased sampling. As a first step, an ML model based on the random forest (RF) algorithm is developed. Random Forest is selected because of its ability to handle complex, non-linear relationships and its robustness in avoiding overfitting, making it ideal for the challenges associated with predicting mechanical properties from compositional data. The RF model is initially trained and then validated using a regression approach, which allows for an assessment of its ability to accurately predict mechanical properties such as YS, tensile strength (UTS), and El based on the steel composition. In addition, the study investigates the impact of various alloying elements and their concentration on the mechanical properties of steel. By analyzing the relationships between different steel compositions and the corresponding output properties, the study aims to gain deeper insights into how each element influences key characteristics of steel. This step will allow for a better understanding of how the input features interact to produce desired material properties and optimize steel design for specific applications. Moreover, to ensure that the predictions made by the RF model are as accurate as possible, ML models based on alternative approaches, such as SVM, extreme gradient boosting machines (XGB) (Chen & Guestrin, 2016), and artificial neural networks (ANN) (Safarian, Ebrahimi Saryazdi, et al., 2021; Safarian et al., 2020; Safarian, Saryazdi, et al., 2021), are also developed. These models will be trained and validated using the same dataset to allow for a comprehensive comparison of performance. By comparing the results of each approach, the study seeks to identify the model that yields the highest accuracy in predicting the mechanical properties of steel, helping to choose the most effective method for future applications in material design and production. Ultimately, this study aims to contribute to the growing body of knowledge on applying ML to materials science, offering a practical tool for optimizing steel production and predicting material behavior.

Table 2
Statical values of mechanical properties in steel samples.

Property	YS (MPa)	UTS (MPA)	El (%)
Min	1005.9	1019	2
Max	2510.3	2570	35
Mean	1420.998	1641.653	14.00726
Std Dev	301.41	345.9195	5.087558

2. Methodology

2.1. Data description

Data on 312 distinct steel samples was utilized for this study, to develop the ML model. The dataset includes experimental measurements of strength ductility, which were extracted and carefully processed from the Citrine dataset (Bajaj, 2017). The data cleaning process involved de-duplication and consistency checks to ensure the reliability and uniformity of the records, providing a robust foundation for model training and validation. The Citrine dataset is a highly regarded resource in materials science that aggregates data from a wide array of sources, including published experimental studies, industrial reports, and proprietary research. It ensures high accuracy and validity by integrating data that meets strict experimental and reporting standards. Each entry in Citrine is linked to verified sources, and the dataset undergoes rigorous quality control processes to identify and reduce inaccuracies. This makes it an asset for predictive modeling in materials science, including applications in compositional relationships and steel properties. Its comprehensive scope and careful management increase the reliability of any derived ML models.

The dataset used in this study includes information on 13 key alloying elements found in steels, namely: aluminum, carbon, cobalt, chromium, manganese, molybdenum, nitrogen, nickel, niobium, silicon, titanium, vanadium, and tungsten. These elements play vital roles in determining the mechanical properties of steel. In addition, the dataset provides a list of stress-strain mechanical properties for each steel sample. These include ultimate UTS, YS, and El, which are key indicators of steel's performance and usability across various applications (Diao et al., 2022; Guo et al., 2019; Guo & Sha, 2004; Shen et al., 2019; C. Wang et al., 2020; Xie et al., 2021; Xiong et al., 2020). By combining detailed compositional data with a diverse set of experimentally measured mechanical properties, the dataset provides a robust basis for ML analysis.

The dataset's compositional and mechanical property statistics are summarized in two tables. Table 1 presents the minimum, maximum, and mean values for the 13 alloying elements included in the steel samples. These values provide an overview of the compositional range

