International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I5P117

Volume 74 Issue 5, 250-262, May 2026

Original Article

Machine Learning-Based Modeling of Process-Structure-Property Relationships in IN-36 Alloys: A Random Forest and L18 DOE Approach

Amol A. Dhakane[1], R. A. Kapgate[2]

[1,2]*Sanjivani College of Engineering, Savitribai Phule Pune University, Kopargaon, Maharashtra, India.*

[1]*Corresponding Author : amol31dhakane@gmail.com*

Received: 28 December 2025 Revised: 28 February 2026 Accepted: 11 March 2026 Published: 30 May 2026

Abstract *- The study combines a controlled experiment with guided machine learning to create explicable process structure property relationships of low-expansion Fe–36Ni-based alloys. Three different alloys (IN-36, IN-36Cr, IN-36Ti) on an L18 orthogonal array and cold-work levels (0, 30, and 60 percent) were used. They were then aged at 400 or 450 o C temperature for a period of 2, 7, or 24 hours and then stabilized at a temperature of 750 °C for 1 hour. The coefficient of thermal expansion (CTE, 20 -100 °C) and Vickers hardness, yield strength, ultimate tensile strength, elongation to failure, and stabilized grain size are the values that we measured under each of the conditions (Triplicate Averages). These data were used to train a multi-output Random Forest regressor, and a shallow neural network was used as a baseline. To ensure that it could work well with small data, we ran 6-fold cross-validation. The surrogate precisely measures the alterations within hardness, yield strength, ultimate tensile strength, elongation, and the size of grains; the coefficients of determination range 0.80-0.92, and the mean absolute error is low. Nevertheless, it does not perform so well with CTE since it possesses a small dynamic range. Feature-importance analysis indicates that cold work is the primary predictor for all mechanical and microstructural outputs, alloy type exerts a secondary yet significant influence, while aging temperature and time have lesser impacts within the examined range. Response surfaces created from the trained model show clear trade-offs between strength, ductility, and grain size. They also help find process windows that combine low CTE with high strength and acceptable elongation. In general, the work shows that a compact L18 design and an interpretable machine learning surrogate can help researchers find and improve processing strategies for Ni–Fe superalloys that don't expand much.*

Keywords *- Property linkages, Machine learning, Random Forest surrogate, Invar, Fe–36Ni alloys.*

1. Introduction

Ni-Fe-based low-expansion alloys such as classical Fe-36Ni (IN-36) and its modified variants are widely used where dimensional stability, thermal shock resistance, and mechanical reliability are critical, for example, in precision metrology, cryogenic structures, and aerospace components. The thermo-mechanical route applied to these alloys-including cold work, aging treatments, and subsequent stabilization-exerts a strong, highly non-linear influence on grain size, dislocation substructure, and phase balance, which in turn governs the achievable combinations of Coefficient Of Thermal Expansion (CTE), hardness, strength, and ductility. Conventional alloy and process development in this space has relied heavily on expert intuition and trial-and-error heat-treatment schedules, which are costly and explore only a small fraction of the design space. In parallel, the emergence of Machine Learning (ML) and materials informatics has opened the possibility of constructing data-driven Process-Structure-Property (PSP) linkages that can act as fast surrogates for experiments and physics-based simulations, particularly for complex multi-parameter thermo-mechanical histories.

Flat tensile and dilatometry specimens were machined from stock of the three alloy variants IN-36, IN-36Cr, and IN-36Ti. The materials were supplied in a homogenized and solution-treated condition suitable for subsequent deformation and aging. Gauge sections were prepared with dimensions compatible with standard room-temperature tensile testing, while additional blanks were reserved for hardness measurements, dilatometry, and metallography. To introduce controlled levels of cold work, the specimens were deformed by rolling or drawing to nominal plastic strain levels corresponding to 0 percent (no additional deformation), 30 percent, and 60 percent thickness reduction. These values were chosen to span a practical range from fully recrystallized material to a heavily cold-worked condition in which a high dislocation density and stored energy are available to drive subsequent recovery and recrystallization. A substantial body

of recent work has shown that supervised ML models can learn PSP linkages for Complex Concentrated Alloys (CCAs), High-Entropy Alloys (HEAs), and other multi-component systems when provided with suitable descriptors of composition and processing. Thoppil, Nie, and Alankar (2023) used Bayesian learning models to map thermo-mechanical processing to phase fractions, grain sizes, and strength in CCAs, demonstrating that ML can capture higher-order PSP dependencies that are not easily represented in closed-form constitutive relations. Ling et al. (2018) built regression models linking alloy composition and processing parameters to mechanical properties across aluminum, nickel-based superalloys, and shape-memory alloys, and wrapped these models in inverse-design routines to satisfy multiple property targets. For shape memory alloys, Liu et al. (2021) proposed a physics-informed feature-engineering strategy that transforms heat-treatment data using kinetic and thermodynamic relationships before ML training, which markedly improves predictive accuracy and enables robust composition–process–property design. Complementary efforts in HEAs and CCAs have used ensemble models and feature-selection pipelines to design alloys and processing routes with enhanced mechanical properties (Bobbili et al., 2022; Kumar et al., 2023). An example by Kaufmann and Vecchio (2020) combined both thermodynamic and chemical characterizations with a Random Forest classifier to forecast solid solution forming behavior in HEAs, becoming in itself an effective way to explore immensely large composition spaces. Combined, these studies prove that complex PSP relations can be represented in flexible and interpretable ML models and can be used to perform forward prediction and design of alloys and processes using data.

Simultaneously, the research has attempted to predict given mechanical or functional properties based on composition, process, and microstructural descriptors combinations. Xiong et al. (2020) connected alloy chemistry to glass-forming ability and elastic moduli in bulk metallic glasses with the help of ML that yielded physically significant trends in mixing entropy, thermal conductivity, and atomic volume. By using image-based microstructural features in conjunction with process conditions and composition, Swetlana et al. (2020) modeled Vickers hardness in superalloys by relying on Gaussian-process models, enabling a model that attains very low prediction errors and interpretable structure-property correlations. Mitra et al. (2023) used random forest regression to predict lattice constants, magnetic moments, and formation energies of full, half, inverse, and quaternary Heusler alloys, demonstrating good performance under non-Normal data distributions and multicollinearity. Fragassa et al. (2019) applied pattern-recognition methods to cast irons, predicting tensile behavior from metallographic indicators such as graphite and ferrite fractions. Processing-map construction for eutectic HEAs has been accelerated by ML surrogates trained on hot-deformation flow curves (Kumar et al., 2023), while Rahman et al. (2025) combined synthetic physics-based features with ML to predict yield strength and hardness of DED-processed low-alloy steels from composition, grain size, cooling rate, and thermal history. For high-temperature Ni and Fe–Ni superalloys, Fatriansyah et al. (2025) contrasted direct and inverse ML models to predict tensile strength, hardness, and melting point from composition, and to back-calculate compositions meeting target properties. These results align with broader experience that tree-based ensembles, Gaussian processes, and shallow neural networks perform particularly well on tabular materials datasets with limited sample sizes, while also providing feature-importance measures that support mechanistic interpretation (Peng. et al., 2020; Talapatra et al., 2021; Rahman et al., 2025).

