

Article

A Transfer-Learning and Continuous Optimization-Based Framework for Predicting Heat Treatment-Dependent Mechanical Properties of DED-Processed Low-Alloy Steels

Atiqur Rahman , Sung-Heng Wu, Ranjit Joy  and Frank Liou * 

Department of Mechanical and Aerospace Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA; a.rahman@mst.edu (A.R.); swm54@mst.edu (S.-H.W.); rjyww@mst.edu (R.J.)

* Correspondence: liou@mst.edu

Abstract

Directed energy deposition (DED) of low-alloy steels involves strongly coupled effects among alloy composition, solidification behavior, and post-deposition heat treatment, making mechanical property prediction difficult when target-domain data are limited. This study develops a transfer-learning and continuous optimization framework for predicting heat treatment-dependent yield strength (YS), ultimate tensile strength (UTS), hardness (HV), and as-solidified phase fractions of martensite, ferrite, and austenite in DED-processed low-alloy steels. A CALPHAD-based dataset was generated for 125 low-alloy steel compositions. A multilayer perceptron (MLP) surrogate was first trained as a baseline model, then fine-tuned through transfer learning and progressively updated as staged continuous optimization; the composition pool increased from 72 to 125 compositions using Random, Greedy, and Bayesian upper-confidence-bound acquisition strategies. The heat treatment prediction accuracy improved from an average R^2 of 0.757 for the baseline model to 0.929 after transfer learning and to approximately 0.997 after continuous optimization, with a nearly 78% reduction in RMSE relative to transfer learning. For the solidification outputs, the average R^2 increased from 0.770 after transfer learning to approximately 0.859 after optimization. Bayesian-UCB provided the most stable and data-efficient improvement by balancing predicted performance with model uncertainty. The optimized prediction system showed low case-study errors for both solidification and heat treatment properties, demonstrating its potential as a rapid screening tool for alloy composition and tempering-condition selection in DED low-alloy steel development.

Keywords: DED; solidification; heat treatment; transfer learning; continuous optimization



Academic Editor: Thomas Niendorf

Received: 26 June 2026

Revised: 5 August 2026

Accepted: 6 August 2026

Published: 10 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Metal additive manufacturing (AM) has created new opportunities to produce complex engineering components with reduced material waste, shorter lead times, and improved repair capability. Among the available metal AM technologies, directed energy deposition (DED) is particularly attractive for the repair, remanufacturing, cladding, and near-net-shape fabrication of large metallic components, as the feedstock powder or wire can be delivered directly into a melt pool generated by a focused energy source [1,2]. These characteristics are especially relevant for low-alloy steels, which are widely used in gears, bearings, structural components, tooling, pipelines, and repair applications due to their desirable combination of strength, hardenability, toughness, weldability, and relatively low cost [3].

However, the material response of DED-processed low-alloy steels is substantially different from that of conventionally processed steels. Localized heating, rapid solidification, repeated reheating by subsequent layers, and heat accumulation during deposition generate complex thermal histories that vary throughout the manufacturing process. Typical cooling rates in laser-directed energy deposition are on the order of 10^3 – 10^4 °C/s, and the resulting microstructure may contain martensite, bainite, ferrite, retained austenite, or mixed transformation products depending on alloy chemistry and local thermal conditions [3–5].

The sensitivity of DED steels to composition and thermal history makes predicting mechanical properties challenging. Bartels et al. investigated Bainidur AM low-alloyed case-hardening steel processed by DED-LB/M using 16MnCr5 substrates. The additively manufactured material developed a bainitic–martensitic microstructure with retained austenite, demonstrating the non-equilibrium character of the as-built state. The Bainidur AM specimens retained a comparatively homogeneous hardness of approximately 400 HV1 after tempering treatments up to 600 °C, whereas conventionally hardened material exhibited a more pronounced hardness reduction [3]. Similarly, Kelley et al. showed that cooling-rate-driven changes between lath martensite and upper bainite were a principal source of hardness sensitivity in laser-DED steels. Their results further demonstrated that alloy composition can strongly affect the robustness of hardness, yield strength, and ultimate tensile strength against variations in laser power and interlayer delay time [4]. Ebrahimnia et al. reported that direct laser metal deposition of 24CrNiMo low-alloy steel was strongly affected by build condition; single-layer deposits were produced without major cracks or pores, whereas multilayer deposits showed crack formation [6]. Wang et al. showed that HSLA low-alloy steel could be successfully fabricated by direct laser deposition without larger defects, further confirming the processability of DED low-alloy steels [7]. These findings indicate that the properties of DED-processed low-alloy steels cannot be adequately described by nominal material composition alone. Instead, their final performance is governed by an interconnected composition–thermal history–microstructure–property relationship.

Post-deposition heat treatment is often required to control the final state of DED steel components and to relieve processing-induced residual stresses. In low-alloy steels, tempering alters the martensitic or bainitic structure through carbon redistribution, carbide precipitation, recovery of dislocations, and relaxation of residual stress. Consequently, heat treatment changes the balance between yield strength, ultimate tensile strength, hardness, ductility, toughness, and service reliability. This effect is particularly important in DED because the as-built condition may include high residual stress, local micro-segregation, refined substructures, and non-equilibrium phase fractions. For example, Deirmina et al. [8] compared direct tempering and quenching-plus-tempering routes for laser-DED H13 tool steel. At a similar hardness level of approximately 500 HV1, the quenching-plus-tempering condition achieved a higher fracture toughness of 89 MPa $\sqrt{\text{m}}$ than the directly tempered condition, which reached 70 MPa $\sqrt{\text{m}}$ [8]. In a DED 316L study, Yan et al. observed that the cellular structure became progressively coarser with increasing annealing temperature and disappeared at temperatures above 1000 °C, accompanied by a decrease in mechanical strength [9]. Liu et al. further demonstrated that post-manufacturing heat treatment of DED 316L reduced cellular substructures and dislocation density, improving hydrogen-embrittlement resistance, and the hydrogen-embrittlement index decreased from approximately 19% in the as-deposited condition to approximately 10% after heat treatment at 1200 °C [10]. Although these studies involve different alloys and heat treatment objectives, they consistently show that post-processing is a primary engineering variable for tailoring the performance of DED components.

The large number of interacting variables involved in DED and heat treatment limits the efficiency of conventional trial-and-error development. Experimental fabrication of

many alloy compositions, followed by multiple heat treatment schedules, microstructural characterization, and mechanical testing, is costly and time-consuming. Machine learning (ML) offers a practical route to learning nonlinear process–structure–property relationships from available data and to rapidly estimating material properties across broad design spaces. In AM, Xie et al. developed a mechanistic data-driven framework that used process-induced thermal histories to predict location-dependent ultimate tensile strength, yield strength, and elongation in DED-fabricated components [11]. In conventional alloy-property prediction, Fu et al. showed that incorporating alloy composition, grain size, and hardness into ML models substantially improved tensile-strength prediction for aluminum alloys, and their gradient-boosted model achieved an R^2 value of 0.914 [12]. Similarly, Shaheen et al. used deep neural networks to predict the elevated-temperature mechanical properties of high-strength steels from composition and temperature, demonstrating the ability of ML to represent coupled chemical and thermal effects [13]. These studies support the use of nonlinear regression algorithms, including neural networks and gradient-boosted decision trees, for predicting mechanical properties from structured composition and processing data.

Despite this progress, ML performance is fundamentally limited by the size, quality, and coverage of the available dataset. Experimental materials datasets are often small, biased toward successful compositions or processing windows, and insufficient to represent the full alloy-design space. Calculation of Phase Diagram (CALPHAD)-based computational tools provide an important route for expanding this design space while retaining thermodynamic and metallurgical consistency. Such tools can estimate phase stability, solidification behavior, transformation response, and property-related descriptors across a broad range of compositions and thermal conditions. Luo et al. demonstrated a CALPHAD-based integrated computational materials engineering framework in which thermal histories from steel-casting simulations were linked to micro-segregation and microstructure prediction; the predicted micro-segregation trends agreed with experimental observations [14]. Li et al. integrated CALPHAD calculations with ML models to design reduced-activation ferritic–martensitic steels with targeted microstructures and tensile properties, achieving an ultimate tensile strength of approximately 100 MPa higher at 600 °C than conventional reference steels for nuclear applications [15]. These studies show that CALPHAD-derived data can provide a physically meaningful source domain for ML, particularly when direct experimental coverage is limited.

However, a model trained only on an initial source dataset may not fully generalize to new alloy compositions or heat treatment conditions because the target domain can contain different composition ranges, tempering responses, and phase-transformation behavior. Transfer learning provides a strategy for addressing this source-to-target discrepancy. In transfer learning, knowledge learned from a source dataset is reused and adapted to a related target domain, enabling the construction of a strong predictor without training each new model from scratch. Tang et al. reviewed transfer-learning applications in AM and emphasized that this approach is particularly valuable when only limited data are available for a new product, machine, material, or processing condition [16]. Liu et al. demonstrated cross-machine knowledge transfer for metal AM, showing that information from existing printer data can provide a more effective starting point for the qualification and optimization of a new AM system [17]. In alloy-property modeling, Sun et al. developed a framework in which models were pre-trained using JMatPro and Thermo-Calc simulation data and then fine-tuned using experimental data; the transfer-learning approach improved prediction performance by approximately 10% compared with conventional ML models trained only on experimental data [18].

Recent alloy-design studies further demonstrate the value of transfer learning when the target dataset is small or lies outside the original training domain. Jia et al. transferred knowledge from an aluminum-alloy dataset containing 1089 records to a particle-reinforced aluminum matrix composite dataset containing 192 entries, reporting R^2 values of 0.924 ± 0.037 for ultimate tensile strength and 0.908 ± 0.044 for elongation [19]. Yang et al. applied deep transfer learning to the prediction of ultra-high-temperature creep rupture life in single-crystal superalloys. By transferring knowledge from a larger conventional superalloy dataset to a target domain containing only 19 ultra-high-temperature creep records, their physically constrained transfer-learning model achieved an extrapolation-space mean absolute error of 21 h [20]. A recent review of transfer learning for metallic materials also reported that simulation-pretrained models can substantially improve predictions when only limited high-fidelity target data are available. For example, a hot-wire laser metal deposition study reduced the melt-pool-dimension prediction error from 21.16% to 2.03% after transfer learning using low-fidelity simulation data and limited experimental data [21]. These results suggest that a simulation-trained source model can provide useful prior knowledge for predicting mechanical properties in new DED low-alloy steel systems, provided that the learned representation is subsequently adapted to the target domain.

In addition to transfer learning, the model should improve as additional simulation and experimental data are generated during alloy and heat treatment development. This adaptive capability is referred to in the present work as continuous optimization: new target-domain observations are progressively incorporated into the framework, enabling the model to update its representation of composition–heat treatment–property relationships and to indicate which new conditions are most informative to evaluate next. Related active-learning and Bayesian-optimization studies have demonstrated that iterative model updating can reduce the number of costly experiments required for alloy discovery. Ghorbani et al. used Bayesian optimization to identify magnesium alloy compositions and processing conditions associated with target mechanical properties, while balancing exploration of uncertain regions and exploitation of high-performing regions [22]. Gao et al. integrated high-throughput CALPHAD calculations, ML, Bayesian optimization, experiments, and data feedback to explore Al-Si-Mg-Sc casting alloys over a high-dimensional composition space [23]. Rao et al. further emphasized that active learning improves alloy development efficiency by iteratively selecting the most informative new data and balancing exploration and exploitation [24]. At a broader level, closed-loop materials platforms have demonstrated the potential of continuously updated models: Kusne et al. reported an approximately tenfold reduction in the number of experiments required for phase mapping and property optimization, while Liang et al. coupled active learning with real-time CALPHAD updating and reduced the experiments needed for phase-diagram mapping by approximately sixfold [25,26]. Nevertheless, these adaptive frameworks have mainly focused on alloy discovery, phase mapping, and composition optimization rather than on the continuous refinement of heat treatment-dependent mechanical-property prediction for DED low-alloy steels.

Accordingly, an important gap remains at the intersection of DED low-alloy steel processing, post-deposition heat treatment, CALPHAD-assisted data generation, transfer learning, and adaptive model updating. Previous studies have separately shown that DED steels exhibit strong composition- and thermal-history-dependent variations in properties, and that heat treatment is essential for tailoring the final microstructure and mechanical response. Also, CALPHAD and transfer learning have been shown to improve data-driven materials prediction when experimental data are limited. However, a robust framework that integrates these capabilities to predict heat treatment-dependent YS, UTS,

HV, and solidification phase fractions in DED-processed low-alloy steels has not yet been fully established.

Therefore, this study develops a transfer learning and continuous optimization framework for predicting the mechanical properties and solidification response of DED-processed low-alloy steels. A CALPHAD-based dataset was generated for 125 low-alloy steel compositions, including solidification outputs and heat treatment outputs over tempering temperatures of 200–650 °C and holding times of 1–3 h. A multilayer perceptron (MLP) model was first trained as the baseline surrogate, then adapted through transfer learning and progressively refined through continuous optimization as additional compositions were added. For continuous optimization, Random, Greedy, and Bayesian-UCB acquisition strategies were compared to evaluate data-efficient candidate selection. The framework predicts YS, UTS, HV, and as-solidified martensite, ferrite, and austenite fractions using R^2 (coefficient of determination), MAE (mean absolute error), and RMSE (Root mean square error) as the main evaluation metrics. By combining simulation-assisted data generation, transferable learning, and uncertainty-guided optimization, the proposed approach provides an adaptive pathway for rapid alloy screening, heat treatment selection, and future development of low-alloy DED steel.

2. Data and Features

2.1. Data Generation and Source

The datasets used in this study were generated using the CALPHAD-based simulation software JMatPro (Version 16). The low-alloy steel composition set was first collected and selected from the existing composition databases and literature sources, including publicly available steel composition datasets, the NIMS/MITS materials database, published low-alloy steel studies, and standard steel handbooks [27–33]. These sources were used to define realistic composition ranges for low-alloy steels. After cleaning and filtering the composition data, each selected composition was simulated in JMatPro to generate two complementary outputs: solidification and heat treatment outputs. The solidification output describes the as-solidified state and provides the room-temperature phase fractions (martensite, ferrite, and austenite) together with the corresponding mechanical properties (yield strength, ultimate tensile strength, and hardness). The heat treatment output describes the tempered state and provides the heat treatment-dependent yield strength, ultimate tensile strength, and hardness.