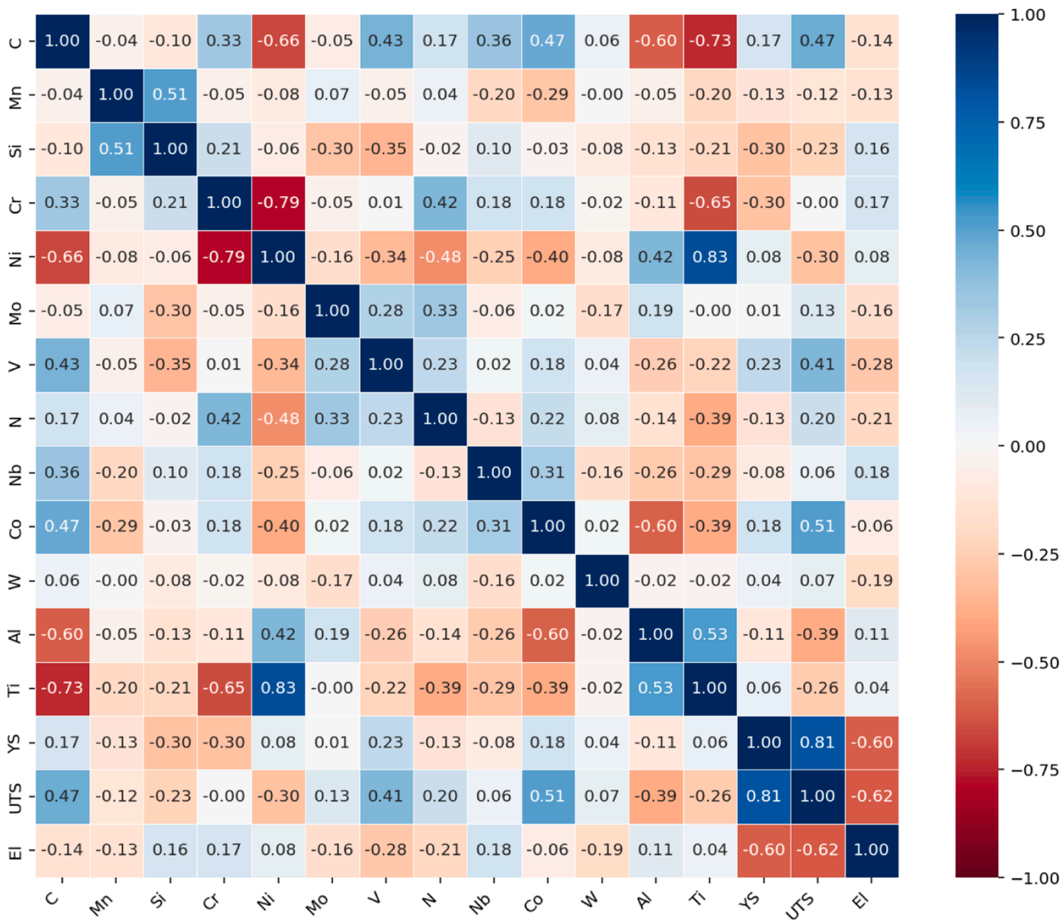

Fig. 1. The heatmap of correlation between composition and mechanical properties UTS, YS, and El. The color scale on the right represents the range of Spearman's rank correlation coefficient values.

across the dataset and highlight the variability in element concentrations. As mentioned earlier we do not focus on single grade rather consider stainless steel (austenitic), tool steel, maraging steel, and high-strength low-alloy steel grades. Similarly, Table 2 details the minimum, maximum, and mean values for the three mechanical properties of the steel samples. This table also provides an understanding of the performance range within the dataset, demonstrating the mechanical diversity of the steel samples used in the study.

2.2. Machine learning model development

In this study, ML models were developed to predict the mechanical properties of various steels. Regression models were also used for predicting mechanical properties, which are numerical. Popular supervised ML algorithms, including RF, ANN, SVM, and XGB, were evaluated as potential candidates. To validate and check the prediction ability of the models, the database was randomly divided into 80/20 train/test splits.

A comprehensive hyperparameter tuning process was implemented for each model, utilizing a combination of random search and grid search strategies. The random search initially explored a wide range of potential hyperparameter values by randomly sampling across a defined search space. The insights obtained from this phase helped refine the parameter range, enabling the use of grid search to fine-tune the hyperparameters further. Model performance during this tuning was evaluated using 5-fold cross-validation within the training dataset. This approach substantially improved prediction accuracy for all models, except for RF which demonstrated strong performance with its default hyperparameters, showing robustness against changes in hyperparameters. This ease of implementation, combined with its reliability and effectiveness, established RF as the preferred base method for predicting material properties.

The RF algorithm, capable of handling regression tasks, was applied to predict mechanical properties. To address the potential impact of random splits in data, models were first tested with a single random seed that ensured a representative split of the dataset. Subsequently, performance robustness was validated across 100 different random seeds, comparing the mean and standard deviation of evaluation metrics on unseen test sets. Additionally, leave-one-out cross-validation (LOOCV) was employed to further validate the RF models. LOOCV trains the model on all but one sample, using the excluded sample for testing, and iteratively performs this process across the entire dataset. The combination of multiple-seeds validation and LOOCV ensured robust and reliable results, solidifying RF's effectiveness as the primary model for predicting alloy properties.

3. Results and discussion

3.1. Correlation between inputs and outputs

For this study, 312 steel experimental data points across major components with crucial mechanical properties were gathered and cross-validated. The thirteen main steel elements: Al, Si, V, C, Mn, N, Nb, Cr, Ni, Mo, Co, Ti, and W were used to create these steels and three mechanical properties were contained and described as continuous variables in the dataset: UTS, YS, and El. The statistical relationships between different compositions and mechanical properties were examined by using Spearman correlation as shown in Fig. 1. This analysis highlights several valuable rank correlations among the variables. El