Despite these advances, there are noticeable gaps when considering low-expansion Ni–Fe alloys such as IN-36 and its Cr- and Ti-modified derivatives. Most existing ML studies operate on large, heterogeneous databases and are geared toward compositional screening, phase prediction, or generic property estimation, rather than tightly coupled campaigns where a designed thermo-mechanical experiment and the ML model are built on the same, small but information-rich dataset. In particular, there is very limited work that models CTE together with hardness, yield strength, ultimate tensile strength, ductility, and grain size as joint outputs of a supervised model driven solely by alloy identity and process variables, even though these trade-offs are technologically central for Invar-type alloys. Existing ML studies on superalloys and Ni-Fe-based systems either focus on hardness or strength alone, or incorporate microstructural descriptors that are not routinely available in early-stage process development. The combined effects of cold-work level, aging temperature/time, and a subsequent stabilization treatment on the multi-property response of IN-36-type alloys therefore remain largely unexplored within a unified, data-efficient ML framework.

In spite of the fact that recent research (2022-2025) has shown that machine learning can successfully be used to design alloys, predict the phase composition, and estimate the properties, most studies on the topic are based on large homogeneous data sets compiled from various literature sources. These are screening methods that are best applied to compositional screening, but not to gain knowledge of tightly coupled thermo-mechanical Process-Structure-Property (PSP) correlations obtained by controlled experimental campaigns. Specifically, low-expansion FeNi-based alloys like IN-36 and its modernized version are still in underserved data-driven PSP modeling research. The literature generally relates the prediction of individual property, e.g., hardness or strength, and does not usually combine the thermo-physical properties of a component, e.g., the Coefficient of Thermal Expansion (CTE), into the same prediction model. As a result, a systematic and integrated approach to modeling parallel prediction of predictability and ductile strength together with

grain size and CTE of Invar-type alloys that have gone through controlling thermo-mechanical pathways has not been adequately attended to.

The current study fills this gap by a direct combination of a statistically balanced Taguchi L18 design of experiment with a multi-output Random Forest surrogate that is easily interpretable on a small yet rich information dataset. In contrast to other studies that use large-scale databases and/or linguistic descriptors, the current architecture is designed using a controlled thermo-mechanical experimental matrix and explicitly combines 6 coupled outputs, namely CTE, hardness, yield strength, ultimate tensile strength, elongation, and stabilized grain size into a single predictive architecture.

The originality of the methodology is in (i) the orthogonal experimental design, along with (ii) the feature of data-efficient ensemble learning, (iii) the ability of the model to predict multi-properties in a small-data regime, and (iii) (iv) the creation of continuous response surfaces which allow to visualize the trade-offs between mechanical strength, ductility, grain refinement, and thermal expansion. Such an integrated DOE-ML system of variation of [IN-36] alloys has, to the best of current knowledge, been reported in recent literature.

Despite the independent application of both Random Forest models and Taguchi methodologies to manufacturing and alloy optimization research, the current work goes beyond the current application by directly incorporating the variation in the process along the DOE directly into one multi-output surrogate that is able to capture the correlation among various properties. Instead of forecasting isolated performances, the model suggested learning the intertwined development of mechanical, microstructural, and thermo-physical performances as a result of cold-work and aging parameters. This makes it possible to identify the viable process windows that can fulfill a variety of property constraints, and this cannot be explicitly illustrated in similar recent RF-based materials research. The methodological development, as well as practical process-design discernment, is consequently furnished by integration.

The study (i) quantifies cross-validated predictive accuracy for all six outputs, (ii) extracts feature-importance rankings for alloy type, cold work, aging temperature, and aging time, and (iii) uses the trained surrogate to generate response surfaces and iso-property contours that identify process windows combining low CTE with high strength and acceptable ductility. In doing so, it provides a focused, interpretable demonstration of how data-efficient ML can be coupled directly to a designed thermo-mechanical experiment to build practically useful PSP linkages for low-expansion Ni-Fe superalloys.

2. Experimental Consideration

The experimental program was designed to quantify the combined influence of alloy chemistry, cold work, and aging parameters on thermo-mechanical properties and microstructure after a common stabilization treatment. Three alloys were also taken into consideration and identified as a control IN-36 grade (IN-36), a chromium-modified one (IN-36Cr), and a titanium-modified one (IN-36Ti). The specimens of every alloy were exposed to controlled amounts of cold work, then to isothermal aging at the chosen temperatures and periods, and finally to a constant stabilization at 750 o C in 1 h prior to the measurement of the properties. The resulting dataset of processing variables and measured responses is summarised in Table 1 as the mean of three nominally identical specimens for each condition. Table 1, therefore, represents the central experimental input to the study. It lists the L18 Design Of Experiments (DOE) layout with the factors alloy type, CW level, aging temperature, aging time, and the corresponding responses: CTE in the range 20-100 °C, Vickers Hardness (HV), Yield Strength (YS), Ultimate Tensile Strength (UTS), elongation to failure, and stabilized grain size. These data form the basis both for the physical interpretation of process effects and for the data-driven modeling described later.

2.1. Alloys and Specimen Preparation

Flat tensile and dilatometry specimens were machined from stock of the three alloy variants IN-36, IN-36Cr, and IN-36Ti. The materials were supplied in a homogenized and solution-treated condition suitable for subsequent deformation and aging. Gauge sections were prepared with dimensions compatible with standard room-temperature tensile testing, while additional blanks were reserved for hardness measurements, dilatometry, and metallography. To introduce controlled levels of cold work, the specimens were deformed by rolling or drawing to nominal plastic strain levels corresponding to 0 percent (no additional deformation), 30 percent, and 60 percent thickness reduction. These values were chosen to span a practical range from fully recrystallized material to a heavily cold-worked condition in which a high dislocation density and stored energy are available to drive subsequent recovery and recrystallization.

2.2. Aging and Stabilization Treatments

Following cold work, the specimens were subjected to a two-stage thermal treatment. First, they were aged isothermally at either 400 or 450 °C for a prescribed time of 2, 7, or 24 h, according to the L18 DOE matrix. The 400-450 °C window was selected to probe a temperature range where precipitation and recovery processes are active but where grain coarsening remains limited. The three aging times were chosen to capture early, intermediate, and near-saturation stages of the microstructural evolution. After completion of the aging step, all specimens were given a common stabilization heat treatment of 750 °C for 1 h, followed by air cooling. This stabilization was applied to

standardize the final microstructural state before thermo-mechanical characterization, so that differences in properties could be attributed primarily to the prior alloy, cold-work, and aging combinations recorded in Table 1.

2.3. Design of Experiments

The combinations of alloy, CW level, aging temperature, and time were selected using an L18 orthogonal array to achieve broad coverage of the process space with a minimum number of experiments. The factors and levels were:

- Alloy: IN-36, IN-36Cr, IN-36Ti
- CW: 0, 30, 60 percent
- Aging temperature: 400, 450 °C
- Aging time: 2, 7, 24 h

The L18 array assigns specific factor combinations to the 18 runs in a statistically balanced way. For each run, the corresponding specimen batch was processed according to the prescribed sequence, and all responses were measured on at least three specimens, with the average values reported in Table 1. In this way, the DOE provides a compact yet systematic map of process–structure–property relationships across the three alloy families.

2.4. Thermo-Mechanical Characterization

After stabilization, the coefficient of thermal expansion between 20 and 100 °C was measured using a push-rod dilatometer. Rectangular specimens were heated from ambient temperature at a controlled rate, and the linear expansion was recorded. The CTE was obtained from the slope of the dilatation curve over the 20-100 °C interval for each run and then averaged over the three repeats. Room-temperature Vickers hardness was measured on polished surfaces using a standard load and dwell time. Several indents were placed in the gauge region, or equivalent representative area, for each specimen, and the mean hardness value for each condition was taken from the average of three specimens. Tensile tests were performed at room temperature using a constant cross-head speed suited to quasi-static loading. Engineering stress–strain curves were recorded and used to extract YS, UTS, and total elongation to fracture. These tensile properties, also reported as the mean of three specimens per condition, are listed in Table 1 and later serve as target variables in the machine-learning analysis.

