Article

Chemistry-driven Predictive Modelling of Mechanical Performance in 42CrMo4 (AISI 4140) Steel with Machine Learning

Sustainable Manufacturing and Foundry Practices
1(1) 64–79, 2026

DOI: 10.1177/IIF.261438647
Journal.indianfoundry.org
THE INSTITUTE OF INDIAN FOUNDRYMEN

Subhasis Das Gupta[1], Abhinav Anand[2] and Ram Krishna[1]

Abstract

The 42CrMo4 low-alloy steel shafts, gears, and other high-strength parts used in the automotive industry offer a strong balance of toughness, strength, and surface quality. Traditionally, alloy development has relied on empirical knowledge and repeated experimental trials to adjust chemical composition and achieve targeted mechanical properties. In this study, we developed a machine learning model to predict the mechanical properties of 42CrMo4 steel from its chemical composition. Our dataset comprised 1,000 heat measurements listed as weight percentages, along with tensile strength, yield strength, proof stress at 0.2%, and surface roughness (R_a) from machining. After removing anomalous data with extreme or inconsistent compositional values, nine key alloying elements, Ti, Ni, Cr, Mo, Cu, Mn, P, Si, and C, were selected as input parameters for the machine learning model. Different machine learning models were trained separately for each identified property, ensuring tailored predictions. We tune their hyperparameters using a five-fold cross-validation grid search, helping us find the best settings. The Gradient Boosting Regressor algorithm performs well, showcasing its effectiveness and reliability, with R^2 values ranging from 0.64 to 0.73 across the four targets, indicating reliable predictions. Additionally, error metrics such as root

[1]Department of Metallurgical and Materials Engineering, National Institute of Technology Jamshedpur, Jharkhand, India.
[2]Department of Electronics, Electronics and Computer Science Engineering, KIIT Bhubaneswar, Odisha, India

Corresponding author:
Ram Krishna, Department of Metallurgical and Materials Engineering, National Institute of Technology Jamshedpur, Jamshedpur, Jharkhand 831014, India.
E-mails: krishnamme@gmail.com; krishna.met@nitjsr.ac.in; sdg.habra@gmail.com

mean squared error and mean absolute error indicate that proposed predictions remain within practical engineering limits, providing confidence in their accuracy. To improve the model, we employed important grouped features and correlation heatmaps for our analyses. These show that Ti, Ni, Cr, and Mo contribute most to strength properties, while Mn, P, and Si have a greater impact on R_a. Further, the results are comparable to the metallurgical information on solid-solution strengthening, carbide formation, and surface finish sensitivity. The suggested framework demonstrates that composition-based machine learning models can support alloy design, reduce experimental trials, and provide a digital tool for predicting the mechanical properties of 42CrMo4 steel.

Keywords

42CrMo4 steel, low-alloy steel, machine learning algorithms, Gradient Boosting Regressor, mechanical properties

Introduction

Low-alloy, quenched-and-tempered steels such as 42CrMo4 are commonly used in the automotive, energy, and general engineering sectors. Widely used components include shafts, axles, gears, and heavy-duty fasteners, for which steel designers demand high strength, good fatigue resistance, and toughness (Bilbao et al., 2023; Dong et al., 2020; Duda et al., 2021). The final mechanical properties of 42CrMo4 steel are governed by the combined effects of chemical composition, heat-treatment conditions, and prior thermomechanical history (Mudda et al., 2025). In practice, alloy steel design frequently depends on operators' experience, empirical formulae, and trial-and-error adjustments to elemental compositions, followed by mechanical processing. These techniques demand significant time and resources and present challenges in achieving the appropriate equilibrium of properties.

42CrMo4 steel is a medium-carbon chromium-molybdenum alloy steel in which elements such as chromium, molybdenum, and nickel contribute to improved hardenability, strength, and toughness. These characteristics make the alloy suitable for demanding engineering applications, particularly in automotive and heavy mechanical components. Adding microalloying elements, such as titanium or niobium, refines the grain size, while residual elements, including copper, phosphorus, and sulphur, affect surface quality and ease of machining. Even minor adjustments to the composition alter the material's transformation behaviour, the way carbides form, and the overall microstructure; ultimately, they influence strength, ductility, and surface finish after machining (Gladman, 1997; Honeycombe & Bhadeshia, 2006; Totten, 2006; Trent & Wright, 2000). A clear understanding of these interactions is essential for the systematic design and optimisation of alloy steels.

The complex relationships among composition, processing, and properties make it challenging to develop a standard model for these steels (Yang et al., 2022). Therefore, advanced emerging techniques like machine learning show great promise. This approach improves understanding of the relationships between alloy composition, processing conditions, and resulting properties. This helps

develop new materials faster and achieve specific properties more easily, making the process more efficient and successful (Colla et al., 2023). Several machine learning algorithms have been applied in materials engineering to analyse complex materials and datasets. These techniques include powerful tools such as support vector machines, random forests, and artificial neural networks. They are highly reliable at predicting key material properties, such as hardness, toughness, and tensile strength (Yarasu & Podgornik, 2025).

Data-driven approaches use historical input datasets to create models. These methods help identify relationships between compositional inputs and the resulting mechanical properties. Although this process can sometimes be challenging with traditional methods, the effort is justified by the benefits of the employed methodologies. These approaches are more efficient for designing and developing alloys and optimising their targeted properties (Lee et al., 2021). Data-driven machine learning approaches have emerged as effective tools for analysing complex relationships in steel alloy and other material systems. They provide information on the complex, non-linear relationships between composition and material properties, enabling a reduction in the number of experiments (Guo et al., 2020).