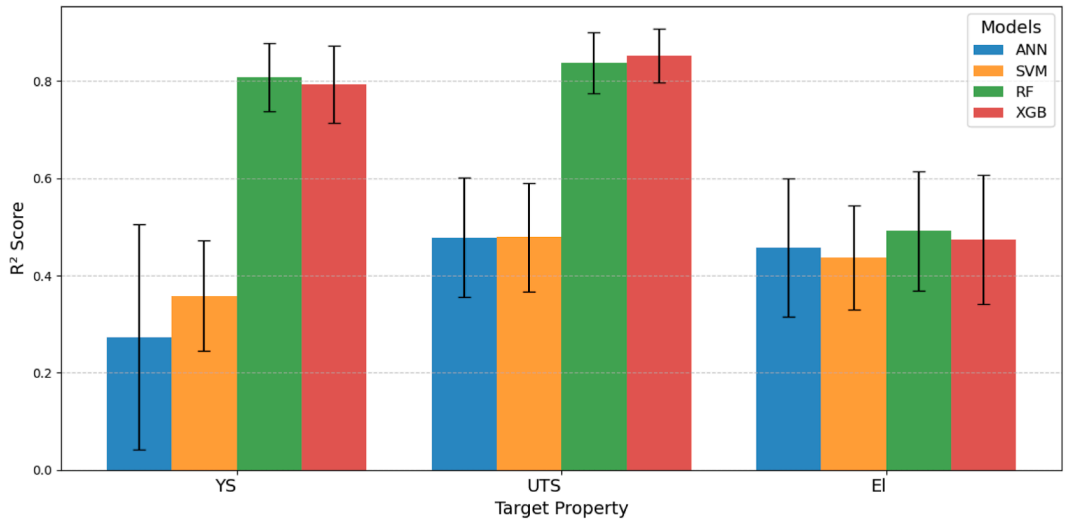

Fig. 2. Comparison of RMS scores for ML models for YS, UTS, and El. The error bars indicate the range of scores over 100 random splits (100-seeds).

shows a considerable negative correlation with YS and UTS. Diao et al. (2022) divided between features and target properties and Wang et al. (2020) only presented features correlation while others neglected to report such a figure. Nonetheless, Diao et al. (2022) observed a similar negative correlation between YS and El. It is worth mentioning that the correlation is not 1 to -1 as observed in e.g. coldwork experiment or as high as others predicted. This is most likely because we incorporated different steel grades in our training data set and stayed away from biases. The UTS and YS exhibit a strong positive correlation with each other, as indicated by the dark blue.

Additionally, except for Mn and to some extent Nb and Ti, which will be discussed later in the feature importance, every other element that is positively correlated with UTS and YS presents a negative correlation with El and vice versa. This is because these elements such as carbon, vanadium, and cobalt contribute to strengthening mechanisms by e.g. solid solution strengthening, precipitation hardening, and increased martensite formation, which enhance YS and UTS. However, these same mechanisms typically reduce ductility, as the steel becomes harder and less capable of undergoing plastic deformation, leading to a trade-off between strength and El. Silicon and chromium are negatively correlated with YS and UTS, indicating that higher levels of these elements cause softening. This is because silicon promotes softer ferritic microstructures, and chromium can form phases like ferrite or carbides that may not enhance YS.

3.2. Machine learning model selection and evaluation

By obtaining the notable rank correlations presented in Fig. 1, the potential of developing ML models to predict these properties from compositions was explored to achieve a reasonable bias-variance trade-off. The bias-variance trade-off is a concept in ML in which increasing bias reduces variance and vice versa. Bias errors result from incorrect assumptions in the learning algorithm; high-bias methods often generate simpler models that fail to capture the relationships between input and output streams. Variance error arises from a model's sensitivity to small variations in the training data. High-variance algorithms may fit the training data well but struggle to adapt to unseen datasets. Although the inherent trade-off between bias and variance, the goal is to develop an ML model that achieves a balanced approach to a given problem.

In this work, the ML algorithms were trained using 80% of the dataset, while the remaining 20% was reserved for testing on unseen data to evaluate model performance. The training and test sets were randomly split to ensure both subsets were representative of the entire dataset, preserving similar data distributions. While this approach is effective for large datasets, smaller datasets, like the one used in this study, can lead to biased splits where the training and test distributions differ significantly. Such imbalances may result in models that perform well on the training data but fail to generalize effectively to the test data.

To ensure the strength of the established ML models and account for the effects of randomness, three evaluation strategies were employed: (1) training and testing with a single random train-test split using a fixed seed to maintain similar train/test sets (1-seed), (2) conducting 100 random train-test splits using 100 unique seeds to analyze model consistency (100-seeds), and (3) utilizing LOOCV to evaluate performance across the entire dataset. The 100-seeds analysis provided comprehensive insights into model bias and variance by evaluating performance across multiple randomized splits. In the LOOCV experiments, each model has been trained on the dataset while excluding one sample as the test case. This process was carried out for all 312 data points, resulting in the creation of 312 models, each generating predictions for its respective unseen test sample. The overall model performance was evaluated by calculating the RMS across the entire dataset. LOOCV is a well-established method for providing reliable and unbiased performance estimates, especially for small datasets, as it ensures nearly the entire dataset is utilized for training during each iteration. Regression metrics, including RMS, were also used to assess the accuracy of the models in predicting mechanical properties.

In this investigation, several prominent ML algorithms on a small dataset were evaluated: ANN, SVM, RF, and XGB. Modeled on the human brain and neural systems, ANN is a computational learning technique that typically requires large datasets for robust results. SVM works by fitting an optimal line to the data and handles both linear and nonlinear problems effectively. RF algorithm builds a set of decision trees by randomly selecting subsets of inputs, requiring minimal data preprocessing, and efficiently handling diverse feature scales. By combining multiple trees, RF reduces prediction variance. XGB, an advanced version of gradient boosting, improves the speed and performance of successive decision tree ensembles.

Hyperparameter tuning for each model was carried out using 5-fold cross-validation within the training set for each property. The prediction performances of hyperparameter-optimized models were analyzed and shown in Fig. 2. This figure compares the performance of various ML models (ANN, SVM, RF, and XGB) in predicting mechanical properties (yield strength, tensile strength, and elongation), as measured by root mean square (RMS) scores. Error bars represent the variability in RMS scores from the 100-seeds experiments. RF and XGB exhibit similar and robust performance with higher RMS scores for YS and UTS, whereas ANN displays lower RMS scores and higher variability, particularly for

YS, indicating less reliability. These observations are consistent with prior studies on the comparative performance of ANN, SVM, and RF.