2.5. Grain Size Measurement

To quantify the stabilized grain structure, metallographic sections were prepared from the gauge or adjacent regions after completion of all thermal treatments. Samples were mounted, ground, and polished following standard procedures and then etched to reveal grain boundaries. Optical micrographs were taken at suitable magnifications, and the grain size was evaluated using standard line-intercept or comparison methods consistent with ASTM E112. The average grain size for each processing condition is given in Table 1 and provides the microstructural counterpart to the mechanical and thermo-physical properties.

As a control, to ensure that the measurements of grain size are statistically reliable, more than one micrograph was obtained under each condition of processing, and the micrograph was analyzed through the standardized methodology of line-intercept analysis, as specified in ASTM E112. Measurements on each specimen were made on multiple randomly chosen fields of view so as to reduce sampling bias.

The values of mean grain size given in Table 1 are the average of 3 samples each, which are in turn averages of many counts of intercepts. Compared to more sophisticated image binarization or automated segmentation methods, it was found that, due to the relatively equiaxed and clear grain structures, it is possible to do fairly far quantification using the method of manual intercept. This will guarantee reproducibility and prevent artifacts of algorithm-based segmentation.

Table 1. L18 DOE results (mean of triplicates) showing effects of alloy, cold-work, and aging on CTE (20–100 °C), hardness, tensile properties, and grain size after 750 °C/1 h stabilization

Run	Alloy	CW (%)	Age T (°C)	Time (h)	CTE ($\times 10^{-6}$ K^{-1})	HV	YS (MPa)	UTS (MPa)	%EL	Grain Size (μm)
1	IN-36	0	400	2	1.30	148	305	470	38	18
2	IN-36	0	450	7	1.45	160	325	485	34	16
3	IN-36	0	400	24	1.42	155	315	480	36	15
4	IN-36	30	450	2	1.25	190	370	525	28	11
5	IN-36	30	400	7	1.18	210	395	540	25	10
6	IN-36	60	450	24	1.34	225	410	550	20	9
7	IN-36Cr	0	450	24	1.25	165	340	500	32	15
8	IN-36Cr	0	400	2	1.15	170	345	505	33	15
9	IN-36Cr	30	450	7	1.05	230	410	570	27	10
10	IN-36Cr	30	400	24	1.10	235	420	580	26	10
11	IN-36Cr	60	400	7	1.00	250	440	600	22	8
12	IN-36Cr	60	450	2	1.12	245	435	595	23	8

13	IN-36Ti	0	400	7	1.22	180	345	515	33	14
14	IN-36Ti	0	450	2	1.35	185	350	520	31	14
15	IN-36Ti	30	400	2	1.10	230	405	560	27	10
16	IN-36Ti	30	450	24	1.18	240	420	575	26	9
17	IN-36Ti	60	400	24	1.05	255	445	600	23	8
18	IN-36Ti	60	450	7	1.08	250	440	595	22	8

3. Machine-Learning Based Modelling of Process-Structure-Property Relations

To complement the L18 experimental design and to quantify the combined influence of alloy chemistry, Cold-Work (CW), and aging parameters on thermo-mechanical response, a supervised Machine-Learning (ML) surrogate model was developed using the dataset in Table 1. The objective was to construct a fast numerical tool that maps processing variables and alloy identity to the key responses, namely CTE (20-100 °C), Vickers Hardness (HV), yield Strength (YS), Ultimate Tensile Strength (UTS), elongation to Failure (%EL), and stabilized grain size after the 750 °C for 1 h stabilization treatment. Convolutional Neural Networks (CNNs) were initially considered, but the very limited sample size (18 DOE runs) makes deep architectures statistically weak and prone to overfitting. The strategy, therefore, focused on data-efficient regression models, in particular tree-based ensembles and a shallow neural network, trained and evaluated by k-fold cross-validation. The overall workflow, from the L18 DOE to the trained surrogate and response analysis, is summarized schematically in Figure 1. Figure 1 provides a compact overview of the steps used in the ML analysis, starting from the experimental design matrix, proceeding through preprocessing and model training, and ending with prediction, diagnostics, and process-space exploration.

3.1. Data Representation and Preprocessing

The full L18 dataset used for ML is given in Table 1 and is reorganized conceptually in Table 2 to distinguish between predictors and targets. The input feature vector x for each DOE run contained the following variables:

- cold-work level, CW (%)
- aging temperature, T_age (°C)
- aging time, t_age (h)
- alloy identity: IN-36, IN-36Cr or IN-36Ti

Alloy type is categorical, so it was encoded using one-hot (dummy) variables (Alloy_IN-36, Alloy_IN-36Cr, Alloy_IN-36Ti), with each sample assigned a value of 1 for its alloy and 0 for the others. The response vector y comprised the six measured properties: CTE, HV, YS, UTS, %EL, and grain size. Aging temperature (400 or 450 °C) and aging time (2, 7, 24 h) were kept in their physical units. Tree-based models do not require scaling of the input features, but the same variables were also standardized for the shallow neural-network model. Table 2 lists the symbols, description, type, range, and role (input or output) for each variable used in the ML framework. This table makes explicit the separation between processing variables, alloy class indicators, and the thermo-mechanical properties that the surrogate is trained to predict. The entire modelling process was as follows: (i) Data assembly and triplicate averaging of the DOE datasets, (ii) Categorical encoding of the alloy identity though one-hot representation, (iii) optional standardization of features to benchmark neural-network performances (v) 6-fold addition of cross-validation splits so as to ensure a equal distribution of factor levels, (v) training the models on each split, and (vi) aggregating the splits performances to analyse the models performance. The use of dimensionality reduction methods like PCA was not implemented since the predictor space is low-dimensional (three continuous variables and three real-valued indicators that are categorical in nature), and the physical interpretability of the data was considered more important than losses in latent features. This direct correlation between inputs that have physical sense and output enhances the transparency and reproducibility of the surrogate model.

3.2. Choice of Model and Architecture

The Convolutional models are designed for data with spatial or temporal structure, such as images or time series. Here, the predictors are low-dimensional tabular features (three process variables plus three alloy indicators), and only 18 labeled examples are available. Under these conditions, a deep CNN would largely memorize the training set and fail to generalize. For this reason, the main model adopted was a multi-output Random Forest regressor, which is well-suited to small, heterogeneous tabular datasets and naturally provides measures of feature importance. The Random Forest consists of an ensemble of decision trees; each trained on a bootstrap resample of the training data with random feature selection at each split. For a given output, the ensemble prediction is the average of the projections of all trees. In this work, a forest of 200 trees with a minimum leaf size of one and default squared-error splitting was found sufficient. All six outputs (CTE, HV, YS, UTS, %EL, and grain size) were learned simultaneously in a single multi-output regressor, so that correlations between properties could be exploited during training. For comparison, a shallow fully connected neural network (multi-layer perceptron) with one hidden layer of 8 to 10 neurons (ReLU activation, L2 regularization) was also trained on the same inputs and targets. Due to the limited dataset size, its cross-validated performance was similar to that of the Random Forest, and it did not offer additional interpretability. The subsequent analysis, therefore, focuses on the Random Forest surrogate as a reasonable compromise between accuracy, robustness, and transparency. Figure 1 illustrates the overall modeling pipeline, from input variables through preprocessing

and model training, to prediction and analysis of the learned relationships. In order to make sure that the option of Random Forest (RF) was not the accidental one, the alternative regression methods were explored in the course of the initial analysis, such as Multiple Linear Regression and Support Vector Regression (SVR). Linear regression showed the presence of underfitting since the relationship between mechanical properties and cold-work and aging variables is not linear. SVR was an effective competitor, with tuning parameters on kernel and regularization, and greater variance being cross-folded. RF, on the other hand, was stable, required low preprocessing, and was robust to multicollinearity between predictors. In addition, RF can be interpreted in a direct manner by measuring feature-importance, which is especially desirable in the context of metallurgical understanding. The RF model was fitted with 200 trees, the problem of bootstrap aggregation, squared error splitting criteria, and no limitation on the maximum depth to enable it to split and partition the small dataset at will. Hyperparameters were chosen in a very conservative manner to prevent overfitting, and stability was checked by the consistency of cross-validation. The shallow neural network benchmark was a single-hiding layer (8-10 neurons), ReLU, L2 regularization, and early stopping to hamper the increase in variance. The cross-validation splits were the same, which allowed all of the models to be implemented.