In metallic alloys, machine learning can be a valuable tool for predicting mechanical properties, process zones, and microstructural features as functions of chemical and material-processing parameters. Models based on compositional-material property predictions offer robust insights. Furthermore, it helps reduce costs by optimising the use of expensive alloying elements while achieving target properties. Machine learning models have demonstrated strong capability in predicting the mechanical properties of steel alloys using compositional descriptors (Cheng et al., 2024; Wang et al., 2019). They simplify achieving required properties by speeding up the alloy design process.

Although many studies have applied machine learning to steels, most focus on predicting bulk mechanical properties such as strength or hardness (Ahmed et al., 2025; Huang et al., 2024). Machining-related responses, particularly surface roughness (R_a), are rarely included in these predictive models (D'Urso et al., 2023; Huang et al., 2024; Kaur et al., 2025; Ling et al., 2017; Wang et al., 2024; Zhang et al., 2020). However, machining behaviour is a critical factor in manufacturing performance and product quality. The present study addresses this research gap by developing a composition-based machine learning framework for 42CrMo4 (AISI 4140) steel that predicts both strength-related properties and R_a after machining. A Gradient Boosting Regressor (GBR) model was employed to evaluate these parameters.

Materials and Methods

Material History and Chemical Composition

42CrMo4 (AISI 4140) is a quenched-and-tempered low-alloy steel highly prized for its capacity to meet necessary application property requirements. It belongs to the chromium-molybdenum steel group and features a medium-carbon level, with

hardness and toughness improved by heat treatment. It is often used in shafts, gears, connecting rods, axles, and bolts, where dependable mechanical performance is crucial (Mazini et al., 2022). The chemical composition is a critical factor controlling the mechanical performance and long-term reliability of the component (Xiong et al., 2020). Elements such as carbon, chromium, molybdenum, and nickel influence hardenability, martensite formation, and tempering characteristics. Manganese, silicon, phosphorus, and sulphur play important roles in determining secondary qualities, such as how easily it machines and its surface finish.

The mechanical properties of this steel depend on its chemical composition. Elements such as carbon, chromium, molybdenum, and nickel influence hardenability, martensite formation, and tempering characteristics. Manganese, silicon, phosphorus, and sulphur play important roles in determining the secondary qualities of steel, such as how easily it machines and its surface finish. At their respective levels can significantly impact strength and smoothness, especially in mechanised settings where many heats are melted and cast to meet customers' requirements. The present study examined the chemical composition of 42CrMo4 (AISI 4140) steel after casting, based on data from 1,000 samples. The compositions of all these samples are listed as weight percent. Additionally, each sample was tested for mechanical properties following the standard Universal Testing Method (UTM). The data replaced the model's thorough review for errors, and entries with absent or uneven values in accuracy. Following this process, nine elements were selected for their known influence on strength, toughness, and surface finish. Carbon, chromium, molybdenum, and nickel are the main alloying elements and strongly influence hardenability and strength. Titanium, copper, phosphorus, silicon, and manganese are present in minor amounts but still exert an effect on grain size, chemical segregation, and machining response. These nine elements collectively provide a clear and appropriate representation in controlling the mechanical properties of 42CrMo4 steel. Table 1 presents the nine chemical elements used for inputs. All values are given in weight percent (%) and correspond to the typical composition range of 42CrMo4 (AISI 4140) steel.

Mechanical Property Measurements

The mechanical properties used in this study were taken from standard tests performed on 42CrMo4 (AISI 4140) steel samples. A few mechanical properties were selected as they represent the most critical performance requirements for this grade: ultimate tensile strength (UTS), yield strength (YS), 0.2% proof stress, and R_a. The principles justify the material's ability to carry a load and resist permanent deformation, and they depend on the microstructure formed during heat treatment.

R_a provides important insights into machining performance and the surface quality of the components we prepare. To understand the tensile properties, we analysed stress–strain curves obtained with a universal testing machine. We carefully measured R_a using a calibrated contact-type profilometer to ensure accuracy. The test matched each mechanical property to its corresponding chemical composition, helping to build a reliable supervised dataset. Table 2 highlights the four fundamental

Table 1. 42CrMo4 (AISI 4140) Steel Shows Chemical Compositions (wt%) and Their Range Used for ML Modelling.

Element	Symbol	Typical Range (wt%)	Role in Steel
Carbon	C	0.38–0.45	Strengthening element, increases hardness and hardenability
Silicon	Si	0.10–0.40	Supports deoxidation, influences ferrite carbide structure
Phosphorus	P	≤0.025	Residual element affects segregation and machinability
Manganese	Mn	0.60–0.90	Improves hardenability and toughness
Copper	Cu	≤0.30	Residual and minor effect on surface quality
Chromium	Cr	0.90–1.20	Enhances hardenability, forms strengthening carbides
Nickel	Ni	≤0.40	Improves hardiness and impact strength
Molybdenum	Mo	0.15–0.30	Improves tempering resistance and strength retention
Titanium	Ti	≤0.050	Forms fine precipitates; helps grain refinement

Source: Honeycombe & Bhadeshia (2006); Totten (2006).

Table 2. The Mechanical Properties and Their Value Range of 42CrMo4 (AISI 4140) Steel,

Mechanical Property	Symbol	Range (Typical)	Unit
Ultimate tensile strength	UTS	900–1,100	MPa
Yield strength	YS	700–900	MPa
0.2% proof stress	0.2% PS	650–850	MPa
Surface roughness	Ra	4.5–8.0	μm

Source: Mazini et al. (2022); Mudda et al. (2025); Bibao et al. (2023).

mechanical properties used as outputs in our models. The ranges listed below reflect typical values for quenched-and-tempered 42CrMo4 (AISI 4140) steel.