We would like to remark that the El score obtained here is lower for all methods meaning it is not an issue of the RF or XGB method. One may suspect combining datasets of various grades as the main reason for low RMS scores. However, we believe that El is not only composition dependent and process parameters must be included to achieve high accuracy.

It is worth noting that random forest is known for its robustness to hyperparameter tuning, often delivering strong performance using default settings without requiring extensive adjustments. As a result, the RF approach with default hyperparameters was chosen as the baseline method in this study.

Three random forest models, using default hyperparameters, were ultimately established to predict each of the three mechanical properties based on the compositional elements. The models were evaluated using three different cross-validation methods: LOOCV, a single random seed, and 100 random seeds as shown in Fig. 3. The results reveal consistent performance for yield strength and tensile strength across all three cross-validation methods, with RMS scores exceeding 0.8. In contrast, elongation shows relatively lower RMS scores for all validation methods, reflecting reduced predictive accuracy. The 1-seed method showed better performance compared to the average of the 100-seeds approach. This improved performance can likely be attributed to the single-seed split, which created an ideal scenario for training the ML models by maintaining a similar data distribution between the training and test sets in terms of both input and output variables. LOOCV systematically trains the model on all but one data point and tests on the excluded point, repeating this process for every data point in the dataset. This ensures that every data point is used for both training and testing, leading to a more robust and unbiased estimate of model performance. In contrast, other approaches depend on random splits, which may introduce variability or bias due to unequal representation of data distributions in the training and testing sets. LOOCV minimizes this risk and provides a more reliable assessment, particularly for smaller datasets.

3.3. Prediction of mechanical properties from compositions

Fig. 4 illustrates the scatter plots for the relationship between predicted and experimental values for (a) yield strength, (b) tensile strength, and (c) elongation obtained from the leave-one-out cross-validation (LOOCV) experiments. Each point on the plots represents a sample, while the dashed diagonal line indicates perfect agreement between predicted and experimental values. The RMS values displayed on each plot quantify the predictive accuracy of the models, with higher values indicating better performance. For YS and UTS, the RMS values are 0.84 and 0.85, respectively, demonstrating strong agreement between the predicted and experimental results. This indicates that the ML models were able to effectively capture the underlying patterns in the data for these properties. However, for EL, the RMS value is comparatively lower at 0.52, showing the model struggled to accurately predict this property. This reduced accuracy may arise from the greater inherent variability of EL or its weaker correlation with compositional.

The RMS value of 0.85 suggests that while elemental composition

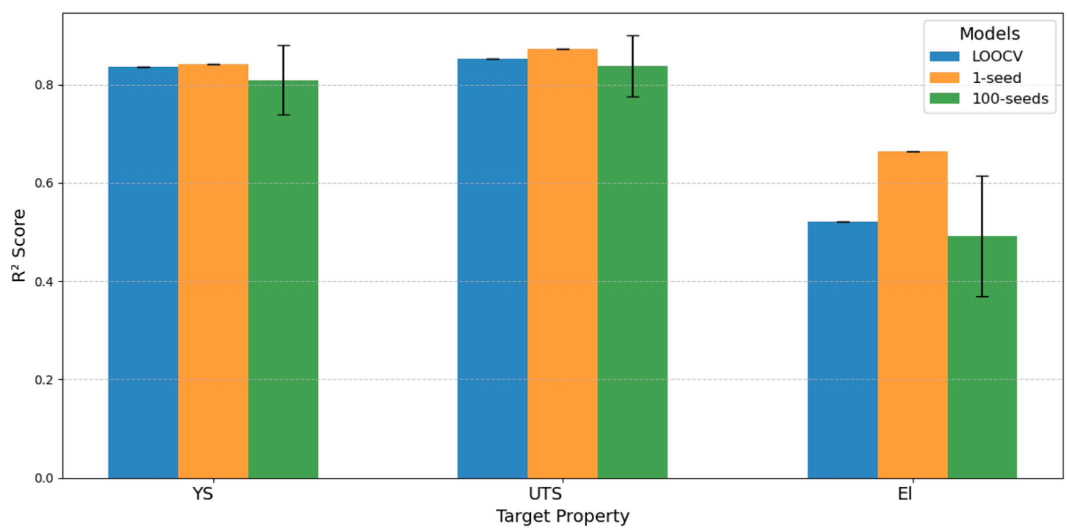

Fig. 3. Comparison of RMS scores for RF models using different cross-validation strategies for three mechanical properties.