The 125 compositions used in this study fall under the class of low-alloy steels (less than 5% alloying elements), with carbon content varying from 0.09% to 0.25% (low-carbon steels) and 0.25% to 0.44% (medium-carbon steels). The compositions were limited to low-alloy steels for investigating the capability of the proposed framework without the additional complexities of other steel classes. All simulations were performed under a fixed baseline processing condition representative of laser-based directed energy deposition (DED): a grain size of 40 μm [34], a room temperature of 25 °C, and a cooling rate of 1000 °C/s [3]. For the heat treatment simulations, each composition was assigned an automatization temperature of 900 °C and a corresponding major phase fraction from solidification output as initial microstructure (martensite) [35,36]. Tempering simulations of the composition were then carried out according to a defined temperature–time schedule. Because the simulations enforce thermodynamic and metallurgical consistency, they provide dense and physically coherent coverage of the composition and tempering space that would be prohibitively expensive to obtain experimentally. High-fidelity experimental measurements are planned for the next study to validate and incorporate into the target and continuous optimization stages. The prior austenite grain-size value was selected as a literature-based upper-bound value reported by Wang et al. [34], who reported that the

EBSID grain-size distribution of L-DED X70 steel included grains reaching up to approximately 40 μm . The cooling rate was selected based on reported laser DED cooling rates on the order of 10^3 K/s [3]. Martensite was selected as the initial heat treatment microstructure because, at the selected high cooling rate of 1000 $^{\circ}\text{C}/\text{s}$, the solidification simulation predicted martensite as the dominant room-temperature phase for most compositions [35].

2.2. Dataset Composition, Structure, and Pre-Processing

In the heat treatment dataset, every composition was evaluated over a regular tempering grid consisting of 19 tempering temperatures from 200 to 650 $^{\circ}\text{C}$ in 25 $^{\circ}\text{C}$ increments and 3 holding times of 1, 2, and 3 h. Therefore, each composition generated 57 heat treatment conditions. The fully cleaned heat treatment dataset comprises 125 compositions and 57 conditions, for a total of 7125 records. The solidification dataset contains one record per composition, totaling 125 records. The descriptive statistics for the common chemical composition descriptors, solidification input descriptors, heat treatment input descriptors, solidification output properties, and heat treatment output properties are summarized together in Table 1. Grain size, room temperature, cooling rate, austenitization temperature, and initial microstructure were fixed because a preliminary sensitivity analysis showed only limited changes in the outputs investigated within the generic DED-based parameter ranges. Zero-variance descriptors are excluded from model training but retained to define the simulation conditions, ensure reproducibility, and establish the applicability range of the predictions.

Fixing these parameters reduced confounding effects and enabled investigation of the influence of alloy composition and tempering conditions consistently across all 125 compositions. The selected values were not arbitrary. The grain size of 40 μm and austenitization temperature of 900 $^{\circ}\text{C}$ were chosen based on reported laser-DED conditions. The tempering heat treatment module in JMatPro only accepts pre-heat treatment microstructure information in terms of the predominant phase (martensite, bainite, pearlite). Based on initial DED cooling-rate sensitivity trials, all compositions produced martensite/ferrite as the major phase at cooling rates of 1000 $^{\circ}\text{C}/\text{s}$ or higher. Hence, only 1000 $^{\circ}\text{C}/\text{s}$ was utilized in the ML model, as the higher cooling rates did not induce a change in the major phase. Cooling rates below 1000 $^{\circ}\text{C}/\text{s}$ do not represent DED processing except for the in situ tempering of the initially deposited layers. Thus, the current framework is limited to properties due to DED processing rather than other conventional manufacturing routes.

The modeling pipeline does not use the full dataset in a single step; instead, it draws nested subsets of increasing composition coverage at successive stages, from an initial source set, through a transfer-learning target set, to a sequence of continuous optimization stages that progressively enlarge the composition pool. The number of compositions and records available at each stage is listed in Table 2; the rationale for each stage is described in Section 3.

The raw simulation dataset exports were first cleaned by removing decorative header rows, empty rows, and empty columns. The correct header was identified from the row containing the elemental symbols. All input descriptors and target properties were converted to numerical format, and malformed numerical values, such as double-decimal entries, were corrected. Records containing missing input or output values were removed, and physically invalid values were excluded before model training.

Table 1. Statistics of the combined heat treatment ($n = 7125$) and solidification ($n = 125$) datasets after data cleaning.

	Variables	Description	Min.	Max.	Mean	Median	SD
Chemical Composition	C	Chemical Composition (wt.%)	0.09	0.44	0.227	0.18	0.1064
	Si		0.18	0.52	0.2969	0.28	0.0829
	Mn		0.145	1.51	0.938	0.79	0.3932
	P		0.006	0.038	0.0185	0.016	0.0086
	S		0.003	0.041	0.0153	0.012	0.0095
	Ni		0	1.08	0.3108	0.13	0.3329
	Cr		0	1.31	0.4352	0.43	0.3936
	Mo		0.005	1.35	0.3783	0.21	0.3566
	Cu		0	0.25	0.0562	0.04	0.0581
	V		0	0.3	0.044	0	0.0866
	Al		0	0.05	0.0091	0.005	0.0124
	N		0	0.015	0.0057	0.0066	0.0039
Solidification Input Descriptors	Grain Size (μm)	Grain Size	40.00	40.00	40.00	40.00	0.00
	Temperature ($^{\circ}\text{C}$)	Room Temperature	25.00	25.00	25.00	25.00	0.00
	Cooling Rate ($^{\circ}\text{C/s}$)	DED Cooling Rate	1000.00	1000.00	1000.00	1000.00	0.00
Heat Treatment Input Descriptors	Austenitization Temperature ($^{\circ}\text{C}$)	Heat Treatment Input	900.00	900.00	900.00	900.00	0.00
	Initial Microstructure	Categorical Heat Treatment Input	Categorical		Martensite		
	Tempering Temperature ($^{\circ}\text{C}$)	Heat Treatment Input	200.00	650.00	425.00	425.00	136.93
	Time (h)	Tempering Time	1.00	3.00	2.00	2.00	0.82
Solidification Output Properties	Martensite (%)	Phase Fraction (%)	23.59	99.94	76.49	74.69	24.75
	Ferrite (%)		0.00	76.39	23.27	25.17	24.91
	Austenite (%)		0.00	0.90	0.19	0.07	0.24
	YS	Yield Strength (MPa)	518.29	1925.02	1187.86	1035.62	438.84
	UTS	Ultimate Tensile Strength (MPa)	725.98	2109.17	1408.62	1278.94	430.40
	HV	Hardness Value (HRC)	19.13	59.11	43.15	41.69	11.46
Heat Treatment Output Properties	YS	Yield Strength (MPa)	342.10	1529.70	871.89	865.80	291.83
	UTS	Ultimate Tensile Strength (MPa)	521.40	1756.60	1115.03	1121.80	300.44
	HV	Hardness Value (HRC)	−1.09	53.18	33.67	36.23	12.11

Note: The chemical composition and solidification input descriptors are common to both datasets; the heat treatment dataset additionally includes the heat treatment input descriptors. Output properties are reported separately because the solidified and tempered states yield different values. Fe and sample number are not included because Fe is the balance element.

Table 2. Number of compositions and records available to the modeling pipeline at each stage.

Stage	Compositions	Heat Treatment Records	Solidification Records
Source (baseline-MLP)	20 (HT)/60 (solidification)	1140	60
Transfer learning	72	4104	72
Continuous optimization stage 1	85	4845	85
Continuous optimization stage 2	95	5415	95
Continuous optimization stage 3	105	5985	105
Continuous optimization stage 4	115	6555	115
Continuous optimization stage 5	125	7125	125

For the heat treatment dataset, the number of records equals the number of compositions $\times$ 57 tempering conditions; for the solidification dataset, there is one record per composition.

For the solidification dataset, the phase fractions were constrained to remain physically valid. Each phase fraction p_i , where i represents martensite, ferrite, or austenite, was first clipped within the range of 0–100%,

$$\tilde{p}_i = \min(\max(p_i, 0), 100) \quad (1)$$

Then, the clipped phase fractions were normalized so that their total became 100% for each composition,

$$p_i^* = \frac{100\tilde{p}_i}{\sum_{j=1}^3 \tilde{p}_j} \quad (2)$$

Thus, the final phase fractions satisfy,

$$p_{martensite}^* + p_{ferrite}^* + p_{austenite}^* = 100 \quad (3)$$

The phase fractions were normalized to ensure that the total amount of martensite, ferrite, and austenite equaled 100%. The same proportional adjustment was applied to all three phases, preserving their relative amounts and preventing any individual phase from being increased or decreased. This correction addressed minor numerical and data-export inconsistencies and ensured physically consistent targets for model training. In future datasets, any other significant phases will be included explicitly in the total phase balance.

Before model training, all numerical input features and target properties were standardized using the training-set mean and standard deviation,

$$z = \frac{x - \mu_{train}}{\sigma_{train}} \quad (4)$$

The same μ_{train} and σ_{train} were applied to the validation and test sets to prevent data leakage from the held-out data.

2.3. Input Features and Target Properties

Each steel is described by its chemical composition and a set of process descriptors. The compositional input consists of twelve alloying and residual elements: C, Si, Mn, P, S, Ni, Cr, Mo, Cu, V, Al, and N (in wt.%) with Fe treated as the balance. The process descriptors for the solidification output are the prior austenite grain size, the room temperature, and the cooling rate. The heat treatment output adds four further descriptors: the Austenitization temperature, the initial microstructure, the tempering temperature, and the tempering time.

Within the present datasets, several of these process descriptors are held at a single value: the grain size (40 μm), the room temperature (25 $^{\circ}\text{C}$), the cooling rate (1000 $^{\circ}\text{C/s}$),

the Austenitization temperature (900 °C), and the initial microstructure (martensitic). As shown by the zero standard deviation of these variables in Table 1, they do not vary across the dataset and therefore carry no learnable signal; the informative inputs are the chemical composition and, for the heat treatment output, the tempering temperature and tempering time.

The target properties depend on the output type. For the solidification output, the targets are the three room-temperature phase fractions (martensite, ferrite, and austenite in %) together with YS, UTS, and HV. For the heat treatment output, the targets are yield strength, ultimate tensile strength, and hardness. The ranges, means, and dispersions of all inputs and targets are reported in Table 1.

2.4. Composition-Aware Splitting and Leakage Control

Because each heat treatment composition generates 57 correlated records, random row-level splitting was avoided to prevent data leakage. A composition group identifier g_c was created from the rounded chemical composition vector,

$$g_c = \text{ID}(\text{round}([C, Si, Mn, P, S, Ni, Cr, Mo, Cu, V, Al, N], d)) \quad (5)$$

where d is the decimal precision used for rounding. All records with the same composition identifier were assigned to the same data partition. Therefore, for any two records i and j ,

$$g_i = g_j \Rightarrow f_i = f_j \quad (6)$$

where f_i and f_j represent the assigned fold or split. This ensures that all 57 heat treatment records from one composition remain together and do not appear in both training and testing sets.

For grouped cross-validation, the dataset was divided by composition groups rather than by individual rows,

$$G_{train}^{(k)} \cap G_{test}^{(k)} = \emptyset \quad (7)$$

where $G_{train}^{(k)}$ and $G_{test}^{(k)}$ are the composition groups used for training and testing in fold k . For fixed train/validation/test splitting, the same rule was applied,

$$G_{train} \cap G_{val} \cap G_{test} = \emptyset \quad (8)$$

with an approximate split ratio of 70/15/15 at the composition level. This approach prevents the model from seeing the same alloy composition during training and testing, giving a more realistic estimate of generalization to unseen steels.

In the transfer-learning stage, an additional leakage guard was applied by removing any record matching the target test compositions from the source domain,

$$D_{source}^{clean} = D_{source} \setminus D_{target, test} \quad (9)$$

Thus, the pre-trained model had no direct or indirect access to the final target evaluation data. In the continuous optimization stage, a fixed unseen test set was kept constant across all stages,

$$G_{test}^{(1)} = G_{test}^{(2)} = \dots = G_{test}^{(n)} \quad (10)$$

while the training pool increased progressively,

$$G_{train}^{(1)} \subset G_{train}^{(2)} \subset \dots \subset G_{train}^{(n)} \quad (11)$$

This design ensures that performance improvements during continuous optimization come from learning additional composition–property relationships rather than memorizing the test data.

3. Methodology

3.1. Overall Framework

The proposed framework is organized as a four-stage pipeline that moves from a conventional data-driven surrogate to an interpretable, continuously updated predictor. In the first stage, data is collected from the CALPHAD-based simulation, and a subset is made for applying in the framework. In the second stage, a baseline multilayer perceptron (MLP) surrogate is trained from scratch on the source simulation data and benchmarked against a panel of established regression algorithms to establish reference accuracy and justify the choice of model. In the third stage, the baseline MLP is reused through transfer learning, and its learned representation is pre-trained on the source domain and then adapted to a broader target composition space. In the fourth stage, the transferred model is progressively refined through an active-learning continuous optimization loop in which new compositions are added in successive batches and the most informative candidates are identified using Random, Greedy, and Bayesian acquisition strategies. The complete pipeline is summarized in Figure 1.

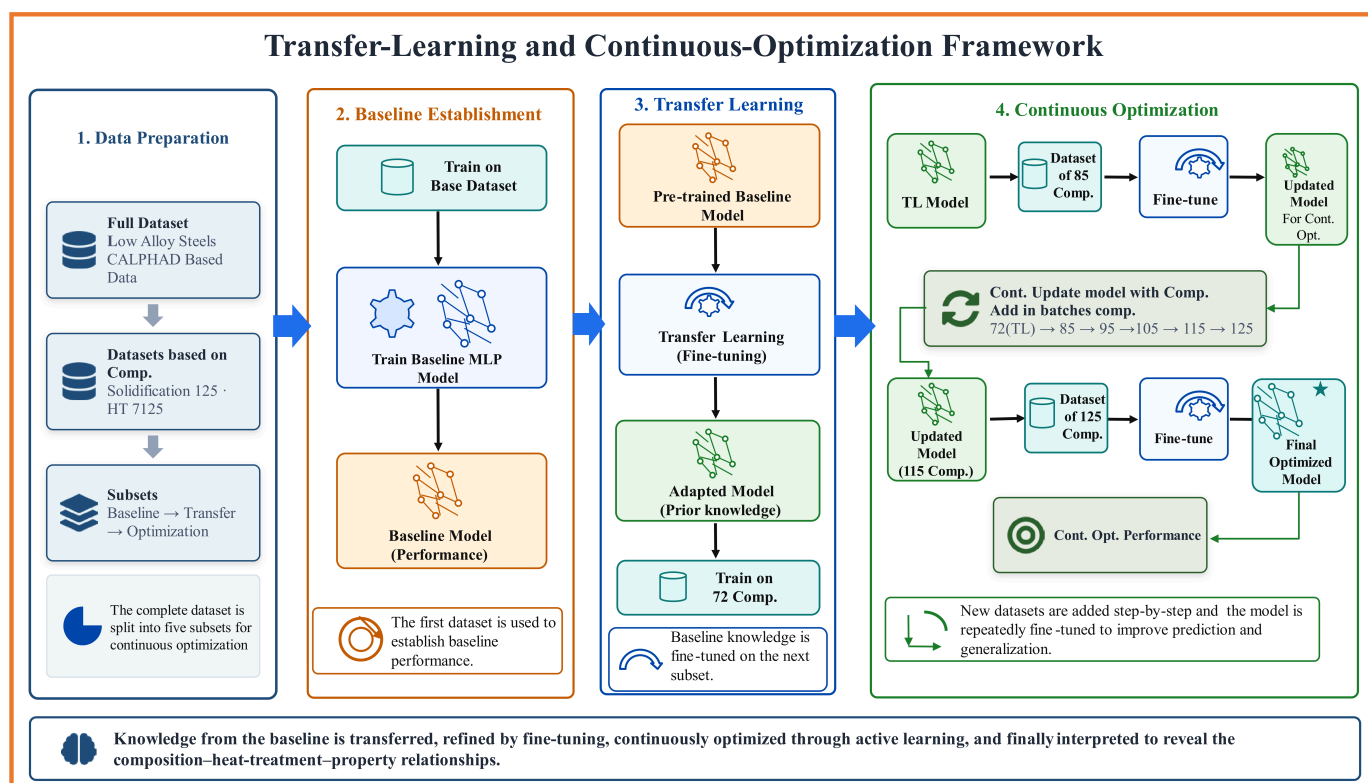


Figure 1. Transfer-learning and continuous optimization framework.