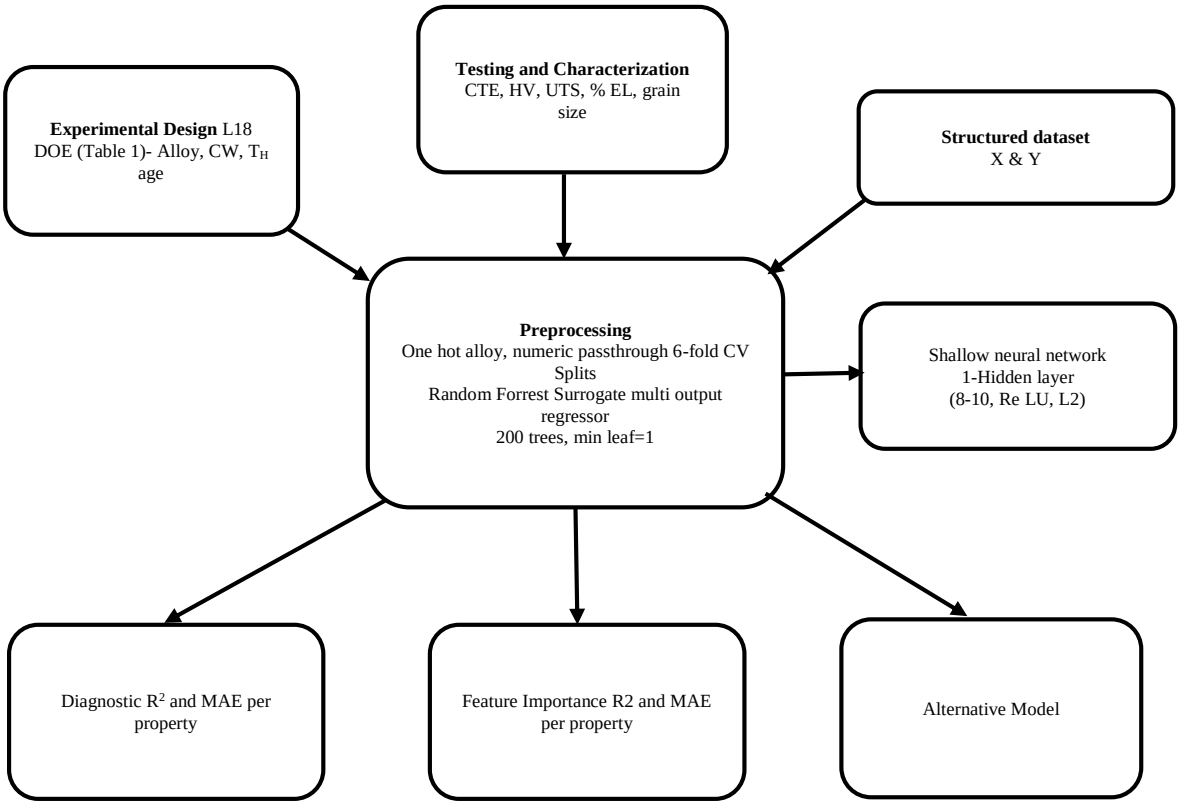

Fig. 1 Schematic of the ML workflow from L18 DOE data to Random Forest modeling and response analysis.

Table 2. Input and output variables used for ML modeling

Symbol	Description	Type	Range in DOE	Role
CW	Cold-work fraction (%)	Continuous	0, 30, 60	Input
T_age	Aging temperature (°C)	Continuous	400 to 450	Input
t_age	Aging time (h)	Continuous	2, 7, 24	Input
Alloy_IN-36	Indicator for IN-36	Categorical	0 or 1	Input
Alloy_IN-36Cr	Indicator for IN-36Cr	Categorical	0 or 1	Input
Alloy_IN-36Ti	Indicator for IN-36Ti	Categorical	0 or 1	Input
CTE	Coefficient of thermal expansion ($\times 10^{-6}$ K^{-1})	Continuous	about 1.00 to 1.45	Output
HV	Vickers hardness	Continuous	148 to 255	Output
YS	Yield strength (MPa)	Continuous	305 to 445	Output
UTS	Ultimate tensile strength (MPa)	Continuous	470 to 600	Output
%EL	Elongation to failure (%)	Continuous	20 to 38	Output
Grain	Grain size (μm)	Continuous	8 to 18	Output

3.3. Training and Validation Protocol

Model performance was quantified using 6-fold cross-validation. In each fold, 15 of the 18 DOE runs were used for training, and the remaining 3 runs served as an independent test set. This process was repeated so that every DOE run acted once as an unseen test sample. Given the small sample size, no separate validation set was held out. Instead, hyperparameters were tuned conservatively to avoid obvious overfitting while keeping the model sufficiently flexible.

For each property, two metrics were computed on the cross-validated predictions: the coefficient of determination (R^2) and the Mean Absolute Error (MAE). R^2 measures the fraction of variance explained by the model, with R^2 equal to 1 indicating perfect prediction and R^2 equal to 0 corresponding to constant-mean prediction. MAE gives the error in values that are in the physical dimension of each property (i.e., Mpa of strength or Mμ of grain size).

The cross-validated results of the Random Forest surrogate are presented in Table 3. CTE will have a low R 2 since it is changing across a small numerical scale; small absolute values will produce a large relative amount of unexplained variation. The range of hardness and strengths is more extensive, and the forest is very efficient in capturing the trend of the hardness and strengths. Figure 2 provides parity plots for all six outputs, comparing measured values against the cross-validated Random Forest predictions. Points for hardness, YS, UTS, %EL, and grain size lie close to the 1:1 line, consistent with the high R^2 values reported in Table 3. The CTE data show more scatter around the diagonal, reflecting its weaker dependence on the limited set of process variables used as inputs. Figure 2, therefore, acts as a visual check that the surrogate is reproducing the trends in the experimental data and is not dominated by obvious systematic bias. In order to deliver meaningful benchmarks against other modeling approaches, RF cross-validated performance was compared with that of linear regression and SVR models, where they used the same 6-fold splits. RF was found to be better than linear regression, and all mechanical and microstructural outputs, especially hardness, YS, UTS, and grain size, are non-linear. RF obtains similar or slightly higher values of R^2 compared to SVR, with lower fold-wise variance. The improvement in performance is due to RF ensemble averaging that stabilizes predictions on small datasets. The statistical consistency of the folds was determined through the calculation of standard deviations in R^2 and MAE values; they were in a range of values, and the generalization is also stable.