Data Cleaning and Preparation

The data was carefully gathered to ensure everything was clear and consistent for the model. First, we reviewed the missing values in the chemical composition and mechanical property columns. Any samples lacking complete information were politely removed, and interpolation was used only when necessary. During data preprocessing, outliers were identified as records with extreme compositional

values that deviated significantly from the dataset's typical range. We looked at outliers- values outside 1.5 times the interquartile range- and removed those resulting from testing issues. These anomalous entries may arise from measurement inconsistencies or data reporting errors and can negatively influence machine learning model training. Such records were removed prior to feature selection and model development to improve dataset consistency.

After cleaning, we standardised the feature labels, merged duplicates, and displayed all compositions as weight percentages for a stronger, additional uniform view. The heatmap was created to visualise relationships between mechanical properties and chemical elements, providing insight.

Reasoning Behind the Selection of the Nine Chemical Features

The nine chemical elements used in this study were carefully selected because they hold a key role in affecting the behaviour of 42CrMo4 steel. Carbon is the primary element that enhances strength and promotes martensite formation during hardening. Chromium encourages hardenability and forms carbides. Molybdenum helps prevent softening and increases toughness. Nickel enhances toughness and enables greater hardening. Titanium forms fine particles, resulting in a finer grain size. Manganese also improves hardenability and surface characteristics. Silicon is vital for balancing ferrite and carbides and for improving the surface finish. Phosphorus, even in small amounts, affects machinability and roughness. Copper provides a small amount of strengthening and affects the steel's surface.

Dataset Summary

After cleaning and checking the data, a final dataset comprising 1,000 entries was prepared for modelling. It contains data on both its chemical composition and corresponding mechanical properties. The dataset includes nine chemical elements as inputs for capturing the strengthening mechanisms and surface-related effects in 42CrMo4 steel. And four mechanical properties as output variables, UTS, YS, 0.2% proof stress, and R_a. Since each sample possesses corresponding composition and property data. The dataset is appropriate for supervised machine learning. This cleaned and organised data was used in the modelling work described in the next section.

Machine Learning Regression Models

In this research, we designed a machine learning framework to generate reliable demonstrations that work effectively to predict the mechanical properties of 42CrMo4 steel from its chemical composition. The model uses nine compositional variables as input features and predicts four mechanical properties as output variables. This way, forecasting capabilities for each property are improved with confidence. The modelling process was simple and followed supervised-learning practices, making the methodology inclusive and manageable (Wei et al., 2021). In this study, four independent regression models were developed, each targeting

a single output variable such as UTS, YS, 0.2% proof stress, and R_a. A multi-output regression framework was not adopted because these properties arise from different metallurgical compositional mechanisms and may respond differently to compositional variations. Training individual models enables the algorithm to better capture property-specific relationships between alloying elements and mechanical responses.

We chose the GBR algorithm as our main model because it performs well on small and medium-sized datasets. It is also highly effective at capturing complex associations among various elements and their mechanical properties, so we can trust its reliability. GBR constructs an ensemble of decision trees, where each tree iteratively reduces the prediction error of the previous model. This makes the method suitable for material datasets containing complex and non-linear relationships (Zhang et al., 2022).

Model validation was performed using a combination of train-test splitting of the dataset and five-fold cross-validation. To train and test, the dataset was split into training and test sets. This approach helps ensure a smoother and more manageable process. Around 80% of the data was utilised for training and model development, including hyperparameter optimisation, while the remaining 20% was reserved to assess the model's performance on new, unseen samples. To ensure our assessment is reliable and reasonable, we used five-fold cross-validation. This means we split the training set into five parts, trained the model on four parts, and tested it on the fifth. We repeated this process five times, each with a different part, and averaged the results to gain a clearer understanding of the model in action. This method helps prevent overfitting and provides a more reliable estimation of predictive accuracy and a sense of how well the model truly performs (Mohanty et al., 2022).

The model's hyperparameters were carefully fine-tuned through a comprehensive grid search, exploring combinations of settings such as the number of trees, learning rate, and maximum depth to identify the optimal configuration. The combination that achieved the highest cross-validation score was chosen for the final model. Once the best settings were identified, the models were retrained on the entire dataset for training (Pathan et al., 2019). Table 3 illustrates the hyperparameter search ranges and selected values for the GBR models.

The performance of each model was evaluated using standard regression metrics, including the coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE). This provides an easy way to see how accurately the predicted values match the actual data. Overall, the models performed quite well across four properties, particularly excelling at predicting 0.2% proof stress and tensile strength. While R_a was also predicted with good accuracy, it showed slightly less accuracy. This makes sense because roughness depends more on machining and processing conditions than simply composition (Ammar, 2021; Sivakumar et al., 2022).

After the training, we used several plots and straightforward tools to better understand how our model behaves. The feature-importance charts highlighted which features had the biggest impact on the predictions. Actual versus predicted plots show how closely the measured values align with the projections, helping

Table 3. Hyperparameter Search Ranges and Selected Values for the Gradient Boosting Regressor Models.

Hyperparameter	Range Tested in Grid Search	Final Selected Value	Description
N estimators	100–1000	500	Number of boosting trees used in the ensemble
Learning rate	0.01–0.2	0.05	Step size controlling the contribution of each tree
Maximum depth	2–6	3	Maximum depth of individual regression trees
Minimum samples split	2–10	4	Minimum samples required to split an internal node
Minimum samples leaf	1–5	2	Minimum samples required at a leaf node
Subsample	0.7–1.0	0.8	Fraction of training samples used for each tree
Maximum features	Options: Auto, sqrt, log2	sqrt	Number of features/ options considered when looking for the best split and reducing overfitting

Source: Pathan et al. (2019); Mohanty et al. (2022).

assess the level of agreement. This analysis was conducted to explore how different elemental compositions affect the model's output (Javed et al., 2024).