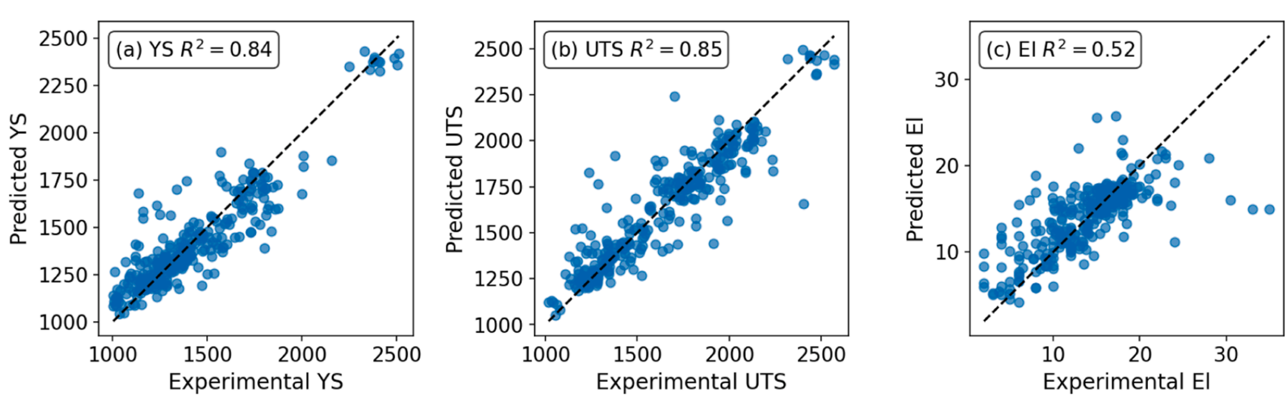

Fig. 4. Scatter plots comparing predicted and experimental values for mechanical properties using LOOCV: (a) YS, (b) UTS, and (c) El.

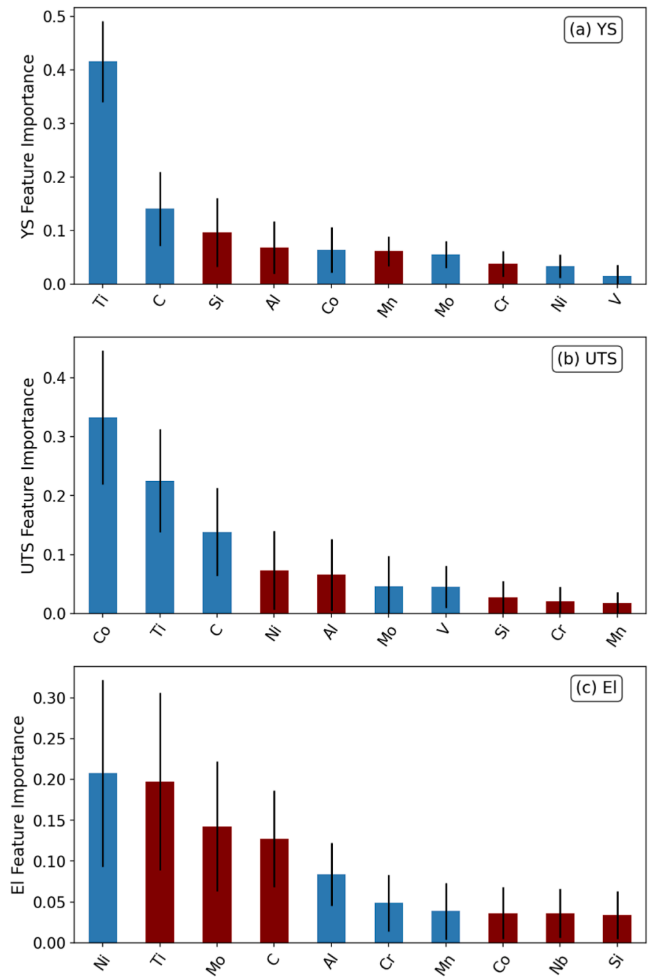

Fig. 5. The RF model's top 10 most influential features for each mechanical property were ranked based on their importance scores: (a) TYS, (b) UTS, and (c) El. The red and blue bars represent negative and positive impacts corresponding to compositions. Thin black lines indicate the error bars

strongly influences the mechanical properties of steel, other factors such as phase and microstructure also play significant roles. These include the distribution and morphology of phases like ferrite, martensite, or bainite, grain size, and the presence of inclusions or defects. Processes like heat treatment further alter microstructures, impacting properties like toughness, strength, and ductility. Thus, the deviation from 1 highlights the complexity of steel behavior, where composition alone cannot fully account for the mechanical response.

To gain deeper insights into how each input (compositions) influences the outputs (mechanical properties), a feature importance analysis was conducted. Fig. 5 presents the feature importance for YS, UTS, and El, with blue bars indicating features that have a positive impact on targets and red bars representing features that have a negative impact. In general, it is a measure of feature contribution to the accuracy of the model and does not distinguish whether a feature is positively or negatively correlated with the target. In RF, relationship between the feature and target can be non-linear meaning depending on the value of the feature and interaction with the rest of the feature it can increase or decrease the target value. Here, we first used RF to determine feature importance and then used linear correlation to determine its sign. One may think utilizing a linear correlation to determine the sign is inaccurate. We regret that in the presence of a non-linear relation, it can be inaccurate. However, we believe this is very important for two reasons: (I) It is a common practice to build a model based on some of the important features. For instance, Wang et al. (2020) only included 7 important features (out of 19) in building their model. They rationalized this with the necessity of having considerably fewer features than the samples (originally 60) in the dataset. Thus, determining a sign allows balancing features with positive and negative impacts. (II) The features sign along with the range of features, e.g. min and max composition, can be used by the experts in the field to verify the ML model. As an example, certain elements may be beneficial up to a limit and deteriorate target property afterward. Experts in the field are aware of that and considering the element content they can verify whether or not the model's prediction can be verified.