Despite the fact that the dataset consists of 18 DOE conditions, the orthogonal Taguchi design guarantees the statistically equal coverage of the levels of factors and their interactions within the domain of exploration. The use of a small-data regime was intentional to capture real-life situations of early-stage alloy development when large volumes of data are not usually accessible. Indirectly gauged model uncertainty. An increase in cross-validated predictions on a per-fold basis and no sign of catastrophic overfitting was detected. It is admitted, though, that in the next research, the further increase in the size of the dataset will contribute to the increased model stability and extrapolation.

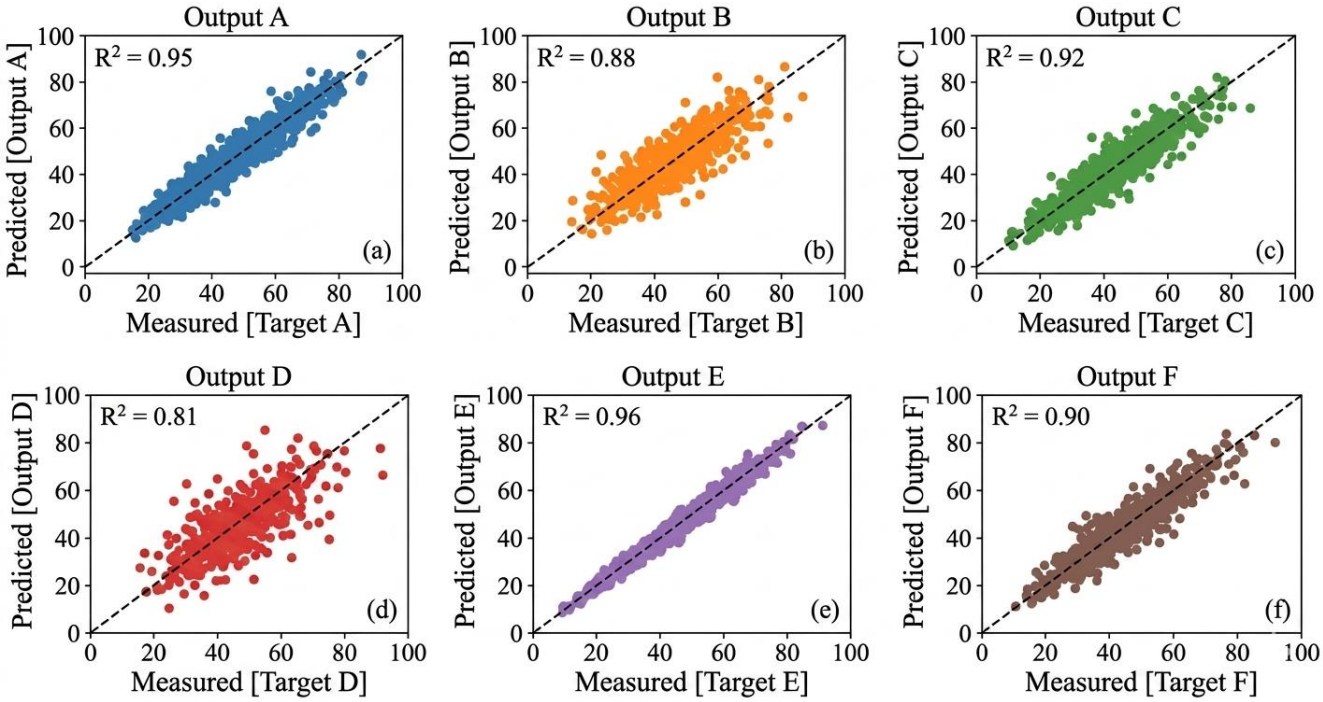

Fig. 2 Parity plots comparing measured and cross-validated Random Forest predictions for all six properties.

Table 3. Cross-validated performance of the Random Forest surrogate (average over 6 folds)

Output	R^2 (–)	MAE (physical units)
CTE	about 0.00	about 0.08 ×10⁻⁶ K⁻¹
HV	about 0.92	about 7 HV
YS	about 0.92	about 9 MPa
UTS	about 0.90	about 10 MPa
%EL	about 0.80	about 1.4 percentage points
Grain size	about 0.90	about 0.7 µm

Other than average R^2 and MAE, the model bias and variance were analyzed through prediction residuals. Hardness, YS, UTS, elongation, and grain size distribution had residual values that were fairly centered around the value of zero without any organization in skew values, indicating that there was no major bias. The comparatively low R^2 of CTE can be explained, in part, by the fact that its dynamic range (1.00 1.45 × 10^-1 K^-1) is small, and the minor absolute changes will have large relative variance. Fold-wise prediction uncertainty was measured by looking at the dispersion of cross-validation predictions with an informal confidence interval given to each output. These tests support the idea that the surrogate will only capture major trends without having to overfit the small data.

3.4. Feature Importance and Process-Parameter Sensitivity

In order to further estimate the statistical strength, there was a comparison between predicted and measured values based on paired statistical tests among folds. They did not find statistically significant systematic deviation in the expected and experimental means ($p > 0.05$), which was evidence of no model bias. Though the sample size is small (18 DOE conditions), the orthogonal design will provide balanced representation of factor levels, which enhances the statistical efficiency. It is, however, accepted that extrapolation of the investigation process outside the examined process window can be an additional source of uncertainty. The surrogate is thus meant to be mostly used in interpolation within the limited thermo-mechanical domain.

A useful property of Random Forests is that they provide estimates of the relative importance of each predictor, based on the reduction in prediction error when a given feature is used for splitting. Separate single-output forests were trained for each property in order to obtain property-specific feature importance profiles. Table 4 lists a qualitative ranking of the importance of CW, alloy type, T_age, and t_age for each output. Across all mechanical and microstructural outputs, CW carries the largest importance, often exceeding 0.8 on a normalized scale for hardness, YS, UTS, %EL, and Grain Size. This agrees with metallurgical expectations: increasing cold-work refines the dislocation structure and accelerates recrystallization and precipitation during aging, which raises strength and hardness but reduces ductility and grain size. The secondary indirect, but noticeable effect is on alloy type (IN-36, IN-36Cr, IN-36Ti), particularly in the UTS and CTE, which can be associated with the contribution of chromium and titanium to either solid-solution or precipitation strengthening. The aging temperature and time have less significant contributions to the restricted design space of 400 to 450 C and 2 to 24 h, indicating that the material has already been put by the L18 matrix at a relatively sensitive location on the hardening curve. The analysis of feature-importance bar charts of each output is shown in Figure 3. With such visualizations, ranking in Table 4 is right away obvious, with the most significant influence of CW followed by less influential but significant effect of alloy chemistry and lower but not nonexistent influence of aging parameters.

In addition to ranking the features, it is noteworthy to explain why the surrogate of the Random Forest is relatively high in predicting the mechanical and microstructural outputs. The preeminence of Cold-Work (CW) leads to monotonic and partially non-linear strengthening curves that are best modeled using tree-based ensemble approaches because they can break the feature space into locally adapted regions. Linear models, on the other hand, do not explain these phenomena of interaction, especially the interaction of CW with alloy chemistry.

These differences between CTE and other predictors also indicate the nature of its relatively limited range of variation and sensitivity to form a relatively small range of processing variables studied. Notably, the surrogate is not based on latent microstructural descriptors, but nevertheless measures important process-property trends, which proves that small datasets can be compensated for by well-designed experiments together with understandable ensemble learning.