Results and Discussion

The machine learning models were trained on treated data, with nine chemical elements as inputs and four mechanical properties as outputs. We developed individual models for each property: UTS, YS, 0.2% proof stress, and R_a. To evaluate the effectiveness of each model, we used R^2, MAE, and RMSE. These measures provided a clear idea of how closely the predictions matched the experimental values (Tiwari et al., 2025). Table 4 presents the performance metrics (R^2, MAE, RMSE) derived from the machine learning algorithm utilised for predicting mechanical properties.

The models validated excellent performance for the strength-related properties. The 0.2% proof stress prediction had the highest accuracy among the four outputs. This result is expected because proof stress depends directly on the combined effects of carbon, chromium, molybdenum, and nickel, all of which are in the dataset. UTS and YS also showed good prediction accuracy, with most values falling close to the measured results. The prediction accuracy for R_a was slightly lower than that of the

strength-related properties. This behaviour is expected because R_a is influenced not only by the alloy composition but also by machining parameters such as cutting speed, feed rate, tool geometry, lubrication, and tool wear. In the present study, machining conditions were kept constant for all samples during testing. Therefore, the machine learning model primarily captures the influence of chemical composition on surface finish. Nevertheless, the absence of machining parameters in the dataset represents an inherent limitation of composition-based modelling for predicting machining responses. Even with these limitations, the model still captured meaningful trends. Elements such as phosphorus, manganese, and silicon exerted a noticeable influence on roughness values, consistent with common metallurgical knowledge of segregation and surface behaviour (Muhammad et al., 2022).

Figure 1 presents helpful graphs highlighting key features of materials. These visuals clarify which factors matter most. The four plots beautifully illustrate the influence of chemical composition on the mechanical properties of 42CrMo4 steel. The model carefully evaluates the influence of elements associated with each property. For a 0.2% proof stress, Ni, Ti, and Si stand out as the most influential. Nickel enhances the strength and toughness of the steel, making it more durable. Titanium strengthens the steel by forming fine precipitates. When considering YS, Mn, C, and Ni are the key players. Carbon primarily boosts strength, with manganese helping to improve hardenability. Nickel also contributes by helping the material maintain its strength as it begins to yield. For UTS, Ti, Mo, and Cu are the most important. Titanium and molybdenum help prevent softening during heat treatment, while copper provides a slight strengthening effect (Li et al., 2022).

Various factors affect how steel responds under high loads. For R_a, manganese (Mn), nickel (Ni), and phosphorus (P) are particularly important. Manganese

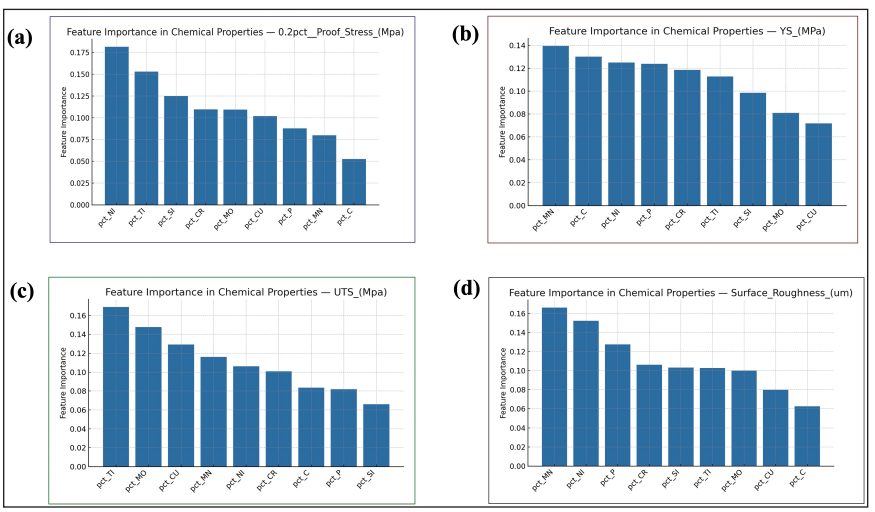

Figure 1. Feature-importance Plots for the Prediction of (a) 0.2% Proof Stress, (b) Yield Strength, (c) Ultimate Tensile Strength, and (d) Surface Roughness of 42CrMo4 Steel. Each of the Nine Elements Impacts Machinability and Strength in Distinct Ways, as Shown by Their Rankings.

affects how the material segregates and the carbide patterns it forms, while phosphorus influences chip formation and the ease with which the material can be machined. Nickel helps stabilise the microstructure, enhancing the material's overall reliability. This is how steel behaves when cut or shaped.

Table 4. Model Performance for Predicting Mechanical Properties of 42CrMo4 Steel Using Chemical Composition.

Property	R^2 (Mean ± Std)	MAE	RMSE	Remarks
UTS (MPa)	0.69 ± 0.04	8.3 MPa	11.5 MPa	Strong correlation and reliability
YS (MPa)	0.67 ± 0.05	7.8 MPa	10.9 MPa	Stable performance
0.2% proof stress (MPa)	0.73 ± 0.03	6.9 MPa	9.7 MPa	Highest model accuracy
Surface roughness (µm)	0.64 ± 0.06	0.28 µm	0.36 µm	Moderate fit, process-dependent