The proposed framework was designed as a four-stage learning pipeline to predict heat treatment-dependent mechanical properties and solidification phase fractions in DED-processed low-alloy steels. The complete dataset can be written as

$$D = \{(x_i, y_i, g_i)\}_{i=1}^N \quad (12)$$

where x_i is the input feature vector, y_i is the target output vector, and g_i is the composition group identifier. The input vector contains chemical composition and process descriptors,

while the target vector contains either mechanical properties or phase fractions, depending on the dataset type.

The complete framework consists of four major stages,

$$D_{source} \rightarrow f_{\theta}^{base} \rightarrow f_{\theta}^{TL} \rightarrow f_{\theta}^{CO} \quad (13)$$

where D_{source} is the baseline source dataset, f_{θ}^{base} is the baseline surrogate model, f_{θ}^{TL} is the transfer-learned model, and f_{θ}^{CO} is the continuously optimized model.

3.2. Baseline Surrogate Model-MLP

A multilayer perceptron was used as the primary baseline surrogate because it can learn nonlinear composition–processing–property relationships and be directly transferred via weight initialization during the transfer-learning stage (Figure 2). The MLP maps the *standardized input vector* x_i to the *predicted target vector* $\hat{y}_i$,

$$\hat{y}_i = f_{\theta}(x_i) \quad (14)$$

where θ represents all trainable network weights and biases. For a network with L hidden layers, the forward propagation is expressed as

$$h_0 = x \quad (15)$$

$$h_l = \sigma(W_l h_{l-1} + b_l), l = 1, 2, \dots, L \quad (16)$$

$$\hat{y} = W_{L+1} h_L + b_{L+1} \quad (17)$$

where W_l and b_l are the weight matrix and bias vector of layer l , and $\sigma(\cdot)$ is the ReLU activation function,

$$\sigma(z) = \max(0, z) \quad (18)$$

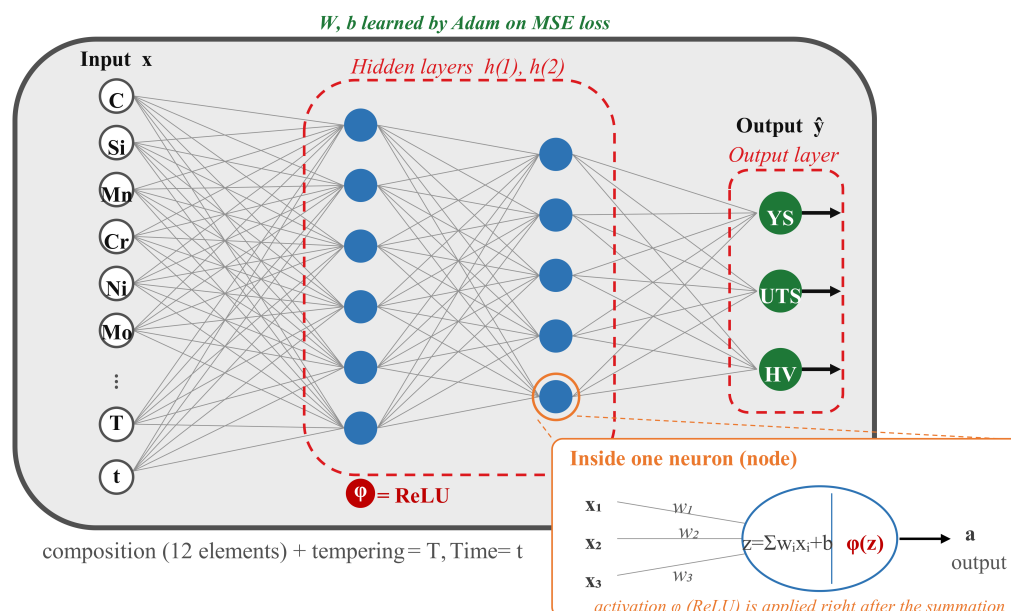


Figure 2. Schematic architecture of the MLP.

The model was trained by minimizing the mean squared error loss,

$$\mathcal{L}_{MSE} = \frac{1}{Nm} \sum_{i=1}^N \sum_{j=1}^m (y_{ij} - \hat{y}_{ij})^2 \quad (19)$$

where N is the number of samples and m is the number of output targets. The model parameters were updated using the Adam optimizer,

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (20)$$

where η is the learning rate, and $\hat{m}_t$ and $\hat{v}_t$ are the bias-corrected first and second moment estimates.

For the heat treatment dataset, grouped cross-validation was used because one composition generates 57 tempering records. Therefore, all records belonging to the same composition were kept in the same fold,

$$G_{train}^{(k)} \cap G_{test}^{(k)} = \emptyset \quad (21)$$

This prevents the same alloy composition from appearing in both training and testing partitions.

The MLP hyperparameters were evaluated using five-fold composition-aware cross-validation, and the configuration with the highest mean cross-validated R^2 across the target properties was selected. The final network consisted of three hidden layers containing 128, 64, and 32 neurons with ReLU activation and a dropout rate of 0.10 after each hidden layer. Training used the AdamW optimizer, mean-squared-error loss, and a weight-decay coefficient of 1×10^{-5} . Full-batch training was employed, and all samples in the training partition were used in each gradient update. The learning rate was 5×10^{-4} for initialization using the 72-composition dataset and 3×10^{-4} during each continuous optimization update. The maximum numbers of epochs were 220 for initialization and 140 for each subsequent optimization stage, with early stopping applied after 35 epochs in validation loss. A random seed of 42 was used for Python (v3.11.5), NumPy (v1.26.4), and PyTorch (v2.2.1).

3.3. Algorithm Benchmarking

To confirm that the MLP was a suitable transferable baseline model, it was compared with several machine learning algorithms, including k-nearest neighbors, support vector regression, and MLP (baseline) regression.

For each algorithm A_q , the cross-validated performance was calculated as

$$Score(A_q) = \frac{1}{Km} \sum_{k=1}^K \sum_{j=1}^m R_{k,j}^2 \quad (22)$$

where K is the number of folds and m is the number of targets. The algorithm with the highest score is useful for comparison; however, the MLP was retained as the main baseline because its learned weights can be transferred and fine-tuned in the next stage. Tree-based models such as k-nearest neighbors and support vector regression are strong predictors, but they do not support direct neural weight transfer in the same way as MLP.

MLP has become a standard choice for this study for CALPHAD-based data instead of modern gradient-boosting algorithms like XGBoost, LightGBM, and CatBoost. XGBoost [37,38] combines strong regularization with flexible modeling of nonlinear feature interactions, LightGBM [39] scales efficiently to large tabular datasets through histogram-based training, and CatBoost [40] handles informative categorical variables with little preprocessing. Transformer architectures adapted to tabular data [41,42] that capture higher-order cross-feature interactions. For the present framework, MLP is not only a convenient choice but also the best for the baseline model. The MLP model selected here is based on the requirement of the complete pipeline rather than a single-stage accuracy model.

The framework requires four capabilities simultaneously, like direct parameter transfer from the source-domain model, property-specific fine-tuning, progressive retraining as the composition extends, and MC-Dropout uncertainty estimation to drive the Bayesian-UCB acquisition step. The MLP is the only candidate that delivers all four within a single architecture. Its hidden layers learn a reusable distributed representation of the composition-based processing property mapping that can be initialized from pretrained weights, fine-tuned incrementally, and calibrated uncertainty through dropout sampling. Boosted gradient algorithm ensembles fail the first three requirements outright. Modern algorithms can only encode knowledge in axis-aligned split structures rather than transferable weights. Also, adapting them to a new composition domain reduces full retraining and offers no dropout-based predictive uncertainty for the acquisition step. Transformers support weight transfer in principle, but their advantages emerge at data scales far beyond the present study with 7125 data points from 125 compositions. All inputs are continuous, and added architectural complexity and tuning could raise the risk of overfitting without any gain over the MLP algorithm. This advantage is confirmed empirically, as kNN and SVR match or exceed the MLP on isolated baseline targets. Only MLP is a transferable representation that enabled the gains of the pipeline for this study and most suitable model. A detailed comparison with XGBoost, LightGBM, CatBoost, and transformer models on larger experimental datasets will be carried out in future work.

3.4. Transfer Learning

Transfer learning is used to adapt knowledge learned from the source dataset to the smaller target dataset. The source model was first trained on the baseline dataset (Figure 3),

$$\theta_s = \arg \min_{\theta} \mathcal{L}(D_{source}; \theta) \quad (23)$$

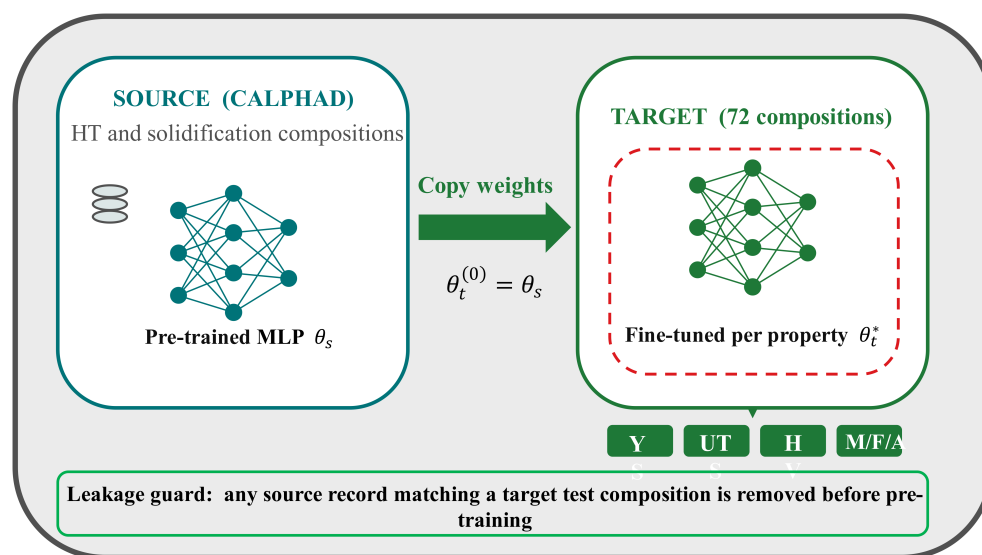


Figure 3. Schematic architecture of transfer learning.

The trained source weights θ_s were then used to initialize the target model,

$$\theta_t^{(0)} = \theta_s \quad (24)$$

The model was then fine-tuned using the target-domain data,

$$\theta_t^* = \arg \min_{\theta_t} \mathcal{L}(D_{target}; \theta_t) \quad (25)$$

In this study, separate transfer-learning models were trained for each target property so that each property could have its own optimized training behavior,

$$f_{\theta}^{YS}(x), f_{\theta}^{UTS}(x), f_{\theta}^{HV}(x) \quad (26)$$

and, for the solidification dataset,

$$f_{\theta}^M(x), f_{\theta}^F(x), f_{\theta}^A(x) \quad (27)$$

where M , F , and A represent martensite, ferrite, and austenite fractions.

A leakage guard was applied before pre-training by removing any source-domain record that matched the target test data,

$$D_{source}^{clean} = D_{source} \setminus D_{target,test} \quad (28)$$

This ensured that the source model did not indirectly learn from the target evaluation set. During fine-tuning, early stopping was used based on validation performance. For mechanical-property models, training was stopped when the validation accuracy reached a predefined cap or when no further improvement occurred,

$$R_{val}^2 \geq R_{cap}^2 \quad (29)$$

This avoided excessive overfitting on the small target dataset.

In this study, transfer learning is implemented as within-domain knowledge transfer rather than simulation-to-experiment adaptation. The source and target datasets represent different composition subsets and associated property distributions. The learned parameters were used to initialize property-specific models, which were fine-tuned using the expanded 72-composition dataset. Thus, the approach transfers learned composition–property representations across subsets with increasing composition coverage within a single simulation domain rather than performing simulation-to-experiment, cross-machine, or cross-process adaptation. The subsequent expansion from 85 to 125 compositions represents progressive fine-tuning of the same transferred model as additional CALPHAD-generated compositions were incorporated. Therefore, the methodological scope of the present framework is limited to knowledge transfer across related CALPHAD-generated composition subsets.

3.5. Active-Learning Continuous Optimization

After transfer learning, continuous optimization was used to progressively improve the model by adding new compositions in stages. The initial transfer-learning dataset contained 72 compositions, and the composition pool was expanded as follows:

$$72(TL) \rightarrow 85 \rightarrow 95 \rightarrow 105 \rightarrow 115 \rightarrow 125 \quad (30)$$

In the Figure 4, continuous optimization refers to the staged expansion of the dataset and progressive updating of the transferred model. Active learning denotes the model-guided selection of informative candidate compositions. Bayesian optimization is implemented through Bayesian-UCB, which combines predicted performance with MC-Dropout uncertainty to balance exploration and exploitation. Random selection serves as an unguided baseline, while Greedy selection represents pure exploitation.

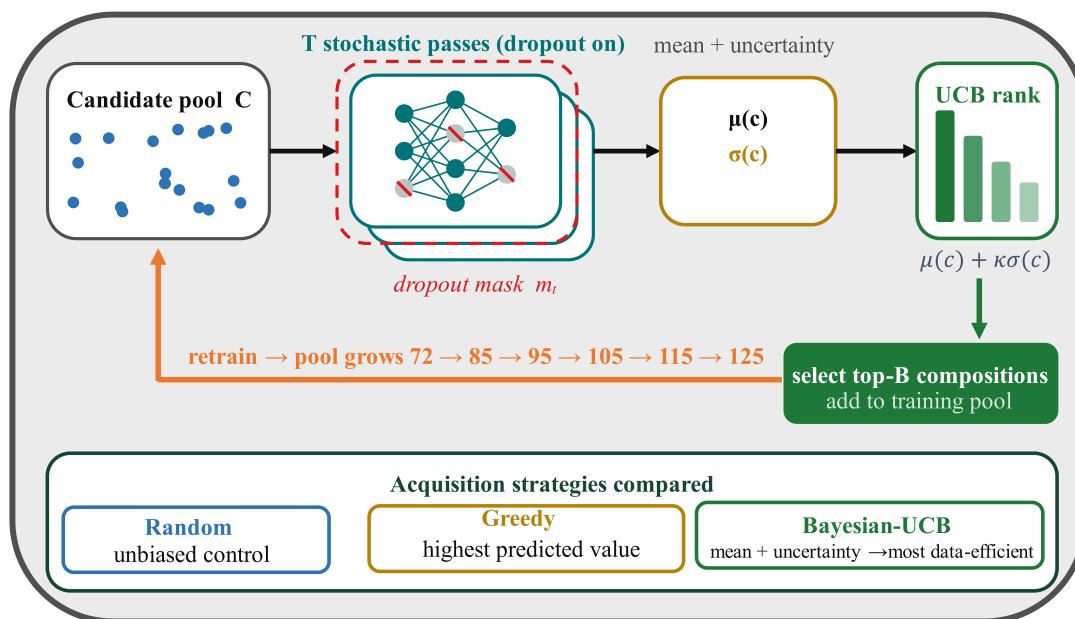


Figure 4. Schematic architecture of continuous optimization.