As shown in Fig. 5a, titanium shows the highest positive impact on YS, leading to improving mechanical properties through austenite grain refinement and precipitation hardening (Baker, 2019; Wang et al., 2022; Zaitsev & Arutyunyan, 2021) in low-carbon steel i.e. due to its ability to form small, finely distributed carbide precipitates that are pinning grain boundaries and dislocations and enhance strength. It is worth noting that, Ti can also form nitride but we observe that N is unimportant which can be true considering its low content (<0.15 wt%). As expected, carbon also has a positive impact, improving the YS. On the other hand, silicon and aluminum content reduces YS. At concentrations above approximately 1.5 wt% Si can improve YS and below that the change are negligible while higher content cause embrittlement in certain steels (Girault et al., 2001; Pang et al., 2017). This probably because Si is a strong ferrite-stabilizers and in low-carbon steels it will effectively reduce harder phases such as martensite or perlite. A predominantly ferritic matrix is softer and thus has lower YS compared to mixed-phase steels. For aluminum this the improvement is a few times less than Si and more than 4 wt% is required to see the effect. High Cr content in ferritic and martensitic steels can lead to grain coarsening if it is not compensated by a refining element such as vanadium. Cr can form Cr_7C_3 or $Cr_{23}C_6$ carbides, which might reduce overall toughness and localized strength. Larger grains and fewer grain boundaries reduce dislocations pinning and reduction in YS. Diao et al. (2022) observed Cr as the most important feature for YS. They concluded that Cr participates in carbide formation which improves the YS. However, they considered the carbon steel dataset with Cr 0.01-1.12 wt% while noticeable chromium carbide formation occurs at chromium concentrations around 5 wt% or higher (Funakawa & Ujiro, 2010). This implies that the formation of chromium carbides is minimal or non-existent in steels with chromium content near and below 1 wt%. Wang et al. (2020) also found Cr to be the most important element besides carbon. In their dataset, Cr content ranged between 4.61 to 9.3 wt% which is consistent with experimental results (Funakawa & Ujiro, 2010). The effect of Mn (<3.0 wt%) can be more complex due to a stronger correlation between mechanical properties and heat treatment. However, in agreement with our result, it has been shown that in the low-carbon Ni-rich steel, YS decreases by increasing Mn from 2.0 to 3.0 wt% (Zhan et al., 2024).

Fig. 5b shows the features' importance for UTS is affected by the same elements as in YS. The difference in the order of feature importance between them can be explained by the distinct mechanisms and material behaviors that govern these properties. While both YS and UTS depend on the microstructure and alloying elements of steel, they are influenced differently by specific features. In particular, YS is defined by the stress at which the material begins to plastically deform, dominated by mechanisms that resist dislocation motion e.g. precipitations, grain boundaries, and phases. On the other hand, UTS is the maximum stress the material can sustain before necking, influenced by the material's strain-hardening capability, ductility, and resistance to fracture. For instance, it appears that Co suppresses carbide coarsening and enlarges the strain hardening window giving rise to UTS. Ni encourages austenite, which is softer than ferrite but it improves uniformity that delays yielding. However, stabilized austenite by Ni may reduce UTS by resisting strain-induced phase transition (e.g. to martensite). Ni has also been found to be responsible for embrittlement in certain steel grades.

In the case of El (Fig. 5c), we observed similar features except that Nb replaced V in the list. As mentioned above Ni stabilizes the softer

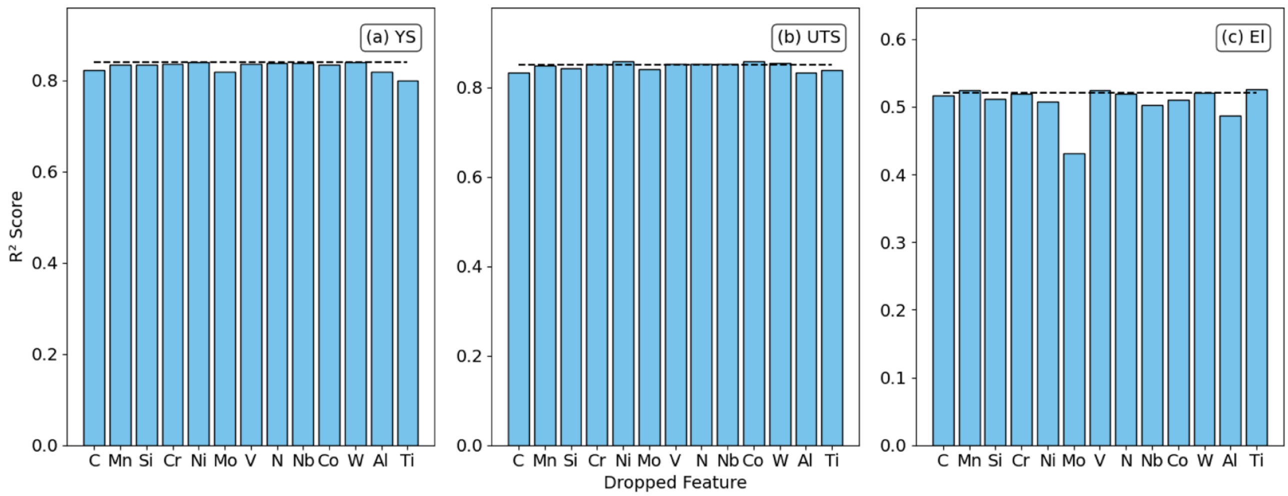

Fig. 6. The RF model's score versus LOFOCV for each element for targets (a) TYS, (b) UTS, and (c) El. The dashed lines indicate a score without LOFOCV.

austenite phase that effectively improves El. Elements such as Ti, Mo, and C which contributed to improving YS and UTS through precipitate hardening, now negatively affect ease of plastic deformation. On the other hand, Al, Cr, and Mn that reduced YS, have a positive impact on El.