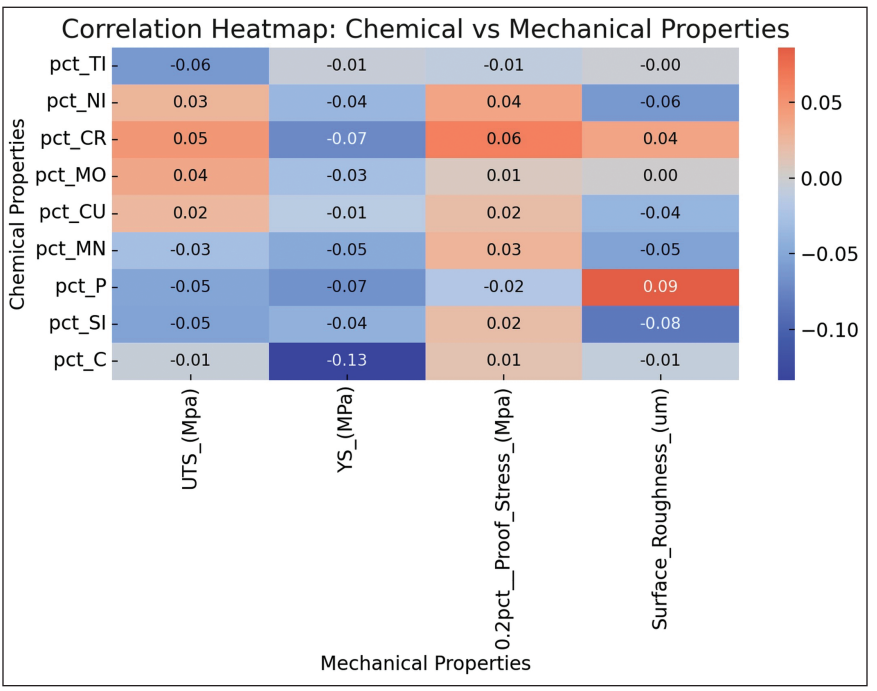

Figure 2. Correlation Heatmap Showing the Linear Relationships Between the Nine Chemical Elements and the Four Mechanical Properties of 42CrMo4 Steel. Most Values Are Low, Indicating Weak Direct Correlations, Which Is Expected for a Multi-element Alloy System. Small Positive and Negative Trends Are Visible for Certain Elements, Such as C, P, and Cr, but None Dominate the Overall Behaviour.

Elements such as chromium and molybdenum, along with carbon, are essential for predicting strength because they control hardenability, martensite formation, and resistance to tempering. Nickel also boosts toughness and supports the steel's strength. Relating to R_a, phosphorus and manganese affect how the material microsegregates and how carbides are distributed, both of which influence cutting performance (Wei et al., 2021).

Figure 2 shows a heatmap of the relationships between composition and the four mechanical properties of 42CrMo4 steel. Most correlations are quite low, which makes sense because these properties are shaped by many different aspects, not just their composition. Carbon shows a slight negative correlation with YS. Phosphorus shows a slight positive association with R_a, consistent with its known effect on machinability. Chromium shows a weak positive trend with proof stress and R_a. The remaining elements do not affect all properties. The heatmap provides a quick overview of these small trends, and a strong linear relationship exists before modelling (Leni et al., 2023).

Figure 3 compares the bar charts showing the effects of chemical elements on all four mechanical properties of 42CrMo4 steel. The bars show the influence each component has on the prediction models. The plot makes it clear that no individual element dominates all properties. Titanium, nickel, and chromium show strong contributions within most strength properties. Molybdenum and copper also play steady roles, mainly in UTS and proof stress. Manganese, phosphorus, and silicon have a larger effect on R_a than on strength. Carbon has a modest influence, but its impact differs within properties. They reflect the behaviour expected of alloyed steels, in

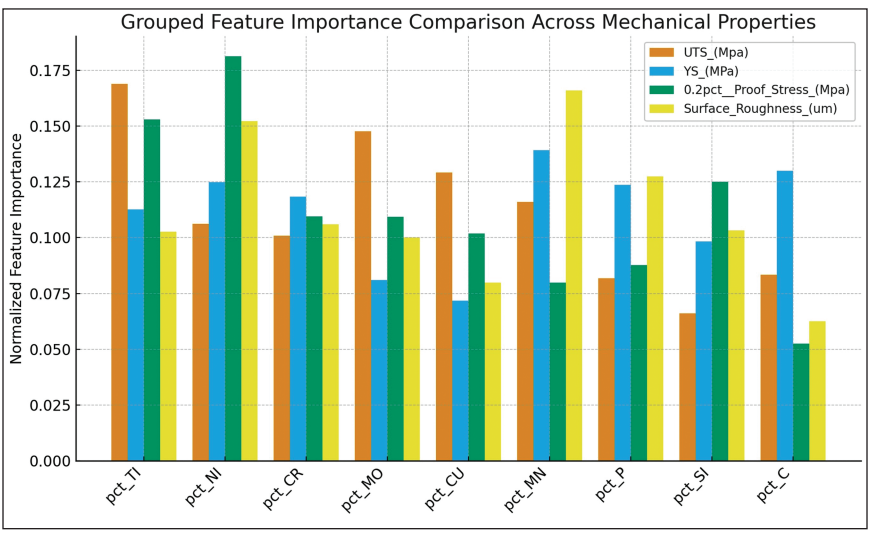

Figure 3. Grouped Feature-importance Comparison for the Prediction of Ultimate Tensile Strength, Yield Strength, 0.2% Proof Stress, and Surface Roughness of 42CrMo4 Steel. The Plot Shows the Contributions of Different Elements to the Four Properties, Highlighting Those That Influence Strength (Ti, Ni, Cr, Mo) and Those That Have a More Substantial Effect on Machinability (Mn, P, Si).