At each stage s , the training set was updated as follows:

$$D_{train}^{(s+1)} = D_{train}^{(s)} \cup B_s \quad (31)$$

where B_s is the batch of newly selected compositions. A fixed held-out test set was kept unchanged across all stages:

$$D_{test}^{(1)} = D_{test}^{(2)} = \dots = D_{test}^{(s)} \quad (32)$$

This allows stage-to-stage performance improvement to be evaluated fairly.

Three acquisition strategies were compared, and those are as follows:

Random selection,

$$B_s \sim U(C_s) \quad (33)$$

where C_s is the remaining candidate composition pool.

Greedy selection,

The greedy method selected compositions with the highest predicted property values,

$$a_{greedy}(c) = \frac{1}{m} \sum_{j=1}^m \hat{y}_j(c) \quad (34)$$

$$B_s = \arg \max_{c \in C_s} a_{greedy}(c) \quad (35)$$

Bayesian-UCB (upper confidence bound) selection with MC-Dropout uncertainty.

For Bayesian uncertainty-guided selection, MC-Dropout was used to estimate prediction uncertainty. The model was evaluated T times with dropout activated,

$$\hat{y}^{(t)}(c) = f_{\theta}^{(t)}(c), t = 1, 2, \dots, T \quad (36)$$

The predictive mean and uncertainty were calculated as follows:

$$\mu(c) = \frac{1}{T} \sum_{t=1}^T \hat{y}^{(t)}(c) \quad (37)$$

$$\sigma(c) = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (\hat{y}^{(t)}(c) - \mu(c))^2} \quad (38)$$

The upper confidence bound acquisition function was then defined as follows:

$$a_{UCB}(c) = \mu(c) + \kappa\sigma(c) \quad (39)$$

where κ controls the balance between exploitation and exploration. A larger κ gives more weight to uncertain compositions. The compositions with the highest UCB score were selected,

$$B_s = \arg \max_{c \in C_s} a_{UCB}(c) \quad (40)$$

This strategy allows the model to select compositions that are either predicted to perform well or are located in uncertain regions of the composition space. The three acquisition strategies selected represent distinct candidate-selection principles within the continuous optimization framework. Random selection was used as an unguided baseline; Greedy selection prioritized candidates with the highest predicted responses; and Bayesian-UCB balanced exploitation and exploration by combining the predicted mean with MC-Dropout uncertainty. This comparison enabled evaluation of the effect of uncertainty-guided sampling against both Random and prediction-based selection.

Bayesian optimization is chosen over existing models like expected improvement (EI), thompson sampling (TS), and entropy search (ES). Because EI is perfect for optimizing a single scalar objective, where improvement is measured against the best value found so far, but in the present framework candidates are scored across nine output properties simultaneously. The goal of each acquisition step is to improve the accuracy over the whole composition, not the single one. TS selects candidates by using sample functions from the predictive distribution with the MC-Dropout function, which is noisy, and the resulting stochastic selections would make the controlled and stage-by-stage comparison of acquisition strategies difficult to reproduce. ES selects candidates that help locate the single best solution, which is computationally expensive for large candidate pools, and it again assumes a single-optimum objective. But Bayesian-UCB avoids these limitations and uses the predictive mean and MC-Dropout uncertainty by Equation (39). It produces deterministic and reproducible rankings, exposes the exploration–exploitation trade-off through a single interpretable parameter κ . UCB is therefore the acquisition function most compatible with the multi-output, accuracy-driven continuous optimization loop, while Random and Greedy selection provided the unguided reference points that were needed for a controlled comparison.

3.6. Evaluation Metrics

Model performance was evaluated using the coefficient of determination, mean absolute error, root mean square error, and mean absolute percentage error. The coefficient of determination was calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (41)$$

The mean absolute error was calculated as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (42)$$

The root mean square error was calculated as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (43)$$

The mean absolute percentage error was calculated as follows:

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (44)$$

Because phase fractions may contain values close to zero, MAPE was reported only for mechanical properties. For cross-validation and repeated runs, results were reported as mean and standard deviation,

$$Metric = \mu \pm \sigma \quad (45)$$

where

$$\mu = \frac{1}{K} \sum_{k=1}^K Metric_k \quad (46)$$

$$\sigma = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (Metric_k - \mu)^2} \quad (47)$$

This reporting format provides both the average model accuracy and the variability across folds or optimization runs.

4. Results and Discussion

The results and discussion section follows the pipeline in order. Section 4.1 establishes the baseline surrogate and the algorithm comparison that justifies the choice of model, Section 4.2 reports the accuracy gained by transfer learning from the CALPHAD-generated source domain, Section 4.3 examines the further gains as the composition pool is expanded through staged continuous optimization, Section 4.4 compares the three acquisition strategies that drive that expansion, Section 4.5 summarizes the overall progression of accuracy across the full pipeline, Section 4.6 presents the final prediction system and a single-alloy case study.

4.1. Baseline Performance and Algorithm Comparison

Before applying transfer learning and continuous optimization, a baseline model was evaluated to establish the framework's initial predictive capability. This baseline stage provides the reference accuracy for both the solidification and heat treatment datasets. The predicted values were compared with the simulated target values using parity plots and standard regression metrics. These results help identify which properties are already well captured by the MLP model and which outputs require further improvement through transfer learning and optimization, as shown in Figures 5 and 6.

To test whether a different learner would serve better, the MLP was benchmarked against a distance-based model (k-nearest neighbors) and a kernel-based model (support vector regression) under identical composition-aware cross-validation. Table 3 reports this comparison, and Figures 5 and 6 show the out-of-fold parity plots for the baseline MLP on the solidification and heat treatment datasets, respectively.

Baseline MLP — Solidification dataset

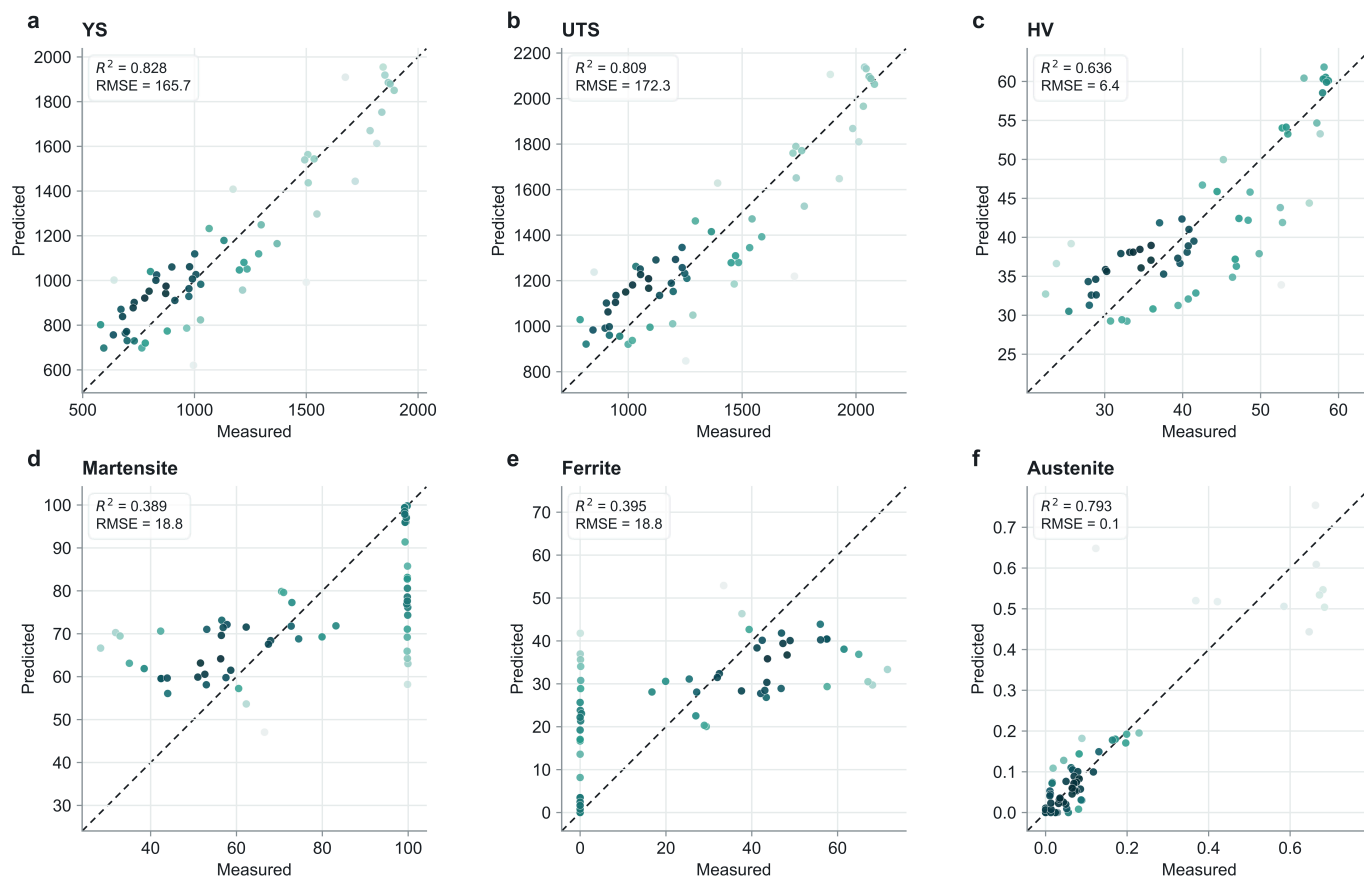


Figure 5. Out-of-fold parity plots of the baseline MLP on the solidification dataset for yield strength, ultimate tensile strength, hardness, and martensite, ferrite, and austenite fractions.

Baseline MLP — Heat-Treatment dataset

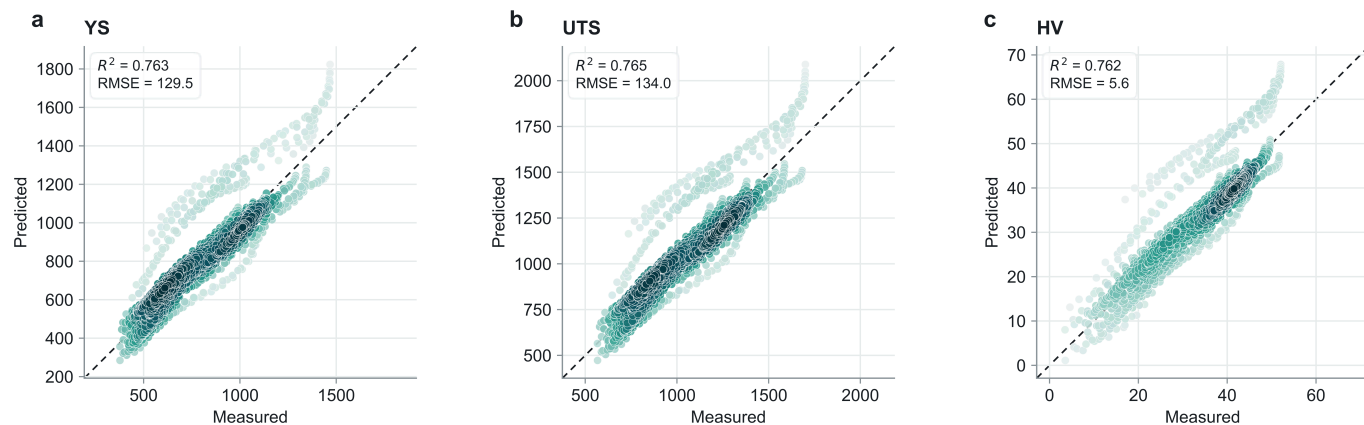


Figure 6. Out-of-fold parity plots of the baseline MLP on the heat treatment dataset for yield strength, ultimate tensile strength, and hardness.

The benchmark in Table 3 shows that no single algorithm dominates every target, but the differences are informative. On the solidification mechanical properties, SVR gave the highest single-target scores (R^2 of 0.90 for yield strength and 0.88 for ultimate tensile strength), with kNN and the MLP close behind, whereas on the heat treatment outputs, kNN scored highest, reaching R^2 of about 0.86 for all three properties. These local similarity and kernel methods perform well when test compositions resemble the training compositions, but their advantage is fragile. kNN interpolates from nearby points rather than learning a global composition–property function, and SVR behaves unevenly across

outputs. More importantly, neither model exposes a reusable internal representation that can be carried into a new alloy domain, which is exactly what the later transfer-learning stage requires. For this reason, the MLP was retained as the baseline surrogate. Its hidden layers learn a distributed representation of the composition–processing–property mapping that can be saved and used to initialize a model for a new target domain, something the tree- and distance-based learners cannot offer. The MLP was therefore selected not on baseline accuracy alone but on its fitness for the full pipeline of baseline training, weight-transfer fine-tuning, and uncertainty-guided active learning.

Table 3. Algorithm comparison for mechanical-property prediction using the baseline dataset.

Solidification Output Algorithm Comparison									
Algorithm	YS (MPa)			UTS (MPa)			HV (HRC)		
	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
SVR(RBF)	0.90	90.01	120.38	0.88	95.12	128.77	0.78	3.18	4.68
kNN	0.81	130.40	158.19	0.79	134.31	163.36	0.73	4.22	5.27
MLP	0.82	132.40	165.70	0.81	138.66	172.30	0.63	5.08	6.40
Algorithm	Martensite (%)			Ferrite (%)			Austenite (%)		
	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
SVR(RBF)	0.35	11.96	18.11	0.35	11.96	18.11	0.87	0.04	0.06
kNN	0.48	12.61	16.45	0.48	12.59	16.46	0.71	0.04	0.08
MLP	0.39	16.04	18.80	0.39	15.41	18.80	0.79	0.06	0.10
Heat Treatment Output Algorithm Comparison									
Algorithm	YS (MPa)			UTS (MPa)			HV (HRC)		
	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
SVR(RBF)	0.75	91.62	117.12	0.77	92.36	116.81	0.75	3.93	4.83
kNN	0.86	70.43	91.87	0.86	73.24	94.91	0.86	3.02	3.98
MLP	0.76	94.97	129.50	0.76	99.01	134	0.76	4.03	5.60

The baseline parity plots in Figures 5 and 6 confirm that the MLP captured the dominant trends. For the heat treatment outputs, the points cluster along the ideal diagonal, with R^2 values of 0.76 for yield strength, ultimate tensile strength, and hardness (RMSEs of 130 MPa, 134 MPa, and 5.6 HRC). The solidification mechanical properties were predicted slightly better (R^2 of 0.83 for yield strength and 0.81 for ultimate tensile strength), but the solidification phase fractions were noticeably weaker, with R^2 of only 0.39 for martensite and 0.37 for ferrite; the scatter at the 0% and 100% bounds in Figure 5 reflects the sharp, near-discontinuous way these fractions switch with hardenability. These deficits in the under-represented, highly nonlinear regions of the design space are precisely what motivated the transfer learning and staged optimization that follow.