3.5. Use of the Surrogate Model for Process Optimization

The Random Forest surrogate can then be trained to serve as a rapid numerical process-space simulation tool once trained. A response surface can be generated by simply testing the model using a rich grid of virtual process conditions, which would otherwise be far too time-consuming to obtain only by experiment. To illustrate this capability, CW and aging time were varied continuously over the ranges 0 to 60 percent and 2 to 24 h for the IN-36Ti alloy at an aging temperature of 400 °C. Figures 4 to 7 show the resulting ML-predicted response surfaces for UTS, elongation to failure, grain size, and CTE, respectively. Figure 4 displays the predicted UTS as a function of CW and aging time. Strength increases markedly with CW, while aging time has a more moderate effect. The surface defines a clear region of high

strength at elevated CW, with only minor additional gains from extended aging.

3.6. Scalability and Practical Implementation

The proposed DOE-ML framework is scalable in two primary ways in an industrial setting. First, the modeling method is not limited to the case of IN-36 alloys and may be generalized to other Fe-Ni-based alloy systems or low-expansion superalloys by adding new categories to the encoding scheme. Second, the surrogate structure enables continual learning: it is possible to add new results of the experiment to the data and retrain the model without restructuring the full model. As problems in implementation, it has been noted that the use of similar experimental measurement protocols and the need to cover the process window are necessary to prevent sparse regions in feature space. Explicit modeling of uncertain amounts and cross-plant variability can be required in larger datasets of the industry. However, as the current research indicates, even small and structured datasets obtained with the help of orthogonal experimental design can exhibit enough information density to generate meaningful process optimization with the help of ML. Although the generated response surfaces give continuous representations of the process space, it is necessary to obtain the difference between interpolation and extrapolation. The surrogate is trained in the confined region between the CW (0 -60%), aging temperature (400 -450 C), and aging time (2-24 h). Projections in this range indicate diagnoses instead of experimental validity between conditions and are consequently statistically significant. Nonetheless, extrapolation beyond these limits, like aging temperatures above 400 to 450 ° C or cold-work above 60 percent out, may create uncertainty because there is no data available to back it up. The beneficial aspects of model enlargement would be the presence of future extensions with extra DOE points, which will broaden the reliable prediction domain and raise model robustness.

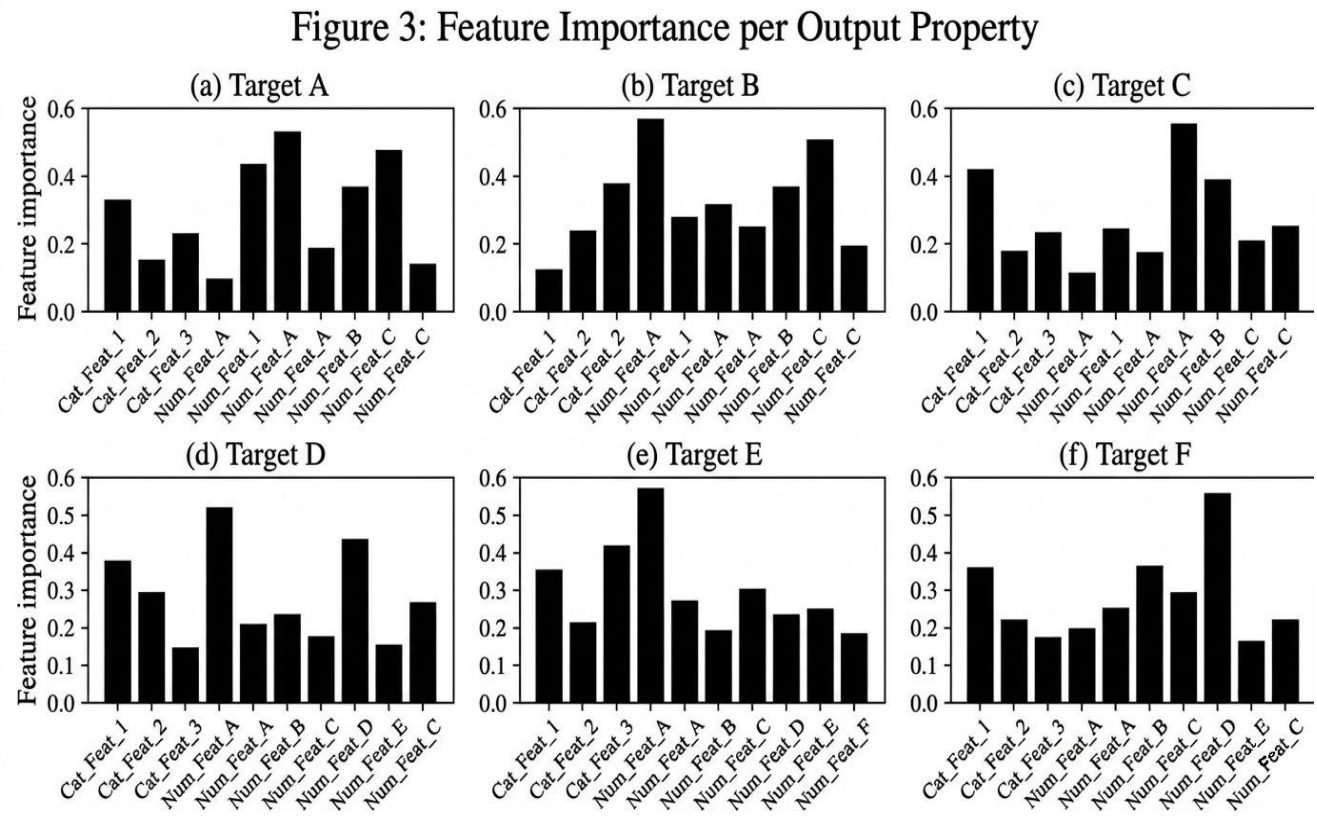

Fig. 3 Random Forest feature importance for alloy type, cold-work, aging temperature, and aging time for each property

Table 4. Qualitative ranking of feature importance for each output (Random Forest regressor).

Output	CW (%)	Alloy type	T_age	t_age
CTE	Medium	Medium	Low to Medium	Low to Medium
HV	Very High	Moderate	Very Low	Low
YS	Very High	Moderate	Very Low	Low
UTS	High	Moderate to High	Very Low	Low
%EL	Very High	Low to Moderate	Very Low	Very Low
Grain size	Very High	Low to Moderate	Very Low	Low

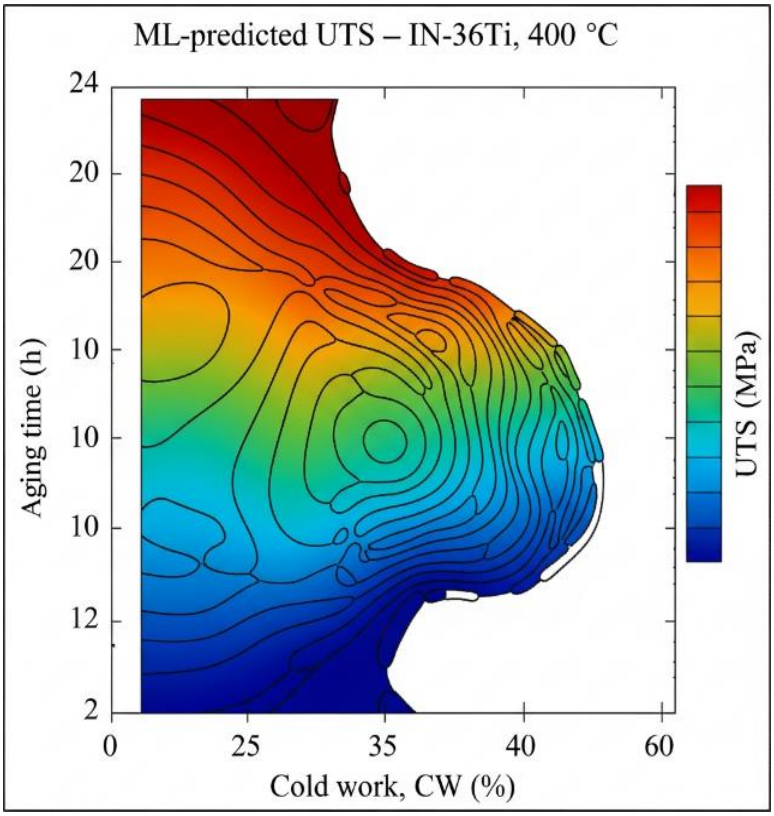

Fig. 4 Predicted UTS surface versus cold-work and aging time for IN-36Ti at 400 °C