Furthermore, we performed leave-one-feature-out cross-validation (LOFOCV) i.e. for each feature in the dataset, leave it out, train the model using the remaining features, and compute the performance metric on the validation set. Thus, features whose removal causes the least drop in performance are less important and vice versa. This method is less popular since it is more computationally expensive and does not favor granular features that have low predictive power individually but collectively have meaningful contributions. However, it is feasible for small datasets, applicable to all methods, and provides further insights into the importance of each feature in predicting the outcome. It can also be used to determine signs of the feature's importance depending on the increase or reduction of the score.

The LOFOCV results are shown in Fig. 6 for different target properties. It can be seen that for the case YS (Fig. 6a), features Ti, Al, Mo, and C have the highest effect on the R^2 score of YS meaning Si, Co, and Mn are not as important as Fig. 5a showed earlier. For the UTS shown in Fig. 6b, elements C, Al, Mo, and Ti are the key players in the model accuracy. Surprisingly, leaving out Co and Ni, 1st and 4th important features in Fig. 5b, slightly improves the model accuracy. Such a behavior is more than a change in the order and contradicts earlier feature importance. Finally, as shown in Fig. 6c, the El score shows a strong dependence on the Mo, Al, and Nb and to a lesser extent on Ni, Co, and Si. It can be seen that leaving out Ti, V, and Mn have a slight effect on the score improvement. Aside from changes in the order, improving accuracy by excluding Ti is surprising. Overall, LOFOCV results highlight the importance of Mo and Al in terms of model accuracy in particular for the prediction of El.

4. Conclusions

This study utilized a dataset of 312 steel experimental data points, encompassing 13 key elements and three critical mechanical properties, YS, UTS, and El, to investigate the application of ML in steel design. The ML models were systematically evaluated, with particular emphasis on their robustness and effectiveness in predicting a complete set of steel properties. In this study, 1-seed, 100-seeds, and LOOCV cross-validations were conducted to examine the bias-variance trade-off in the developed ML models for predicting test data points. The results showed that random forest models were both proper and reliable in predicting the full set of steel properties, exhibiting generally low bias and variance. We have shown that supervised ML can present high performance in a large parameter space obtained by mixing several steel grades and still provide meaningful analysis neglecting process parameters introduced by others for handling small datasets. Moreover, the predictions achieved in this work are comparable to and better than, those from previous studies utilizing larger datasets with different steel grades.

In summary, this study demonstrates that ML models, like RF, can effectively predict a full set of properties, including continuous mechanical properties, for steel variants using a realistic small dataset. Furthermore, the methodology employed here can be readily adapted to investigate processing-properties relationships in other datasets as RF is known to handle categorical properties.

CRediT authorship contribution statement

Movaffaq Kateb: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Sahar Safarian:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

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.

Data availability

The data is avaialble in link provided in the references.

References

Alliance Steel. (n.d.). *What Are The Four Main Types of Steel?* | Alliance Steel. Retrieved January 18, 2025, from https://www.alliancesteel.net/what-are-the-four-main-types-of-steel/.

Bajaj, S. (2017). *Mechanical properties of some steels.* https://citrination.com/datasets/153092/show_files/.

Baker, T. N. (2019). Titanium microalloyed steels. *Ironmaking & Steelmaking, 46*(1), 1–55. https://doi.org/10.1080/03019233.2018.1446496

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).

Diao, Y., Yan, L., & Gao, K. (2022). A strategy assisted machine learning to process multi-objective optimization for improving mechanical properties of carbon steels. *Journal of Materials Science & Technology, 109*, 86–93. https://doi.org/10.1016/J.JMST.2021.09.004

Funakawa, Y., & Ujiro, T. (2010). Tensile Properties of Chromium-bearing Extra Low Carbon Steel Sheets. *ISIJ International, 50*(10), 1488–1495. https://doi.org/10.2355/ISIJINTERNATIONAL.50.1488

Girault, E., Mertens, A., Jacques, P., Houbaert, Y., Verlinden, B., & Van Humbeeck, J. (2001). Comparison of the effects of silicon and aluminium on the tensile behaviour of multiphase TRIP-assisted steels. *Scripta Materialia, 44*(6), 885–892. https://doi.org/10.1016/S1359-6462(00)00697-7

Guo, S., Yu, J., Liu, X., Wang, C., & Jiang, Q. (2019). A predicting model for properties of steel using the industrial big data based on machine learning. *Computational Materials Science, 160*, 95–104.

Guo, Z., & Sha, W. (2004). Modelling the correlation between processing parameters and properties of maraging steels using artificial neural network. *Computational Materials Science, 29*(1), 12–28.

Hajihoseini, H., Kateb, M., Ingvarsson, S.Þ., & Gudmundsson, J. T. (2019). Oblique angle deposition of nickel thin films by high-power impulse magnetron sputtering. *Beilstein Journal of Nanotechnology*. https://doi.org/10.3762/bjnano.10.186

Hajihoseini, H., Kateb, M., Ingvarsson, S., & Gudmundsson, J. T. (2018). Effect of substrate bias on properties of HiPIMS deposited vanadium nitride films. *Thin Solid Films, 663*, 126–130. https://doi.org/10.1016/J.TSF.2018.06.060

Kateb, M., Gudmundsson, J. T., & Ingvarsson, S. (2019). Effect of atomic ordering on the magnetic anisotropy of single crystal Ni80Fe20. *AIP Advances, 9*(3).