4.2. Transfer-Learning Performance and Gains over Baseline

Transfer learning was introduced to test whether representations learned on the source domain could improve accuracy on the smaller target domain. The pre-trained MLP weights were used to initialize a separate fine-tuning model for each target property, which was then adapted to the target data. A leakage guard removed any source record that matched a target test composition before pre-training, so the reported accuracy reflects genuine generalization. Accuracy was estimated as a cross-validated mean over eight

target splits because the target sets are small and single-split estimates are noisy. Table 4 summarizes this performance, and Figures 7 and 8 show the corresponding parity plots.

Table 4. Cross-validated transfer-learning performance for the solidification and heat treatment prediction pipelines.

Solidification Output Prediction Performance			
Target	R2	MAE	RMSE
YS (MPa)	0.906	86.11	115
UTS (MPa)	0.89	89.41	123.13
HV (HRC)	0.82	3.02	4.31
Martensite (%)	0.54	10.99	15.99
Ferrite (%)	0.54	10.95	16.05
Austenite (%)	0.90	0.03	0.05

Heat Treatment Output Prediction Performance			
Target	R2	MAE	RMSE
YS (MPa)	0.92	59.46	75.45
UTS (MPa)	0.92	62.73	78.90
HV (HRC)	0.92	2.486	3.155

Note: YS and UTS errors are reported in MPa, HV errors are reported in HRC units, and phase-fraction errors are reported in percentage units. Only R2, MAE, and RMSE mean values are included.

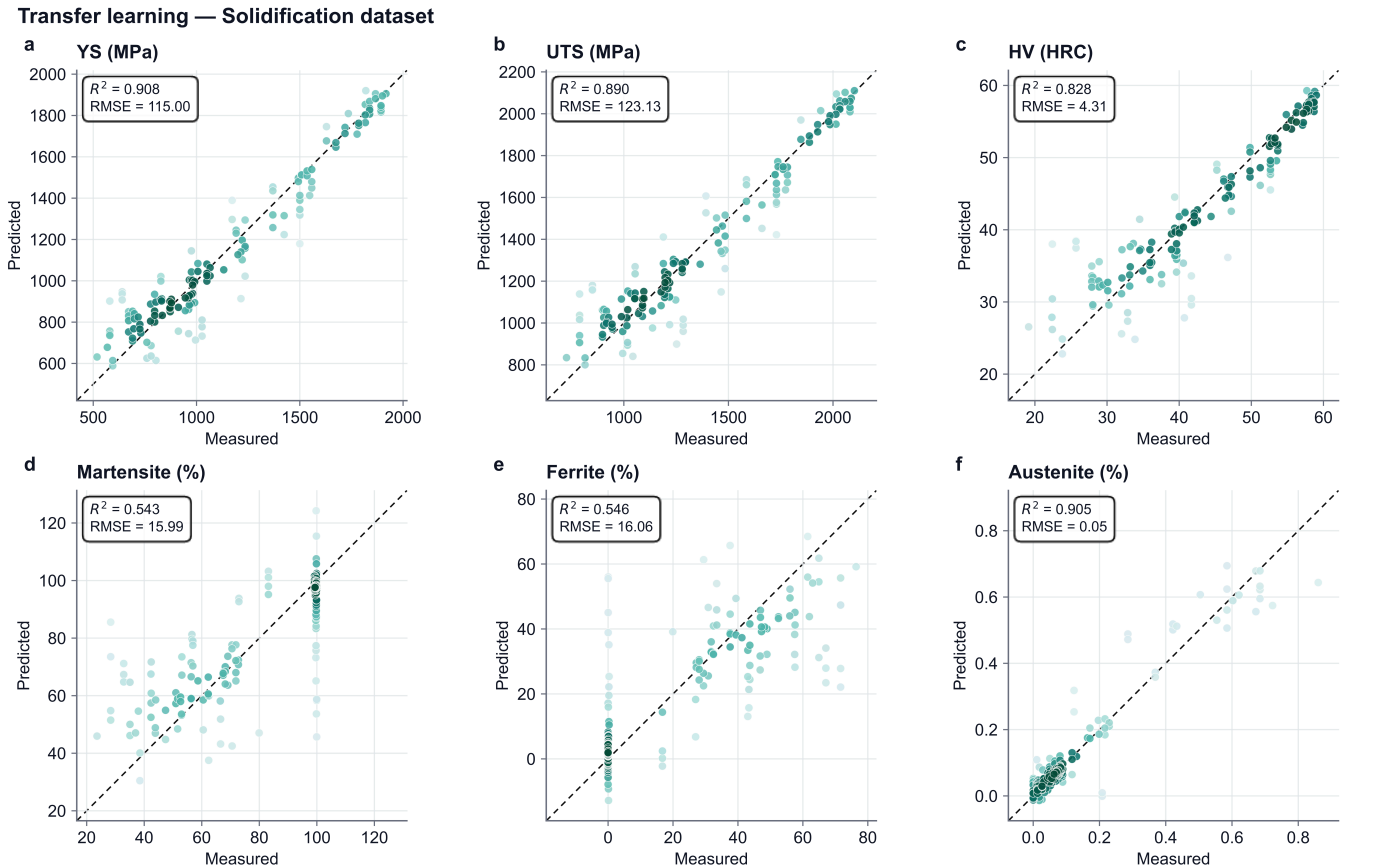


Figure 7. Transfer-learning parity plots on the solidification dataset; metrics are cross-validated means over eight target splits.

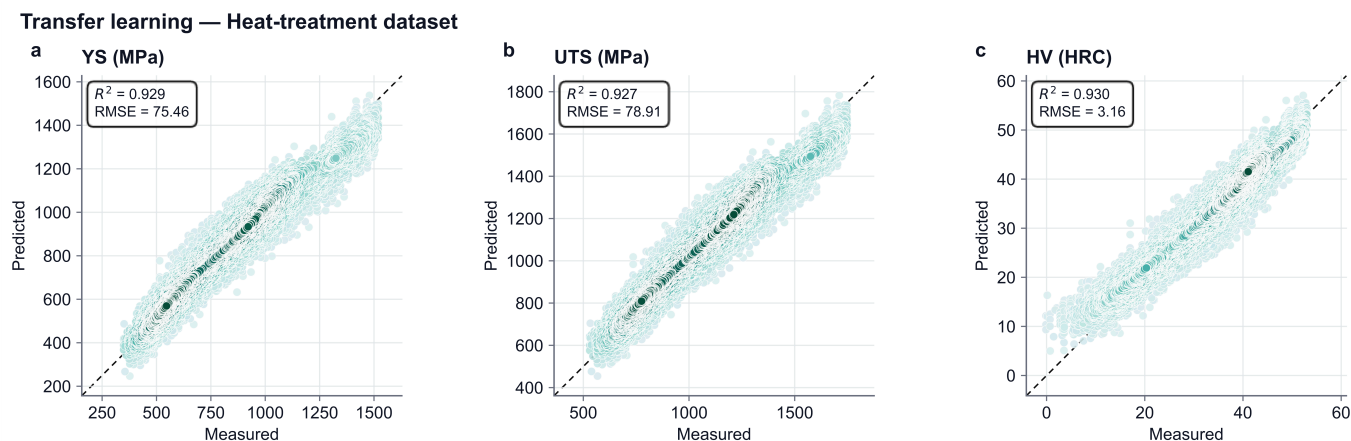


Figure 8. Transfer-learning parity plots on the heat treatment dataset; metrics are cross-validated means over eight target splits.

Relative to a model trained directly on the limited target data, fine-tuning from the source representation produced more stable predictions and lower errors across nearly every output, and the benefit was largest where the target data were scarcest. The mechanism is straightforward: the source model had already encoded the broad composition–property and tempering–property relationships, so fine-tuning only needed to correct for the source-to-target shift rather than relearn the mapping from Random initialization.

For the heat treatment, the mechanical properties were pronounced. Yield strength, ultimate tensile strength, and hardness reached R^2 values of 0.929, 0.927, and 0.930, with RMSEs of 75 MPa, 79 MPa, and 3.2 HRC (Figure 7), a clear advance over the baseline R^2 of about 0.76. On the solidification side, yield strength and ultimate tensile strength reached R^2 values of 0.908 and 0.890, and hardness reached 0.828 (Figure 8). The phase fractions improved in trend but remained the hardest targets: martensite and ferrite held at R^2 near 0.54, while austenite, whose values are small and smoothly distributed, reached 0.905. The residual difficulty with martensite and ferrite is consistent with their abrupt, hardenability-driven transitions and with the saturated 0% and 100% values that compress the learnable signal.

The consistent improvement confirms that the source model genuinely transferred knowledge. This matters for alloy development because every additional simulation or experiment is costly; transfer learning enables the framework to achieve reliable accuracy with far fewer target-domain points than training from scratch would require, directly reducing the data-generation burden that normally limits data-driven property prediction in DED.

4.3. Data-Efficiency Trend Across Optimization Stages

Having shown that transfer learning improves accuracy at fixed data, the next question is how the model behaves as data accumulates. Continuous optimization was evaluated by expanding the composition pool in stages: 85, 95, 105, and finally 125 compositions while holding a fixed, never-seen test set constant across all stages, so that improvements reflect new learning rather than test leakage. Table 5 reports the stage-by-stage metrics, and Figures 9 and 10 show the parity plots for the final 125-composition model. In Figures 9 and 10, point color indicates local data-point density estimated using a Gaussian kernel (warm = high density, cool = low density); the dashed line is the 1:1 parity reference.

Accuracy climbed steeply over the first expansion and then flattened. For the heat treatment outputs, yield strength, ultimate tensile strength, and hardness rose from an R^2 of about 0.995 at 85 compositions to roughly 0.997 from 95 compositions onward, with

the RMSE settling near 16–17 MPa for the strengths and below 0.7 HRC for hardness on the held-out test set. The final 125-composition model reached an R^2 of 0.989, 0.990, and 0.991 (Figure 10). The solidification mechanical properties followed the same pattern at a slightly lower ceiling—an R^2 of 0.97 for yield strength, 0.96 for ultimate tensile strength, and 0.94 for hardness in the final model (Figure 9)—while the phase fractions, though much improved over baseline, plateaued around R^2 of 0.75 for martensite and ferrite and 0.90 for austenite.

Table 5. Optimization-stage performance comparison for solidification-output and heat treatment-output datasets.

Solidification Output Optimization Comparison										
Mechanical-Property Outputs										
Stage	Comps.	YS			UTS			HV		
		R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
Opt. 85	85	0.9838	47.06	60.95	0.9789	47.88	67.95	0.9631	1.78	2.43
Opt. 95	95	0.9841	43.40	60.37	0.9801	47.51	66.05	0.9621	1.81	2.46
Opt. 105	105	0.9829	48.21	62.57	0.9787	47.38	68.42	0.9588	1.82	2.57
Opt. 115	115	0.9821	49.87	64.03	0.9789	45.51	68.06	0.9612	1.74	2.49
Opt. 125	125	0.9817	52.27	64.67	0.9801	48.21	66.09	0.9646	1.79	2.38
Phase-fraction outputs										
Stage	Comps.	Martensite			Ferrite			Austenite		
		R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
Opt. 85	85	0.6639	8.50	15.15	0.6841	7.80	14.80	0.8897	0.06	0.10
Opt. 95	95	0.6594	8.84	15.25	0.6713	8.26	15.09	0.8920	0.06	0.10
Opt. 105	105	0.6550	8.67	15.35	0.6684	8.08	15.16	0.8946	0.06	0.09
Opt. 115	115	0.6491	8.74	15.48	0.6608	8.18	15.33	0.8986	0.06	0.09
Opt. 125	125	0.6604	8.78	15.23	0.6711	8.38	15.10	0.8960	0.06	0.09
Heat Treatment Output Optimization Comparison										
Mechanical-property outputs										
Stage	Comps.	YS			UTS			HV		
		R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
Opt. 85	85	0.9953	17.00	20.96	0.9953	17.57	21.39	0.9955	0.63	0.82
Opt. 95	95	0.9971	13.04	16.62	0.9971	13.27	16.76	0.9971	0.49	0.65
Opt. 105	105	0.9970	13.37	16.86	0.9971	13.61	17.00	0.9969	0.51	0.68
Opt. 115	115	0.9971	13.14	16.48	0.9972	13.25	16.51	0.9973	0.48	0.64
Opt. 125	125	0.9970	13.45	16.88	0.9971	13.67	16.97	0.9969	0.51	0.67

Note: Only optimization stages are included. YS and UTS errors are reported in MPa, HV errors are reported in HRC units, and phase-fraction errors are reported in percentage units.

The shape of this trend is as informative as its endpoint. The largest gains came from the first batches of new compositions, which extended the model into previously sparse regions of the chemical and tempering space once the dominant relationships were captured. Further compositions produced only marginal improvement, and the metrics fluctuated within a narrow band. The most pronounced performance gain occurred during the initial expansion from 72 to 85 compositions, indicating that this batch introduced previously underrepresented information into the transfer-learning dataset. Thereafter, the

performance metrics varied only marginally as the dataset increased from 95 to 125 compositions, suggesting that the dominant composition–property and tempering–property relationships had largely been captured and that subsequent additions provided diminishing marginal benefit. This diminishing return is the central argument for guided rather than indiscriminate data collection: beyond a certain coverage, what a new composition contributes depends far more on where it sits in the design space than on the raw increase in dataset size.