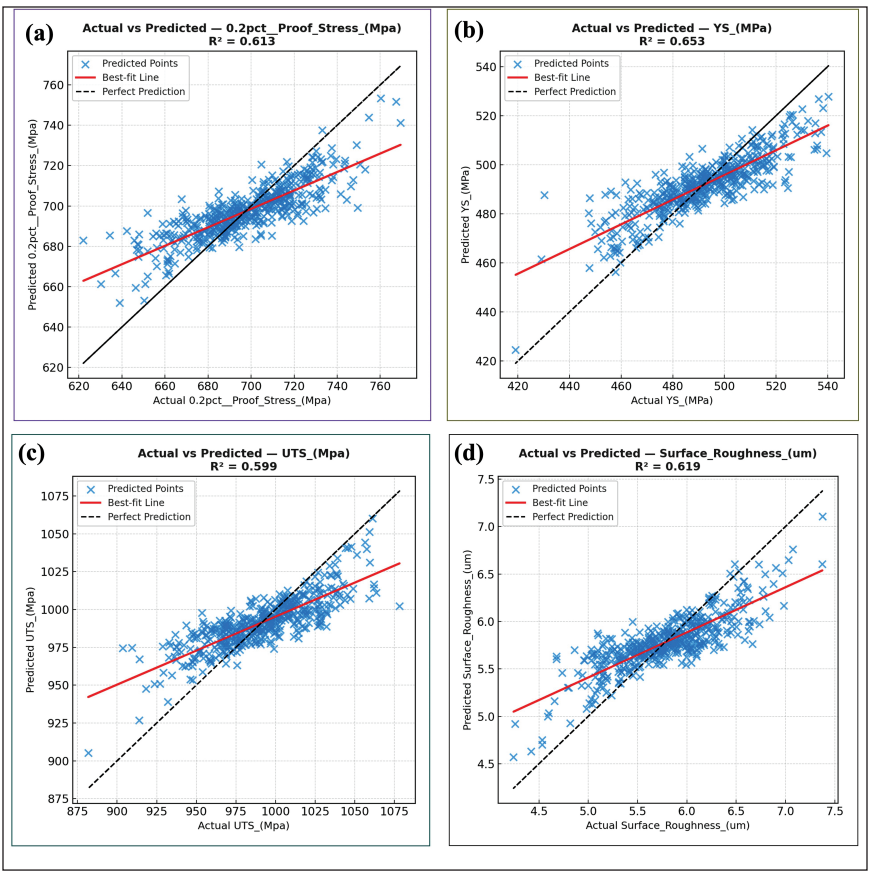

Figure 4. The Plots for (a) 0.2% Proof Stress, (b) Yield Strength, (c) Ultimate Tensile Strength, and (d) Surface Roughness of 42CrMo4 Steel Show the Links Between Actual and Predicted. Each Plot Includes Model Predictions, a Best-fit Line, and a Perfect-prediction Reference Line. The Trends Show Good Agreement for Proof Stress and Yield Strength, with Moderate Fits for UTS and Surface Roughness.

which strengthening and machinability result from several elements acting together rather than from a single element (Romanov & Hawk, 2019).

Figure 4 compares the model-predicted values with the measured values for each mechanical property of 42CrMo4 steel. Each plot includes the predicted points, a best-fit regression line, and a perfect-prediction line. The model's performance will improve as the measured points approach the perfect-prediction line. In (a), 0.2% proof stress, most points fall close to the perfect-prediction line. The R^2 value is the highest among the four properties, indicating that the model predicts this property more effectively than the others. In (b), the YS range is wider and exhibits an upward trend. Numerous points are close to the ideal prediction line. This demonstrates consistently steady model behaviour. In (c), UTS, the scatter is larger. The model still follows the main trend, but the predictions show more variation. UTS is affected

by several factors other than chemistry, which explains the weaker fit. In (d) R_a, the points show a noticeable upward trend, but there is greater scatter, especially at higher roughness values. R_a is influenced by the alloy composition and machining parameters, so a moderate fit is usually expected (Peng et al., 2020). Overall, all four plots show the material's behavioural properties.

Proof stress and YS are predicted with high accuracy, whereas UTS and R_a have moderate accuracy. They highlight the intrinsic properties of the material and its chemical composition, which makes these results quite promising (Tan et al., 2023).

Overall, these findings suggest that analysing the composition of 42CrMo4 steel alone can give us trustworthy insights into its strength properties. We can also model R_a quite effectively, and by including machining parameters, we can achieve even better results. This shows that the machine learning approach is not just accurate; it also aligns with metallurgical principles. It is a really helpful tool that can support composition control, process planning, and early property predictions for different applications. It advances assurance in the processes and makes everything more dependable.

Conclusions

This study demonstrates that the mechanical properties of 42CrMo4 (AISI 4140) steel can be accurately predicted solely based on composition. The data, containing 1,000 heats and nine essential alloying and residual elements, enabled the Gradient Boosting algorithm to establish reliable correlations between composition and strength-related characteristics. The models showed strong correlations for 0.2% proof stress, YS, and tensile strength, with proof stress achieving the highest accuracy. R_a predictions were moderate, which is understandable because machining conditions—though influential—were not included in the dataset. It should be noted that R_a is also affected by machining parameters, including cutting speed, feed rate, and tool condition. Since these parameters were not included in the present dataset, the model predicts roughness mainly from compositional effects. Future work should integrate machining parameters and microstructural descriptors to further improve the predictive capability of the model.

The features and their correlation patterns aligned well with metallurgical principles. Elements such as Ti, Ni, Cr, and Mo significantly impacted strength, whereas Mn, P, and Si mainly affected R_a. These findings reflect known phenomena like strengthening and segregation behaviours in low-alloy steels. The comparison of actual versus predicted values further validated that the model captured the main behavioural trends, especially for strength properties.

Overall, this work demonstrates the machine learning's supportive role in alloy composition management, heat-treatment scheduling, and initial property forecasting in steel manufacturing. Such methods can notably reduce the number of mechanical tests required and accelerate feedback during alloy development. Future enhancements include processing parameters, microstructural details, and machining conditions to improve R_a predictions and to generate a complete digital platform for steel design and production.