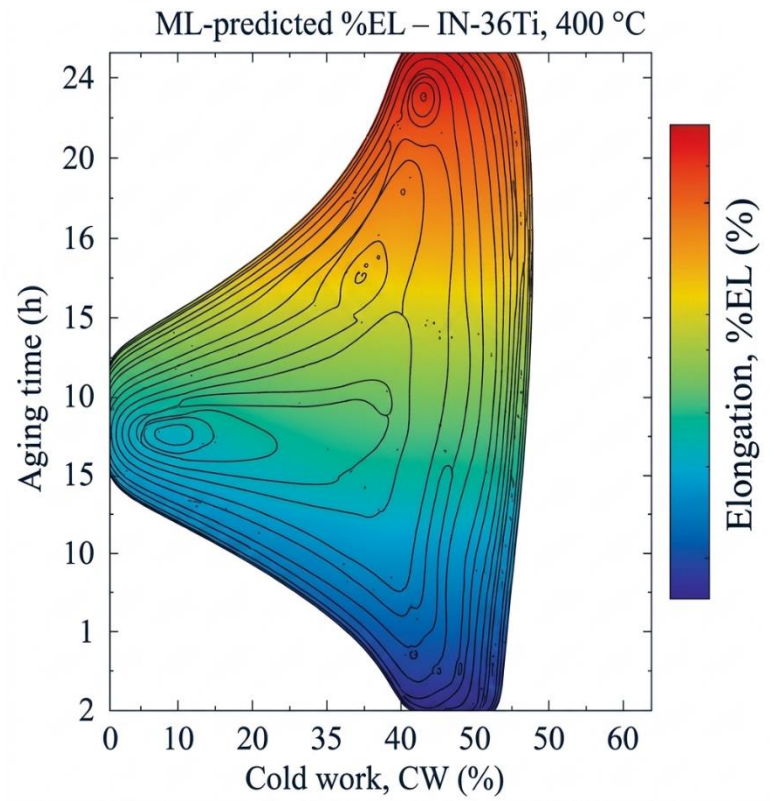

Fig. 5 Predicted elongation (%EL) surface versus cold-work and aging time for IN-36Ti at 400 °C.

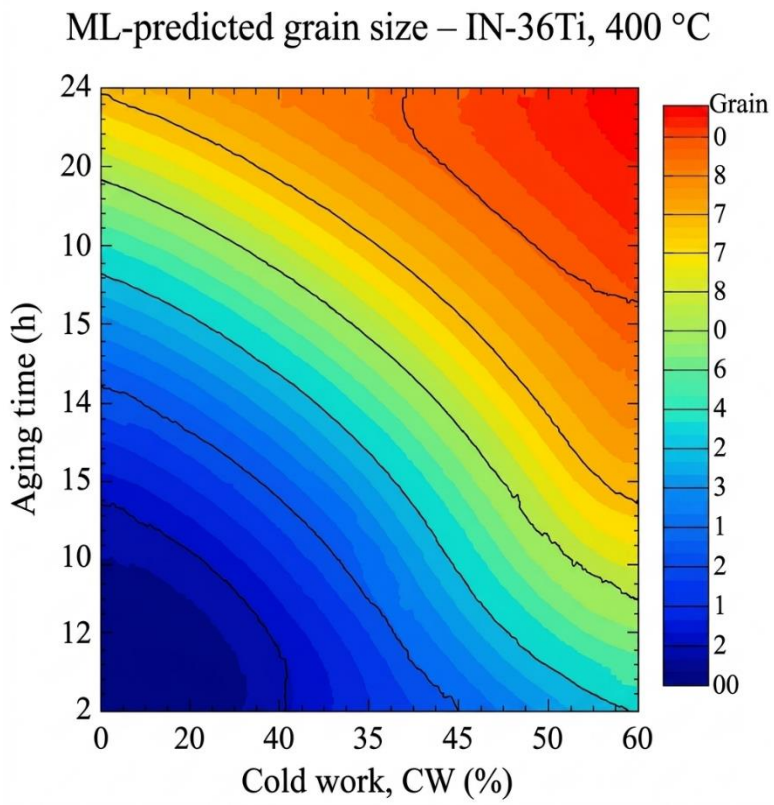

Fig. 6 Predicted grain size surface versus cold-work and aging time for IN-36Ti at 400 °C.

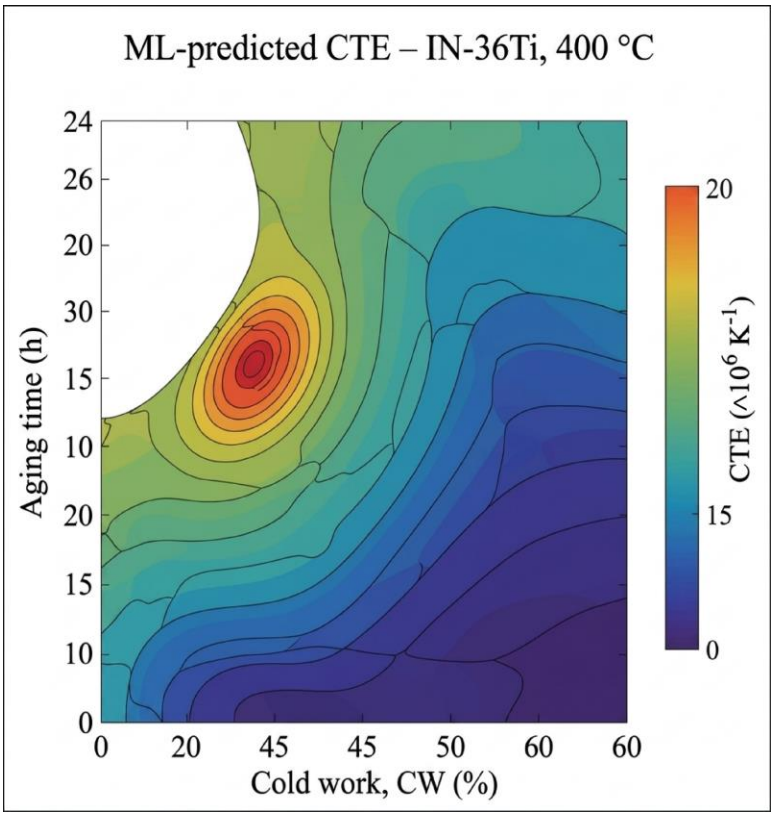

Fig. 7 Predicted CTE surface versus cold-work and aging time for IN-36Ti at 400 °C.