Final optimized model (125 comps, CV) — Solidification

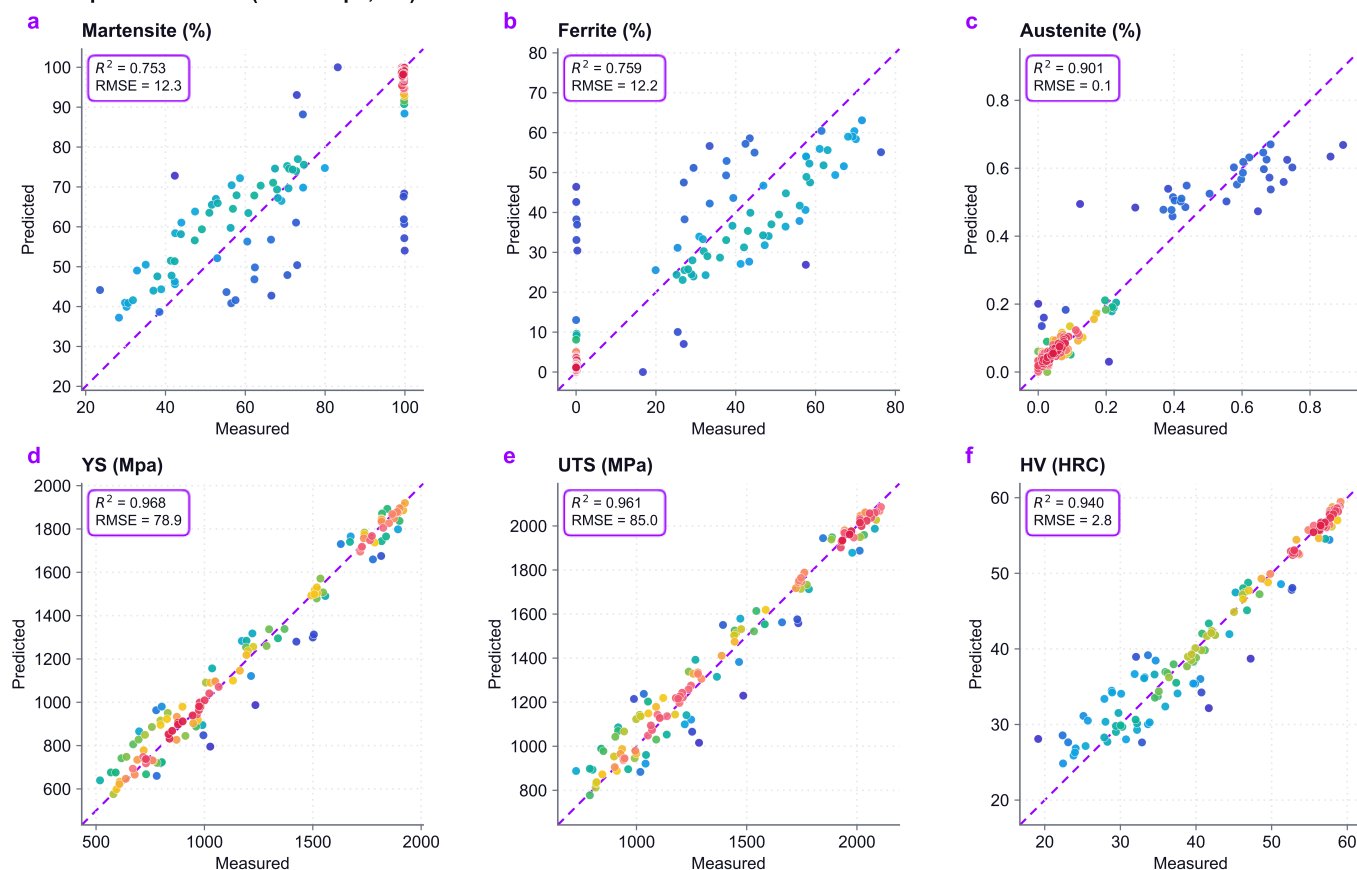


Figure 9. Parity plots of the final optimized 125-composition model on the solidification dataset (held-out test set).

Final optimized model (125 comps, CV) — Heat Treatment

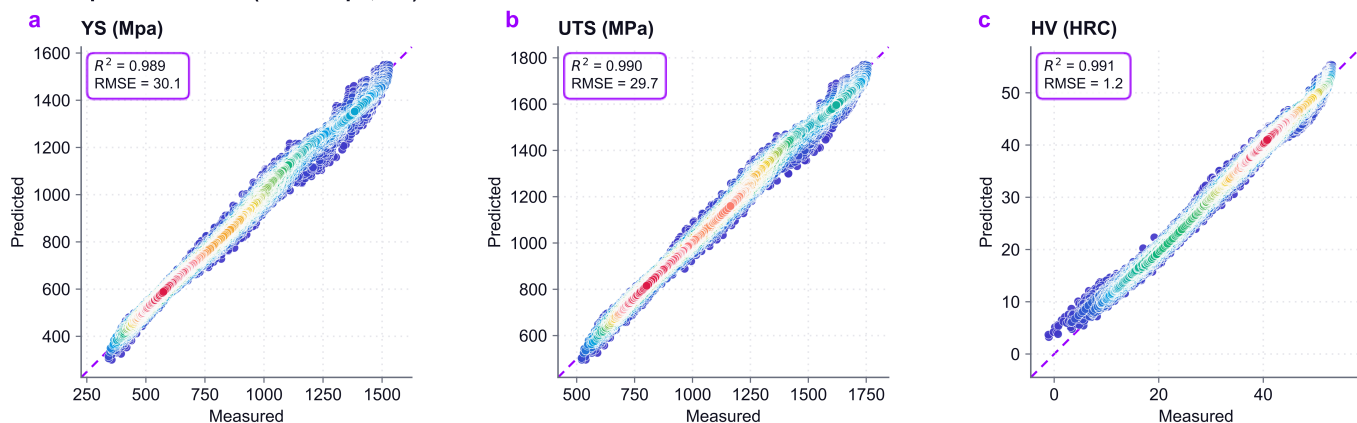


Figure 10. Parity plots of the final optimized 125-composition model on the heat treatment dataset (held-out test set).

The 125-composition model offered the broadest coverage and the most stable behavior across all targets and was therefore adopted as the final optimized model for the prediction system and the case study in Section 4.6.

4.4. Acquisition-Strategy Comparison

Because adding compositions yields diminishing returns, the way candidates are selected becomes the deciding factor for efficiency. Three acquisition strategies were compared as the pool grew from 72 to 125 compositions: Random, which draws candidates without reference to the model and serves as a control; Greedy, which selects the compositions predicted to have the highest target values; and Bayesian-UCB, which ranks candidates by an upper-confidence bound combining the predicted mean with an MC-Dropout estimate of predictive uncertainty. Figures 11 and 12 trace R^2 , MAE, RMSE, and MAPE across the stages for the heat treatment and solidification outputs.

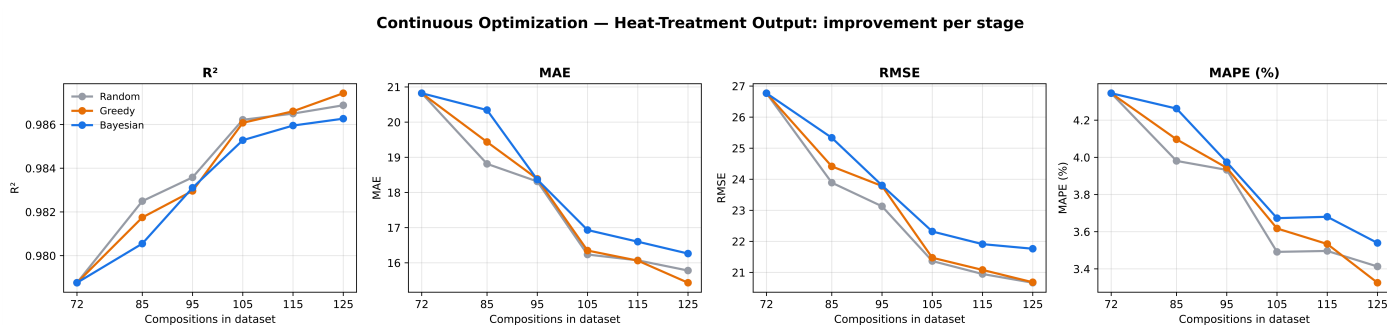


Figure 11. Acquisition-strategy comparison (Random, Greedy, and Bayesian-UCB) for the heat treatment output, showing R^2 , MAE, RMSE, and MAPE as the composition pool grows from 72 to 125.

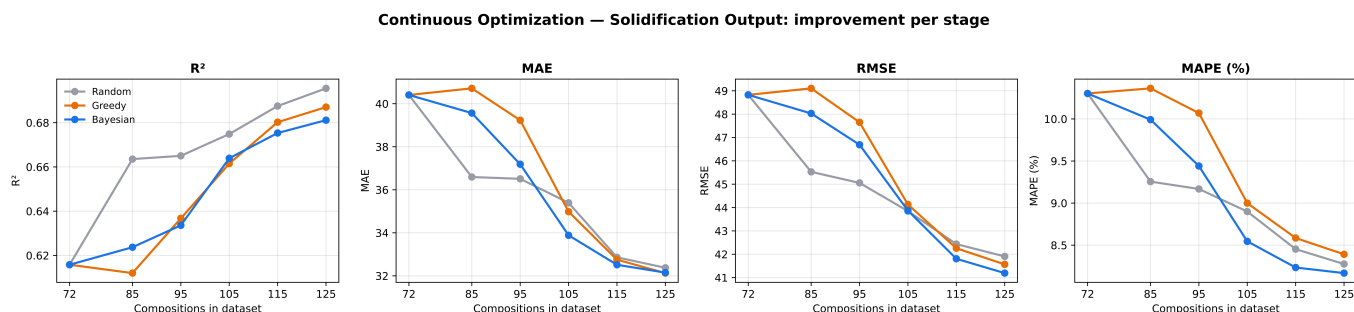


Figure 12. Acquisition-strategy comparison (Random, Greedy, and Bayesian-UCB) for the solidification output across the same staged composition growth.

Random selection improved the model on average but unevenly. Some draws landed in informative regions and helped, while others duplicated information already present and barely moved the metrics; on the heat treatment output, it occasionally edged ahead at individual stages purely by chance. As a strategy, it is simple and unbiased, but it provides no control over what each new experiment contributes to an inefficient basis for guiding costly alloy development.

Greedy selection focuses on compositions predicted to perform well, making it attractive when the goal is to maximize strength or hardness. Its weakness is self-reinforcement: by always sampling where the current model already predicts high values, it tends to neglect uncertain regions that may hold better candidates or the information needed to improve generalization. On the solidification output, this appeared as early stagnation before the later batches recovered.

Bayesian-UCB demonstrated the most balanced behavior. By weighing predicted performance against predictive uncertainty, it sampled candidates that were either promising or poorly understood, so it explored sparse regions of the composition space without abandoning high-performing ones. The MC-Dropout uncertainty made this possible without a separate probabilistic model, and across both outputs, it produced the smoothest, most monotonic reduction in error, the most dependable and data-efficient trajectory of the three (Figures 11 and 12).

Taken together, the comparison shows that active learning improves design efficiency only when selection is guided by both predicted performance and uncertainty. Where each new simulation or experiment is expensive, the normal situation in DED alloy and heat treatment development, an uncertainty-aware acquisition rule extracts more value per added composition than either Greedy exploitation or Random sampling.

4.5. Overall Performance Progression

This subsection brings the stages together to illustrate the pipeline’s cumulative effect. Table 6 lists the average R^2 , RMSE, and percentage error at each stage relative to the transfer-learning model. Figure 13 plots the progression of R^2 , RMSE, and MAPE from the baseline through transfer learning to the optimized stages for both datasets, and Figures 14 and 15 break this progression down property by property.

Table 6. Optimization progression comparison for the heat treatment and solidification output datasets.

Heat Treatment Output Optimization Progression					
Stage	Avg R^2	Avg RMSE	ΔR^2 vs. TL	RMSE ↓% vs. TL	MAPE ↓% vs. TL
Baseline	0.757	80.51	−0.17	−53.33	−18.95
TL-72	0.929	52.51	+0.00	+0.00	+0.00
85	0.995	14.39	+0.07	+72.59	+79.86
95	0.997	11.34	+0.07	+78.40	+84.71
105	0.997	11.51	+0.07	+78.07	+84.08
115	0.997	11.21	+0.07	+78.65	+84.88
125	0.997	11.51	+0.07	+78.08	+84.01
Solidification Output Optimization Progression					
Stage	Avg R^2	Avg RMSE	ΔR^2 vs. TL	RMSE ↓% vs. TL	MAPE ↓% vs. TL
Baseline	0.578	62.02	−0.19	−35.54	-9.458465×10^{15}
TL-72	0.770	45.76	+0.00	+0.00	+0.00
85	0.861	26.90	+0.09	+41.22	−12,051.45
95	0.858	26.55	+0.09	+41.97	−12,260.70
105	0.856	27.36	+0.09	+40.21	−12,379.59
115	0.855	27.58	+0.09	+39.73	−12,540.79
125	0.859	27.26	+0.09	+40.43	−12,311.36

Note: Values are average across target outputs. For solidification, the very large MAPE-based values are caused by near-zero phase-fraction targets; R^2 and RMSE should be emphasized for interpretation.

Continuous-optimization progression: Baseline → Transfer learning → Optimization

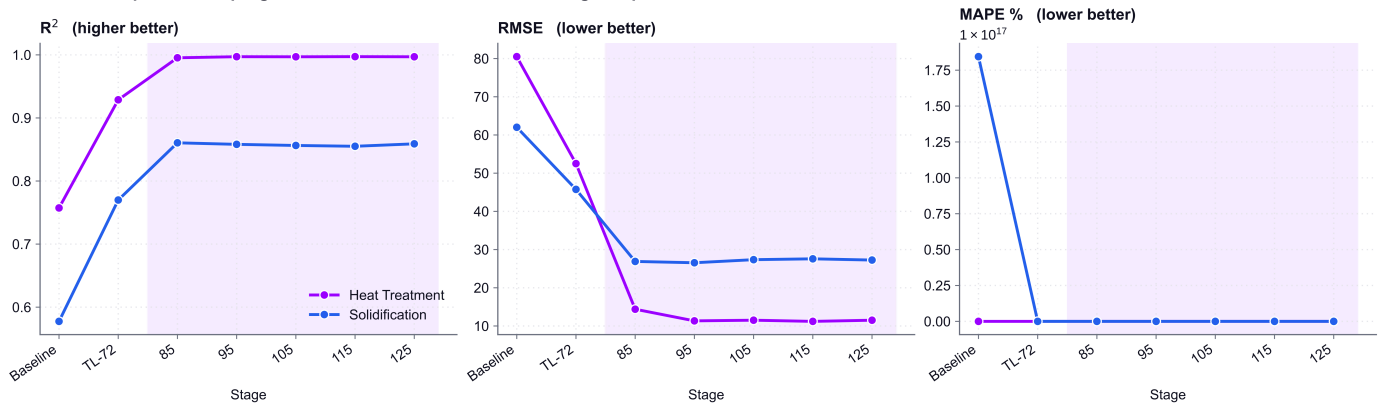


Figure 13. Overall performance progression from the baseline through transfer learning to the optimized stages for the heat treatment and solidification outputs (average R^2 , RMSE, and MAPE).