Authors' Contributions

All authors contributed to this work.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship and/or publication of this article: Financial assistance from the Anusandhan National Research Foundation (ANRF) under the Partnerships for Accelerated Innovation and Research (PAIR) programme, in collaboration with IIT (ISM) Dhanbad, for the project 'Development of Innovative and Cutting-Edge Indigenous Technologies for Critical Minerals Exploration and Smart/Sustainable Mining' (Sanction Order No. ANRF/PAIR/2025/000027/PAIR-B) is gratefully acknowledged.

References

Ahmed, R., Kumar, V., Faisal, N. H., Al-Marri, M. J., & Davies, S. (2025). Influence of alloy composition on the tribomechanical properties of 50% blend of CoCrWMoCFeNiSiMn (Stellite 1) and CoCrMoCFeNiSiMn (Stellite 21) alloys. *Journal of Materials Engineering and Performance*. https://doi.org/10.1007/s11665-025-11034-7

Ammar, M. H. A. (2021). *Advanced digital twins for condition monitoring, diagnosis, and predictive remaining life using artificial intelligence (Doctoral dissertation, Brunel University London)*. Brunel University Research Archive. https://bura.brunel.ac.uk/bitstream/2438/28074/1/FulltextThesis.pdf

Bilbao, O., Loizaga, I., Alonso, J., Girot, F., & Torregaray, A. (2023). 42CrMo4 steel flow behavior characterization for high temperature closed dies hot forging in automotive components applications. *Heliyon*, *9*(11), e22256. https://doi.org/10.1016/j.heliyon.2023.e22256

Cheng, H., He, Z., Ge, M., Che, L., Zheng, K., Si, T., & Zhao, F. (2024). Composition design and optimization of Fe–C–Mn–Al steel based on machine learning. *Physical Chemistry Chemical Physics*, *26*(10), 8219. https://doi.org/10.1039/d3cp05453e

Colla, V., Vannucci, M., Matino, I., & Valentini, R. (2023). A deep learning-based approach to the estimation of Jominy profile of medium-carbon quench hardenable steels. *Steel Research International*, *95*(2). https://doi.org/10.1002/srin.202300374

Dong, J., Pei, W., Ji, H., Long, H., Fu, X., & Duan, H. (2020). Fatigue crack propagation experiment and numerical simulation of 42CrMo steel. *Proceedings of the Institution of Mechanical Engineers*, *Part C: Journal of Mechanical Engineering Science*, *234*(14), 2852. https://doi.org/10.1177/0954406220910458

D'Urso, D., Chiacchio, F., Cavalieri, S., Gambadoro, S., & Khodayee, S. M. (2023). Predictive maintenance of standalone steel industrial components powered by a dynamic reliability digital twin model with artificial intelligence. *Reliability Engineering & System Safety*, *243*, 109859. https://doi.org/10.1016/j.ress.2023.109859

Duda, M., Rozumek, D., Lesiuk, G., Smolnicki, M., Babiarczuk, B., & Warycha, J. (2021). Fatigue crack growth under mixed-mode I + II and I + III in heat treated 42CrMo4 steel. *International Journal of Fracture*, *234*, 235. https://doi.org/10.1007/s10704-021-00585-0

Gladman, T. (1997). *The physical metallurgy of microalloyed steels*. The Institute of Materials.

Guo, K., Yang, Z., Yu, C., & Buehler, M. J. (2020). Artificial intelligence and machine learning in design of mechanical materials. *Materials Horizons*, *8*(4), 1153. https://doi.org/10.1039/d0mh01451f

Honeycombe, R. W. K., & Bhadeshia, H. K. D. H. (2006). *Steels: Microstructure and properties* (3rd ed.). Elsevier Butterworth-Heinemann.

Huang, J., Ando, D., & Sutou, Y. (2024). Heat-resistant aluminum alloy design using explainable machine learning. *Materials & Design*, *243*, 113057. https://doi.org/10.1016/j.matdes.2024.113057

Javed, M. F., Fawad, M., Lodhi, R. H., Najeh, T., & Gamil, Y. (2024). Forecasting the strength of preplaced aggregate concrete using interpretable machine learning approaches. *Scientific Reports*, *14*(1), 8381. https://doi.org/10.1038/s41598-024-57896-0

Kaur, R., Kumar, R., & Aggarwal, H. (2025). Systematic review of artificial intelligence, machine learning, and deep learning in machining operations: Advancements, challenges, and future directions. *Archives of Computational Methods in Engineering*, *32*, 4983–5036. https://doi.org/10.1007/s11831-025-10290-z

Lee, J.-W., Park, C., Lee, B. D., Park, J., Goo, N. H., & Sohn, K. (2021). A machine-learning-based alloy design platform that enables both forward and inverse predictions for thermo-mechanically controlled processed (TMCP) steel alloys. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-90237-z

Leni, D., Earnestly, F., Sumiati, R., Adriansyah, A., & Kusuma, Y. P. (2023). Evaluasi sifat mekanik baja paduan rendah berdasarkan komposisi kimia dan suhu perlakuan panas menggunakan teknik exploratory data analysis (EDA). *Dinamika Teknik Mesin*, *13*(1), 74. https://doi.org/10.29303/dtm.v13i1.624

Li, Z., Wang, Z., Wang, Z., Qin, Z., Liu, F., Tan, L., Jin, X., Fan, X., & Huang, L. (2022). Quantifying solid solution strengthening in nickel-based superalloys via high-throughput experiment and machine learning. *Computer Modeling in Engineering & Sciences*, *135*(2), 1521. https://doi.org/10.32604/cmes.2022.021639