Kateb, M., Hajihoseini, H., Gudmundsson, J. T., & Ingvarsson, S. (2018). Comparison of magnetic and structural properties of permalloy Ni80Fe20 grown by dc and high power impulse magnetron sputtering. *Journal of Physics D: Applied Physics, 51*(28), Article 285005.

Kateb, M., & Ingvarsson, S. (2017). Thickness-dependent magnetic and magnetoresistance properties of permalloy prepared by field assisted tilt sputtering. In *2017 IEEE Sensors Applications Symposium (SAS)* (pp. 1–5).

Kateb, M., & Ingvarsson, S. (2021). Correlation of uniaxial magnetic anisotropy axes and principal resistivities in polycrystalline ferromagnetic films. *Journal of Magnetism and Magnetic Materials, 532*, Article 167982.

Kateb, M., Jacobsen, E., & Ingvarsson, S. (2018). Application of an extended van der Pauw method to anisotropic magnetoresistance measurements of ferromagnetic films. *Journal of Physics D: Applied Physics, 52*(7), 75002.

Diehl Tool Steel, Inc. (n.d.). *Effects of common alloying elements in steel*. Last seen 13 Jan 2025 https://www.diehlsteel.com/technical-information/effects-of-common-alloying-elements-in-steel/.

Metal Supermarkets. (n.d.). *What Are The Different Types of Steel & Steel Grades? - Metal Supermarkets*. Retrieved January 18, 2025, from https://www.metalsupermarkets.com/different-types-steel-steel-grades/.

Pang, J. C., Xu, B. Y., Wang, G. D., Lu, Q., Wang, J. F., & Yi, H. L. (2017). Effect of silicon and aluminium in ferrite on tensile and impact properties. *Materials Science and Technology (United Kingdom), 33*(15), 1806–1810. https://doi.org/10.1080/02670836.2017.1320085/ASSET/IMAGES/10.1080_02670836.2017.1320085-IMG10005.PNG

Quaranta, E., & Davies, P. (2022). Emerging and innovative materials for hydropower engineering applications: Turbines, bearings, sealing, dams and waterways, and ocean power. *Engineering, 8*, 148–158.

Safarian, S. (2023). To what extent could biochar replace coal and coke in steel industries? *Fuel, 339*, Article 127401.

Safarian, S., Ebrahimi Saryazdi, S. M., Unnthorsson, R., & Richter, C. (2021). Modeling of hydrogen production by applying biomass gasification: Artificial neural network modeling approach. *Fermentation, 7*(2), 71.

Safarian, S., Saryazdi, S. M. E., Unnthorsson, R., & Richter, C. (2020). Artificial neural network integrated with thermodynamic equilibrium modeling of downdraft biomass gasification-power production plant. *Energy, 213*, Article 118800.

Safarian, S., Saryazdi, S. M. E., Unnthorsson, R., & Richter, C. (2021). Artificial neural network modeling of bioethanol production via syngas fermentation. *Biophysical Economics and Sustainability, 6*, 1–13.

Shen, C., Wang, C., Wei, X., Li, Y., van der Zwaag, S., & Xu, W. (2019). Physical metallurgy-guided machine learning and artificial intelligent design of ultrahigh-strength stainless steel. *Acta Materialia, 179*, 201–214.

Wang, C., Shen, C., Cui, Q., Zhang, C., & Xu, W. (2020). Tensile property prediction by feature engineering guided machine learning in reduced activation ferritic/martensitic steels. *Journal of Nuclear Materials, 529*, Article 151823. https://doi.org/10.1016/J.JNUCMAT.2019.151823

Wang, S., Gao, Z., Wu, G., & Mao, X. (2022). Titanium microalloying of steel: a review of its effects on processing, microstructure and mechanical properties. *International Journal of Minerals, Metallurgy and Materials, 29*(4), 645–661. https://doi.org/10.1007/S12613-021-2399-7/METRICS

Xie, Q., Suvarna, M., Li, J., Zhu, X., Cai, J., & Wang, X. (2021). Online prediction of mechanical properties of hot rolled steel plate using machine learning. *Materials & Design, 197*, Article 109201.

Xiong, J., Zhang, T. Y., & Shi, S. Q. (2020). Machine learning of mechanical properties of steels. *Science China Technological Sciences, 63*(7), 1247–1255. https://doi.org/10.1007/S11431-020-1599-5/METRICS

Zaitsev, A., & Arutyunyan, N. (2021). Low-Carbon Ti-Mo microalloyed hot rolled steels: special features of the formation of the structural state and mechanical properties. *Metals, 11*, 1584. https://doi.org/10.3390/MET11101584. *202111*(10), 1584.

Zhan, Z., Shi, Z., Wang, Z., Lu, W., Chen, Z., Zhang, D., Chai, F., & Luo, X. (2024). Effect of manganese on the strength–toughness relationship of low-carbon copper and nickel-containing hull steel. *Materials, 17*(5), 1012. https://doi.org/10.3390/MA17051012. *202417*1012.