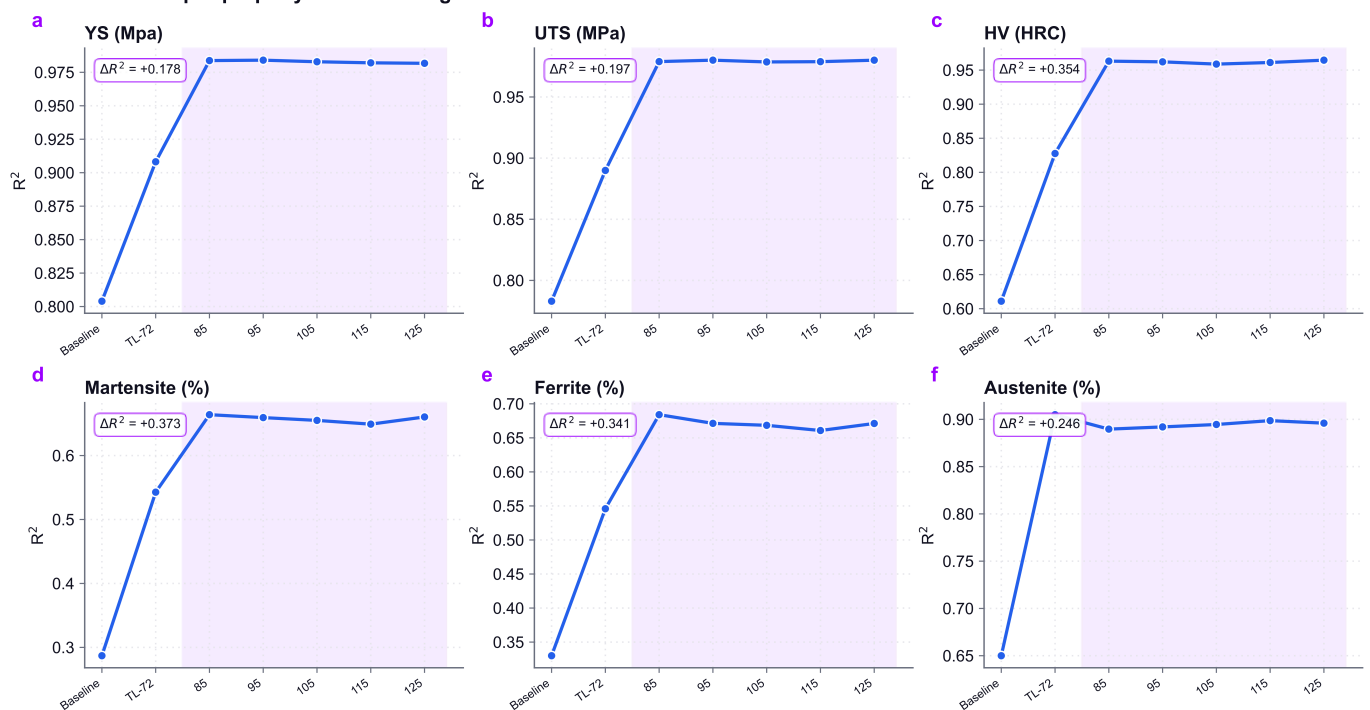
Solidification — per-property R^2 across stages

Figure 14. Per-property R^2 across stages for the solidification output, from baseline through transfer learning to the optimized models.

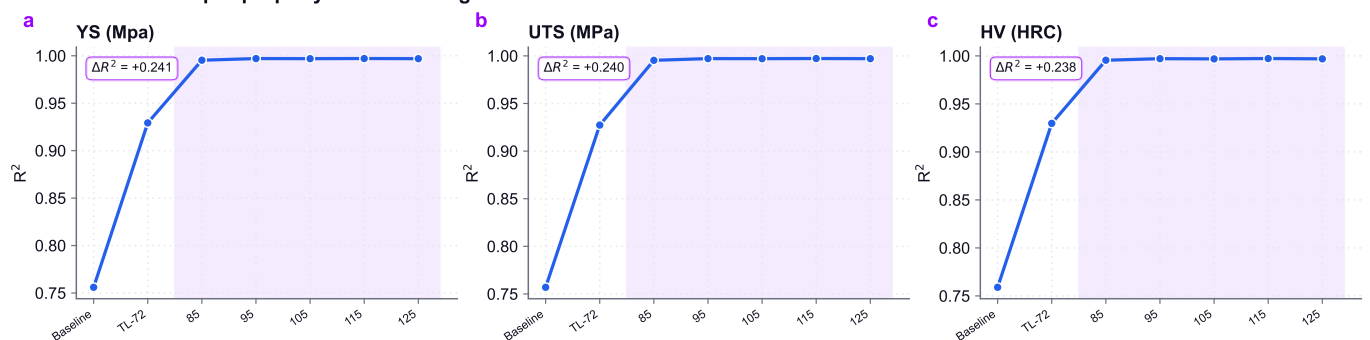
Heat Treatment — per-property R^2 across stages

Figure 15. Per-property R^2 across stages for the heat treatment output, from baseline through transfer learning to the optimized models.

In Figure 13, the overall performance progression from the baseline through transfer learning (TL-72) to the continuously optimized stages (85–125 compositions) for the heat treatment and solidification outputs shows the average R^2 (higher is better), RMSE (lower is better), and MAPE (lower is better). Both datasets improve sharply from the baseline through stage 85, then plateau as R^2 approaches its ceiling and the remaining gains shift into the error metrics. MAPE is averaged over the mechanical properties (YS, UTS, HV) only; phase-fraction MAPE is excluded because near-zero true values for ferrite and austenite make percentage error undefined and produce spuriously large magnitudes. The shaded region marks the continuous optimization stages. The per-property view in Figures 14 and 15 shows that the gains are concentrated between the baseline and the first optimized stage: every property climbs steeply through stage 85 and then plateaus across 95, 105, 115, and 125. This happens because of two compound effects. First, transfer learning already provides a strong priority, and the optimization adds the most informative compositions first. The acquisition strategy targets the highest-value regions early, so each later batch contributes diminishing new information. Second, R^2 is bounded at 1.0, and the model is already near it by stage 85 (≈ 0.97 – 0.98 for solidification, ≈ 1.00 for heat treatment), and R^2 compresses as it approaches this ceiling, becoming insensitive to small absolute improvements. The refinement from the later stages, therefore, does not disappear, and it shifts into the error metrics, with MAE, RMSE, and MAPE continuing to fall even where the R^2 curves are flat. The stable trend through stage 125 confirms the model stays accurate as new compositions are explored, and reaching near-ceiling accuracy with about 85 compositions is itself a data-efficiency result rather than a stalling of the optimization.

The progression is large and consistent. For the heat treatment output, average R^2 rose from 0.757 at baseline to 0.929 after transfer learning and to about 0.997 once optimization began. In contrast, average RMSE fell from roughly 80 to 52 and then to about 11 MPa, a reduction of nearly 78% relative to the transfer-learning model and close to 85% in percentage error. The solidification output improved from R^2 of 0.578 to 0.770 and then to about 0.86, with RMSE falling from about 62 to 46 and then to 27, roughly a 40% reduction beyond transfer learning. The per-property view in Figures 14 and 15 shows where the pipeline did the most work and the hardest baseline targets gained the most, with hardness improving by about $\Delta R^2 = 0.35$ and the solidification phase fractions by ΔR^2 between 0.25 and 0.37, while the already-strong heat treatment properties each gained about 0.24. Most of the climb occurs between the baseline and the first optimized stage, after which the curves are essentially flat, with the same data-efficiency signature as in Section 4.3, now visible for every individual property.

Martensite, ferrite, and austenite were selected as the phase-fraction targets because they represented nearly all the predicted room temperatures. Other phases, including carbides, were present only in trace amounts of less than 0.01% under the investigated compositions and simulation conditions. Their contribution to the overall phase balance was therefore negligible, and they were not included as separate prediction targets. The lower accuracy of martensite and ferrite arises from the small solidification dataset, their abrupt composition-dependent transitions, and the frequent occurrence of values near 0% and 100%. Fixed processing conditions further limit the phase-transformation behavior available for learning. The final heat treatment model achieved a held-out R^2 value of 0.989–0.991, exceeding values reported in related studies. However, direct comparisons should be made cautiously because the datasets, alloy systems, targets, and validation methods differ.

4.6. Final Prediction System and Case Study

The final prediction system is built on the optimized 125-composition model, chosen for its coverage and stability. Given a new alloy composition and processing condition, it returns the corresponding mechanical properties and, for the solidification route, the room-temperature phase fractions enable a candidate to be screened in seconds rather than through a fresh simulation or experiment. Table 7 compares the percentage error of the baseline, transfer-learning, and optimized systems, and Table 8 reports a single-alloy case study with actual, predicted, and absolute-error values at each stage.

Table 7. Error comparison for the baseline, transfer-learning, and continuous optimization prediction systems.

Process	Property	Baseline Err (%) ↓	TL Err (%) ↓	Optimization Err (%) ↓
Solidification	YS	7.716	3.375	0.035
	UTS	8.603	2.340	0.077
	HV	6.795	1.359	0.571
	Martensite	0.244	4.790	0.327
	Ferrite	—	—	—
	Austenite	—	28.515	—
Heat Treatment	YS	3.644	4.653	1.849
	UTS	2.574	4.665	1.232
	HV	0.129	4.053	1.967

Note: The downward arrow (↓) means a lower error is better. Values marked as— are undefined or not reported.

Table 8. Single-alloy case-study comparison with actual, predicted, and absolute-error values.

Process	Property	Baseline			Transfer Learning			Continuous Optimization		
		Actual	Predicted	Abs. Err	Actual	Predicted	Abs. Err	Actual	Predicted	Abs. Err
Solidification	YS	1650	1777.32	127.32	1629.14	1684.11	54.97	1925.02	1924.34	0.67
	UTS	1850	2009.15	159.15	1846.3	1889.51	43.21	2109.17	2107.53	1.63
	HV	55	58.73	3.7	54.84	54.1	0.74	59.11	59.44	0.33
	Martensite	99.84	99.6	0.24	99.61	94.84	4.77	98.93	99.26	0.3
	Ferrite	0	0	0	0	3.72	3.72	0	0	0
	Austenite	0.12	0.39	0.27	0.28	0.36	0.08	0.89	0.73	0.15
Heat Treatment	YS	1400	1451.02	51.02	1306.9	1246.09	60.80	1059.2	1039.61	19.58
	UTS	1650	1692.46	42.46	1551.8	1479.4	72.39	1315.5	1299.28	16.21
	HV	52	51.93	0.06	48.83	46.85	1.97	42.64	41.8	0.83

Note: YS and UTS are in MPa, HV is in HRC, and phase fractions are in %. Abs. Error = |actual – predicted|.

Table 7 shows the percentage error contracting sharply along the pipeline for the solidification properties, for example, yield-strength error falling from 7.7% at baseline to 3.4% after transfer learning and 0.04% after optimization, and ultimate-tensile-strength error from 8.6% to 2.3% to 0.08%. The heat treatment properties also improved overall, though less monotonically: for some properties, the transfer-learning error briefly exceeded the baseline before optimization brought it well below both, a reminder that the value of transfer learning is realized in combination with subsequent data expansion rather than in isolation. The single-alloy case study in Table 8 tells the same story in absolute terms: for the optimized model, the solidification predictions fell within about 1 MPa of the target strengths and 0.3 HRC of hardness, and the heat treatment predictions within roughly

20 MPa and 0.8 HRC. The directions are also metallurgically sensible: higher carbon and hardenability-promoting additions drove higher predicted strength and hardness, while leaner compositions were predicted to be softer.

The staged expansion of the composition pool did not add data uniformly because each acquisition batch targeted a distinct region of the design space, and the accuracy gains at each stage can be traced in Figure 16 and Table 9. This becomes apparent when the acquired compositions are projected onto carbon content and carbon equivalent (CE_{IIW}). The batch acquired at stage 85 clusters tightly in the medium-carbon, high-hardenability corner, with C between 0.33 and 0.43 wt% and CE_{IIW} between 0.75 and 0.84, which is well above the pool mean of 0.58. Only 14 of the original 72 compositions had occupied this corner, so the first batch nearly doubled its coverage. The later batches then worked through the remaining gaps in sequence and broad infill spanning the transition compositions at stage 95, the lean low-hardenability region at stage 105, and finally the low-carbon, high-alloy and lean transition-zone compositions at stages 115 and 125 (Figure 16a).

Stage-by-stage localization of the model-quality gains (Baseline → TL-72 → 125)

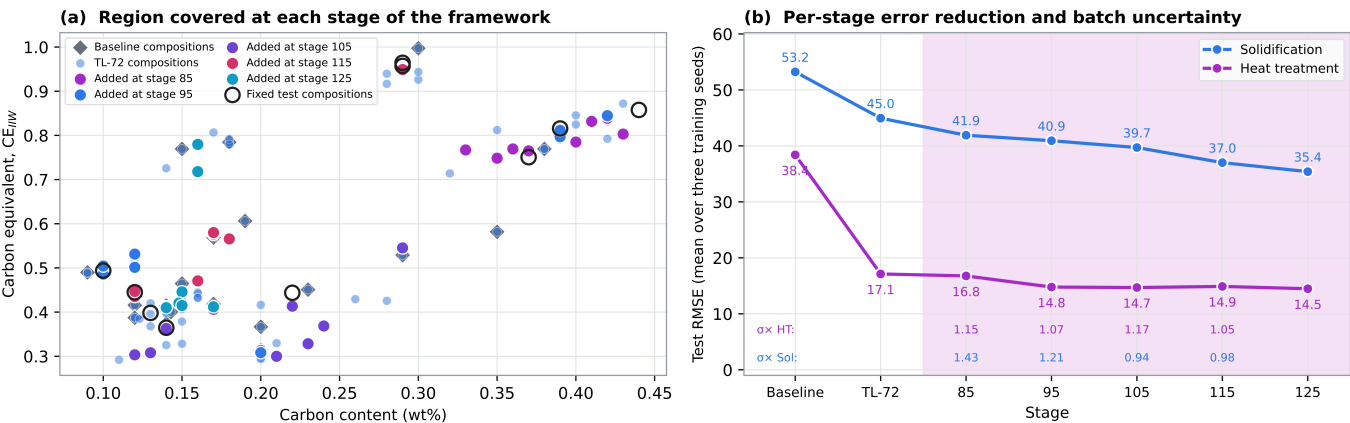


Figure 16. Stage-by-stage localization of the model-quality gains from Baseline to the 125-composition stage.

Table 9. Stagewise summary of composition-space expansion and prediction-error changes.

Stage	C (wt%)	CE_IIW	Region Covered	σ Ratio (HT)	σ Ratio (Sol.)	Δ RMSE (HT)	Δ RMSE (Sol.)
Baseline	0.09–0.39	0.31–1.00	Initial Dataset	-	-	-	-
72→85	0.33–0.43	0.75–0.84	High-hardenability corner	1.16	1.43	0.5	3.0
85→95	0.10–0.42	0.31–0.84	Broad infill (incl. transition)	1.07	1.22	2.1	1.0
95→105	0.12–0.29	0.30–0.55	Lean, low hardenability	1.16	0.94	0.3	1.2
105→115	0.12–0.29	0.44–0.95	Low-C, high-alloy mixed	1.02	0.98	0.0	2.7
115→125	0.14–0.17	0.41–0.78	Lean transition-zone	–	–	0.1	1.6

Note: CE_{IIW} = Carbon equivalent by the IIW (International Institute of Welding) formula. So, CE_{IIW} = C + Mn/6 + (Cr + Mo + V)/5 + (Ni + Cu)/15 (all in wt%). It compresses the effect of all the alloying elements into one number that expresses the steel’s hardenability—how strongly it forms martensite on cooling. Carbon counts at full weight; Mn at 1/6; Cr, Mo, V at 1/5; Ni and Cu at 1/15, reflecting each element’s relative potency.