Ling, J., Hutchinson, M., Antono, E., DeCost, B., Holm, E. A., & Meredig, B. (2017). Building data-driven models with microstructural images: Generalization and interpretability. *Materials Discovery*, *10*, 19–27. https://doi.org/10.1016/j.md.2018.03.002

Mazini, J. P., Filho, A. I., Ávila, B. M. R., da Silva, R. V., & de Oliveira, P. G. B. (2022). Microstructure and mechanical properties of microalloyed steels containing molybdenum. *Materials Research*, 25. https://doi.org/10.1590/1980-5373-mr-2021-0608

Mohanty, T., Chandran, K. S. R., & Sparks, T. D. (2022). Machine learning guided optimal composition selection of niobium alloys for high temperature applications. *ChemRxiv*. https://doi.org/10.26434/chemrxiv-2022-1d9k0

Mudda, S., Hegde, A., Sharma, S., Gurumurthy, B. M., Shettar, M., & Shankar, M. C. G. (2025). Effect of various heat treatment methods and optimization of their parameters on mechanical properties of AISI 4140 steel. *Scientific Reports*, *15*(1), 31854. https://doi.org/10.1038/s41598-025-17299-1

Muhammad, W., Kang, J., Ibragimova, O., & Inal, K. (2022). Experimental investigation and development of a deep learning framework to predict process-induced surface roughness in additively manufactured aluminum alloys. *Welding in the World*, *67*(4), 897. https://doi.org/10.1007/s40194-022-01445-8

Pathan, M. V., Ponnusami, S. A., Pathan, J., Pitisongsawat, R., Erice, B., Petrinic, N., & Tagarielli, V. L. (2019). Predictions of the mechanical properties of unidirectional

fibre composites by supervised machine learning. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-50144-w

Peng, J., Yamamoto, Y., Hawk, J. A., Lara-Curzio, E., & Shin, D. (2020). Coupling physics in machine learning to predict properties of high-temperature alloys. *npj Computational Materials*, *6*(1). https://doi.org/10.1038/s41524-020-00407-2

Romanov, V., & Hawk, J. A. (2019). *Data analytics for designing Fe-9Cr steels*. U.S. Department of Energy Office of Scientific and Technical Information. https://www.osti.gov/biblio/1569972

Sivakumar, E. R., Krishna, R., Senthilkumar, P., & Çolak, İ. (2022). Influence of mechanical properties of coating and substrate on wear performance of HDLC and TiN-coated AISI 5140 alloy steel. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*. https://doi.org/10.1177/09544089221132434

Tan, R. R., Li, Z., Zhao, S., & Birbilis, N. (2023). A primitive machine learning tool for the mechanical property prediction of multiple principal element alloys. *arXiv*. https://doi.org/10.48550/arxiv.2308.07649

Tiwari, S., Ahn, H. S., Reddy, M. H., Park, N., & Reddy, N. S. (2025). Mechanical property prediction of industrial low-carbon hot-rolled steels using artificial neural networks. *Materials*, *18*(13), 2966. https://doi.org/10.3390/ma18132966

Totten, G. E. (2006). *Steel heat treatment: Metallurgy and technologies*. CRC Press.

Trent, E. M., & Wright, P. K. (2000). *Metal cutting* (4th ed.). Butterworth-Heinemann.

Wang, C., Shen, C., Cui, Q., Zhang, C., & Xu, W. (2019). Tensile property prediction by feature engineering guided machine learning in reduced activation ferritic/martensitic steels. *Journal of Nuclear Materials*, *529*, 151823. https://doi.org/10.1016/j.jnucmat.2019.151823

Wang, K., Gupta, V., Lee, C. S., Mao, Y., Kilic, M. N. T., Li, Y., Huang, Z., Liao, W.-K., Choudhary, A., & Agrawal, A. (2024). XElemNet: Towards explainable AI for deep neural networks in materials science. *Scientific Reports*, *14*, 25178. https://doi.org/10.1038/s41598-024-76535-2

Wei, Q., Xiong, J., Sun, S., & Zhang, T. (2021). Multi-objective machine learning of four mechanical properties of steels. *Scientia Sinica Technologica*, *51*(6), 722. https://doi.org/10.1360/sst-2020-0475

Xiong, J., Zhang, T., & Shi, S. (2020). Machine learning of mechanical properties of steels. *Science China Technological Sciences*, *63*(7), 1247. https://doi.org/10.1007/s11431-020-1599-5

Yang, X., El-Fallah, G. M. A. M., Tao, Q., Fu, J., Leng, C., Shepherd, J., & Dong, H. (2022). Dimensionality reduction for machine learning using statistical methods: A case study on predicting mechanical properties of steels. *Materials Today Communications*, *34*, 105162. https://doi.org/10.1016/j.mtcomm.2022.105162

Yarasu, V., & Podgornik, B. (2025). An optimized machine-learning tool to predict heat treatment response of hot-work tool steels. *Results in Engineering*, *26*, 105260. https://doi.org/10.1016/j.rineng.2025.105260

Zhang, C., Gu, X., Bhandari, U., Lei, J., Guo, S., Kourouma, M., Karoui, A., & Yang, S. (2022). Machine learning prediction of the mechanical properties of refractory multi-component alloys based on a dataset of phase and first principles simulation. *Frontiers in Metals and Alloys, 1*. https://doi.org/10.3389/ftmal.2022.1036656

Zhang, Y., Sun, C., Gao, L., Yue, Z., Shabbir, S., Xu, W., Wu, M., & Yu, J. (2020). Determination of minor metal elements in steel using laser-induced breakdown spectroscopy combined with machine learning algorithms. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *166*, 105802. https://doi.org/10.1016/j.sab.2020.105802