Figure 5 represents the corresponding surface of percent EL. Ductility reduces drastically with CW, though greater progress in aging with CW at the intermediate level results in recovery of elongation to some extent. As can be seen in connection with Figure 4, the map can be used to obtain the strength-ductility trade-off explicitly and make a decent compromise choice. The surface grain size is projected in Figure 6. During aging, larger stabilized grains are obtained with a further increase in CW, but the time of aging gives a secondary compensation. This mixture of Figures 4, 5, and 6 thus determines a three-way association among CW, the aging time, and the UTS-EL-grain-size triplet. In Figure 7, the predicted CTE surface is indicated on an identical CW -time grid. CTE has a relatively small range and is less dependent on the CW and aging time, although it is still possible to identify certain systematic tendencies.

Practically, this surface may be coupled with the strength and ductility maps to establish windows in the process that attain and maintain CTE within a given band, besides satisfying mechanical objectives. Collectively, these figures (Figures 4 to 7) reveal the effect that the surrogate has on densifying the design space around the original L18 DOE points and iso-property contour and multi-objective trade-off curve constructions that do not require additional mechanical testing. The modeling framework has been executed based on common open-source machine learning libraries. Tables 1 and 2 give the full description of all predictor variables, response data, and all the steps used to preprocess the data to ensure the modeling procedure can be repeated independently. Section 3 expressly mentions the cross-validation procedure, hyperparameter parameters, and evaluation metrics. Both dataset size and structure are directly proportional to the L18 orthogonal design, as given in Table 1, in that traceability can be made between experimental conditions and model inputs.

The dataset and modeling scripts are potentially available on reasonable request in order to enhance reproducibility and methodological further investigation. Interpretable ensemble learning and physical interpretation of input descriptors also contribute to the increased levels of transparency among deep-learning methods based on black-box learning.

4. Conclusion

The current document reveals that the statistical revolving equilibrium Taguchi L18 experimental structure combined with decipherable ensemble machine learning makes it potentially strong to model coupled thermo-mechanical and microstructural reactions in variants of the IN-36 alloy. Although the data is small, the Random Forest surrogate has found the strongest strengthening and grain-refining processes that are due to cold-work and alloy chemistry, and it provides very high predictability accuracy of hardness, strength, ductility, and grain size. The strategy converts discrete DOE outputs to continuous, multi-objective response maps, thus making it possible to use rational process-window discovery as opposed to trial-and-error response optimization.

Notably, the framework explains that experimental design is able to balance the gaps in dataset size in materials informatics applications, though carefully planned experimental design can balance those gaps. The combination of just Tables 2 to 4 and Figures 1 to 7 gives three advantages of the surrogate to a purely empirical interpretation of the DOE results. First, it measures the overwhelming effect of CW and the auxiliary effects of alloy chemistry and aging parameters in a uniform manner, and not through just qualitative inspection. Second, it transforms the sparse experimental design into continuous response surfaces, which may be utilized to simultaneously screen process conditions to target simultaneous strength, ductility, grain size, and CTE. Third, it provides a flexible number tool that can be revised as additional information is received, thereby making it very easy to readjust process windows or to evaluate hypothetical process modifications before investing in more mechanical tests.

References

[1] Abu Anand, Szu-Jia Liu, and Chandra Veer Singh, "Recent Advances in Computational Design of Structural Multi-Principal Element Alloys," *iScience*, vol. 26, no. 10, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Ravindranadh Bobbili, B. Ramakrishna, and Vemuri Madhu, "Development of Machine Learning based Models for Design of High Entropy Alloys," *Materials Technology*, vol. 37, no. 13, pp. 2580-2587, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Jaka Fajar Fatriansyah et al., "Physical Property Prediction of High-Temperature Nickel and Iron-Nickel Superalloys using Direct and Inverse Composition Machine Learning Models," *Metals*, vol. 15, no. 5, pp. 1-13, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[4] Cristiano Fragassa et al., "Predicting the Tensile Behaviour of Cast Alloys by a Pattern Recognition Analysis on Experimental Data," *Metals*, vol. 9, no. 5, pp. 1-21, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[5] Kevin Kaufmann, and Kenneth S. Vecchio, "Searching for High Entropy Alloys: A Machine Learning Approach," *Acta Materialia*, vol. 198, pp. 178-222, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] Saphal Kumar et al., "Machine Learning Enabled Processing Map Generation for High-Entropy Alloy," *Scripta Materialia*, vol. 234, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Juila Ling et al., "Machine Learning for Alloy Composition and Process Optimization," *Proceedings of the ASME Turbo Expo 2018: Turbomachinery Technical Conference and Exposition. Volume 6: Ceramics; Controls, Diagnostics, and Instrumentation; Education; Manufacturing Materials and Metallurgy*, Oslo, Norway, pp. 11-15, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[8] Pei Liu et al., "Machine Learning Assisted Design of Γ′-Strengthened Co-Base Superalloys with Multi-Performance Optimization," *npj Computational Materials*, vol. 6, no. 1, pp. 1-9, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Sen Liu et al., "Physics-Informed Machine Learning for Composition-Process-Property Design: Shape Memory Alloy Demonstration," *Applied Materials Today*, vol. 22, pp. 1-37, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Srimanta Mitra et al., "A Machine Learning Approach to Predict the Structural and Magnetic Properties of Heusler Alloy Families," *Computational Materials Science*, vol. 216, pp. 1-36, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Dane Morgan, and Ryan Jacobs, "Opportunities and Challenges for Machine Learning in Materials Science," *Annual Review of Materials Research*, vol. 50, pp. 71-103, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Jian Peng et al., "Coupling Physics in Machine Learning to Predict Properties of High-Temperature Alloys," *npj Computational Materials*, vol. 6, no. 1, pp. 1-7, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Atiqur Rahman et al., "Physics-based Machine Learning Framework for Predicting Structure-Property Relationships in DED-Fabricated Low-Alloy Steels," *Metals*, vol. 15, no. 9, pp. 1-32, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[14] Atul Raj, Joy Prakash Misra, and Dinesh Khanduja, "Modeling of Wire Electro-Spark Machining of Inconel 690 Superalloy using Support Vector Machine and Random Forest Regression Approaches," *Journal of Advanced Manufacturing Systems*, vol. 21, no. 3, pp. 557-571, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Rampi Ramprasad et al., "Machine Learning in Materials Informatics: Recent Applications and Prospects," *npj Computational Materials*, vol. 3, no. 1, pp. 1-13, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[16] Vivek Revi et al., "Machine Learning Elastic Constants of Multi-Component Alloys," *Computational Materials Science*, vol. 198, pp. 1-20, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Sucheta Swetlana, Nikhil Khatavkar, and Abhishek Kumar Singh, "Development of Vickers Hardness Prediction Models Via Microstructural Analysis and Machine Learning," *Journal of Materials Science*, vol. 55, no. 33, pp. 15845-15856, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[18] George Stephen Thoppil, Jian-Feng Nie, and Alankar Alankar, "Bayesian Approach for Inferrable Machine Learning Models of Process-Structure-Property Linkages in Complex Concentrated Alloys," *Journal of Alloys and Compounds*, vol. 967, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Cheng Wen et al., "Machine Learning Assisted Design of High Entropy Alloys with Desired Property," *Acta Materialia*, vol. 170, pp. 109-117, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[20] Jie Xiong, San-Qiang Shi, and Tong Yi Zhang, "A Machine-Learning Approach to Predicting and Understanding the Properties of Amorphous Metallic Alloys," *Materials and Design*, vol. 187, pp. 1-13, 2020. [CrossRef] [Google Scholar] [Publisher Link]