The MC-Dropout uncertainty of each batch, evaluated against the unselected candidate pool now of selection, tracks these gains closely in Figure 16 and Table 9. For the solidification model, the stage-85 batch carried 1.43 times the pool-mean uncertainty, which is the strongest signal observed in this study and yielded the largest single optimization-stage

reduction, from an RMSE of 45.0 to 41.9 on the fixed held-out test set of ten compositions (all values averaged over three training seeds). The solidification error continued to decrease through stage 125 (35.4), reflecting the greater intrinsic difficulty of the phase-fraction targets, which change abruptly across the hardenability transition. The heat treatment model behaved differently: the transfer-learning expansion provided the dominant gain (38.4 to 17.1), the optimization stages refined the error to 14.5 with essentially the entire gain realized by stage 95, and the later stages produced only fluctuations within seed-to-seed variability coinciding with the decay of the batch uncertainty ratio toward unity. A leave-one-composition-out ablation, repeated over the same training seeds, showed that no single acquired composition dominates any stage; removing any individual alloy was compensated by its neighbors in the same region. So, the improvements are collective, regional-coverage effects. Taken together, the staged gains arose from successively resolving sparsely sampled regions of the composition space, and the eventual plateau marks the exhaustion of the uncertainty signal, the point at which no under-represented region remained for the acquisition rule to exploit.

The practical value of the system lies in its role as a fast first filter. Rather than evaluating every candidate by simulation or experiment, a designer can scan many compositions and tempering schedules in moments and reserve costly validation for the few that look most promising, cutting the number of trial-and-error iterations that dominate conventional development.

The system is best understood as a decision-support tool rather than a replacement for physical validation. Predictions are reliable within the composition and processing window represented in the training data. However, extrapolation beyond that window carries added uncertainty, so the alloys it nominates should still be confirmed by targeted simulation or experiment.

4.7. Limitations and Future Research Directions

Future work will focus on validating the proposed transfer learning and continuous optimization framework using high-fidelity experimental data from DED-processed low-alloy steels. Although the present model was trained on CALPHAD-generated simulation data, experimental measurements of yield strength, ultimate tensile strength, hardness, and phase fractions are needed to further assess the reliability of the predictions and reduce the simulation-to-experiment gap. Incorporating experimental results into the target and continuous optimization stages will strengthen the model's applicability to real DED material qualification.

The current dataset used fixed values for several important process descriptors, including cooling rate, prior austenite grain size, austenitization temperature, and initial microstructure. Therefore, future datasets should independently vary these parameters so that the model can explicitly learn their effects on solidification phase fractions and heat treatment-dependent mechanical properties. Expanding the heat treatment design space to include additional austenitization, quenching, tempering, and multi-step heat treatment routes would also improve the framework's ability to recommend practical post-processing schedules.

Future work should also extend the framework to broader alloy systems and more complex DED conditions, including different low-alloy steel families, multi-principal-element alloys, and variable deposition parameters. In addition, uncertainty-guided active learning can be integrated with multi-objective optimization to identify compositions and heat treatment schedules that balance strength, hardness, toughness, and phase stability. With closed-loop experimental feedback and real-time DED process data, the proposed approach could be developed into an adaptive digital tool for faster material qualification,

intelligent heat treatment selection, and process optimization in industry-ready additive manufacturing systems.

5. Conclusions

This study developed and evaluated a transfer-learning and continuous optimization framework for predicting heat treatment-dependent mechanical properties and solidification phase fractions of DED-processed low-alloy steels. The main findings are summarized as follows:

- A robust framework was developed using CALPHAD-generated simulation data, a baseline multilayer perceptron surrogate, transfer learning, and staged continuous optimization. The framework was applied to both solidification and heat treatment outputs, including yield strength, ultimate tensile strength, hardness, and phase fractions of martensite, ferrite, and austenite.
- For the heat treatment outputs, the average R^2 increased from 0.757 in the baseline stage to 0.929 after transfer learning and approximately 0.997 after continuous optimization. The average RMSE was reduced by nearly 78% relative to the transfer-learning model.
- The final optimized 125-composition heat treatment model achieved strong held-out performance, with R^2 values of 0.989, 0.990, and 0.991 for yield strength, ultimate tensile strength, and hardness, respectively. This confirms that the optimized model accurately captures trends in tempering-dependent mechanical properties.
- For the solidification outputs, the model also improved consistently, with mechanical-property predictions reaching R^2 values above 0.94 after optimization. However, martensite and ferrite remained more difficult to predict because of their sharp transformation behavior and boundary-limited phase fractions.
- The acquisition-strategy comparison showed that Bayesian-UCB provided the most reliable and data-efficient improvement. By combining predicted performance with MC-Dropout uncertainty, Bayesian-UCB produced smoother error reduction than Random selection or purely Greedy selection.
- The final prediction system showed practical value in the case study, reducing solidification prediction errors for yield strength, ultimate tensile strength, and hardness to 0.035%, 0.077%, and 0.571%, respectively. For heat treatment, the corresponding optimized errors were 1.849%, 1.232%, and 1.967%, demonstrating the framework's potential as a rapid screening tool for alloy composition and tempering-condition selection prior to experimental validation.

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

Funding: This research was funded by the US Navy's and Product Innovation and Engineering's (PINE) SBIR/STTR projects support (Contract numbers N68335-24-C-0215, N6833526C0027, and N68335-24-C-0577).

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

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

References

- Li, L.; Zhang, X.; Liou, F. Experimental and Numerical Investigation in Directed Energy Deposition for Component Repair. *Materials* **2021**, *14*, 1409. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rahman, A.; Ali, M.H.; Mahmood, M.A.; Liou, F. Industry 5.0, human–machine interface, and smart manufacturing in additive manufacturing—A recent trend. *Int. J. Adv. Manuf. Technol.* **2025**, *141*, 109–140. [\[CrossRef\]](#)
- Bartels, D.; von Lattre-Hertel, A.; Novotny, T.; Mohr, A.; Hill, H.; Merklein, C.; Schmidt, M. Directed Energy Deposition of Low-Alloyed Steels: An Insight on Microstructural and Mechanical Properties. *Steel Res. Int.* **2023**, *94*, 2200925. [\[CrossRef\]](#)
- Kelley, J.; Newkirk, J.W.; Bartlett, L.N.; Isanaka, S.P.; Sparks, T.; Alipour, S.; Liou, F. Development of Robust Steel Alloys for Laser-Directed Energy Deposition via Analysis of Mechanical Property Sensitivities. *Micromachines* **2024**, *15*, 1180. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rahman, A.; Isanaka, S.P.; Liou, F. A Comprehensive Study of Cooling Rate Effects on Diffusion, Microstructural Evolution, and Characterization of Aluminum Alloys. *Machines* **2025**, *13*, 160. [\[CrossRef\]](#)
- Ebrahimi, M.; Xie, Y.; Chi, C. Effect of cladding parameters on microstructure and defects in direct laser metal deposition of 24CrNiMo steel. *J. Laser Appl.* **2021**, *33*, 012007. [\[CrossRef\]](#)
- Wang, Q.; Zhang, S.; Zhang, C.; Wang, J.; Shahzad, M.B.; Chen, H.; Chen, J. A high strength low alloy steel fabricated by direct laser deposition. *Vacuum* **2019**, *161*, 225–231. [\[CrossRef\]](#)
- Deirmina, F.; Amirabdollahian, S.; Pellizzari, M.; Molinari, A. Effect of Different Post-Processing Thermal Treatments on the Fracture Toughness and Tempering Resistance of Additively Manufactured H13 Hot-Work Tool Steel. *Metals* **2024**, *14*, 112. [\[CrossRef\]](#)
- Yan, Z.; Guo, J.; Zou, X. Effect of Post-Heat Treatment on the Mechanical and Residual Stress Behavior of Pulsed Wave S316L Fabricated by Directed Energy Deposition. *Sensors* **2024**, *24*, 7457. [\[CrossRef\]](#) [\[PubMed\]](#)
- Liu, Q.; Yan, Y.; Iqbal, M.Z.; Tariq, M.; Barnoush, A.; Aissa, B.; Farhat, H. Enhanced hydrogen embrittlement resistance of directed energy deposition 316 L by post-manufacturing heat treatment. *npj Mater. Degrad.* **2025**, *9*, 62. [\[CrossRef\]](#)
- Xie, X.; Bennett, J.; Saha, S.; Lu, Y.; Cao, J.; Liu, W.K.; Gan, Z. Mechanistic data-driven prediction of as-built mechanical properties in metal additive manufacturing. *npj Comput. Mater.* **2021**, *7*, 86. [\[CrossRef\]](#)
- Fu, K.; Zhu, D.; Zhang, Y.; Zhang, C.; Wang, X.; Wang, C.; Jiang, T.; Mao, F.; Zhang, C.; Meng, X.; et al. Predictive Modeling of Tensile Strength in Aluminum Alloys via Machine Learning. *Materials* **2023**, *16*, 7236. [\[CrossRef\]](#) [\[PubMed\]](#)
- Shaheen, M.A.; Presswood, R.; Afshan, S. Application of Machine Learning to predict the mechanical properties of high strength steel at elevated temperatures based on the chemical composition. *Structures* **2023**, *52*, 17–29. [\[CrossRef\]](#)
- Luo, C.; Hansson, K.; Song, Z.; Ågren, D.; Persson, E.S.; Cederholm, F.; Xuan, C. Modelling Microstructure in Casting of Steel via CALPHAD-Based ICME Approach. *Alloys* **2023**, *2*, 321–343. [\[CrossRef\]](#)
- Li, X.; Zheng, M.; Pan, H.; Mao, C.; Ding, W. An integrated design of novel RAFM steels with targeted microstructures and tensile properties using machine learning and CALPHAD. *J. Mater. Inform.* **2024**, *4*, 27. [\[CrossRef\]](#)
- Tang, Y.; Rahmani Dehaghani, M.; Wang, G.G. Review of transfer learning in modeling additive manufacturing processes. *Addit. Manuf.* **2023**, *61*, 103357. [\[CrossRef\]](#)
- Liu, S.; Stebner, A.P.; Kappes, B.B.; Zhang, X. Machine learning for knowledge transfer across multiple metals additive manufacturing printers. *Addit. Manuf.* **2021**, *39*, 101877. [\[CrossRef\]](#)
- Sun, H.; Zhang, H.; Ren, G.; Zhang, C. A Knowledge Transfer Framework for General Alloy Materials Properties Prediction. *Materials* **2022**, *15*, 7442. [\[CrossRef\]](#) [\[PubMed\]](#)
- Jia, Q.; Xu, K.; Li, C.; Kan, G.; Liu, Y.; Ren, H.; Zhang, S.; Lou, M.; Chang, K. Transfer learning-assisted multi-objective optimization of mechanical properties for particle reinforced aluminum matrix composites. *Mater. Futur.* **2026**, *5*, 015602. [\[CrossRef\]](#)
- Yang, F.; Zhao, W.; Ru, Y.; Lin, S.; Huang, J.; Du, B.; Pei, Y.; Li, S.; Gong, S.; Xu, H. Transfer learning enables the rapid design of single crystal superalloys with superior creep resistances at ultrahigh temperature. *npj Comput. Mater.* **2024**, *10*, 149. [\[CrossRef\]](#)
- Liu, Y.; Zhu, D.; Tian, Z.; Liu, J.; Ran, X.; Wang, Z.; Tang, C.; Wang, L.; Xu, W.; Lu, X. Transfer learning-enabled performance prediction of metallic materials: Methods, applications and prospects. *Int. J. Miner. Metall. Mater.* **2026**, *33*, 749–767. [\[CrossRef\]](#)
- Ghorbani, M.; Boley, M.; Nakashima, P.N.H.; Biribilis, N. An active machine learning approach for optimal design of magnesium alloys using Bayesian optimisation. *Sci. Rep.* **2024**, *14*, 8299. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gao, J.; Zhong, J.; Liu, G.; Zhang, S.; Zhang, J.; Liu, Z.; Song, B.; Zhang, L. Accelerated discovery of high-performance Al-Si-Mg-Sc casting alloys by integrating active learning with high-throughput CALPHAD calculations. *Sci. Technol. Adv. Mater.* **2023**, *24*, 2196242. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rao, Z.; Bajpai, A.; Zhang, H. Active learning strategies for the design of sustainable alloys. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2024**, *382*, 20230242. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kusne, A.G.; Yu, H.; Wu, C.; Zhang, H.; Hatrick-Simpers, J.; DeCost, B.; Sarker, S.; Oses, C.; Toher, C.; Curtarolo, S.; et al. On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat. Commun.* **2020**, *11*, 5966. [\[CrossRef\]](#) [\[PubMed\]](#)

26. Liang, H.; Wang, C.; Yu, H.; Kirsch, D.; Pant, R.; McDannald, A.; Kusne, A.G.; Zhao, J.-C.; Takeuchi, I. Real-time experiment-theory closed-loop interaction for autonomous materials science. *Sci. Adv.* **2025**, *11*, eadu7426. [CrossRef] [PubMed]
27. Rahman, A.; Ali, H.; Mahmood, M.A.; Malik, A.W.; Liou, F. *A Machine Learning Model to Predict Mechanical Property of Directed Energy Deposition Processed Low ALLOY Steels*; University of Texas at Austin: Austin, TX, USA, 2025.
28. Rahman, A.; Ali, M.H.; Malik, A.W.; Mahmood, M.A.; Liou, F. Physics-Based Machine Learning Framework for Predicting Structure-Property Relationships in DED-Fabricated Low-Alloy Steels. *Metals* **2025**, *15*, 965. [CrossRef]
29. Reddy, N.S.; Krishnaiah, J.; Hong, S.-G.; Lee, J.S. Modeling medium carbon steels by using artificial neural networks. *Mater. Sci. Eng. A* **2009**, *508*, 93–105. [CrossRef]
30. Woolman, J.; Mottram, R.A. (Eds.) *The Mechanical and Physical Properties of the British Standard EN Steels (B.S. 970–1955)*; Elsevier: Amsterdam, The Netherlands, 1969. [CrossRef]
31. Qi, C.; Wang, C.; Fu, D.; Shao, L.; Zhou, K.; Zhao, Z. A hybrid knowledge-guided and data-driven method for predicting low-alloy steels performance. *Comput. Mater. Sci.* **2025**, *249*, 113602. [CrossRef]
32. NIMS Materials Database (MatNavi). Available online: <https://mits.nims.go.jp/> (accessed on 1 November 2025).
33. Data for Mechanical Properties of Low-Alloy Steels. 2021. Available online: <https://www.Kaggle.Com/Discussions/General/242077> (accessed on 1 November 2025).
34. Wang, Z.; Wang, C.; Wu, L.; Sun, G. Investigation on the Microstructure and Mechanical Properties of X70 Pipeline Steel Fabricated by Laser-Directed Energy Deposition. *Materials* **2025**, *18*, 4997. [CrossRef] [PubMed]
35. Borisov, S.; Borisova, Y.; Tkachev, E.; Kniaziuk, T.; Kaibyshev, R. Tempering Behavior of a Si-Rich Low-Alloy Medium-Carbon Steel. *Metals* **2023**, *13*, 1403. [CrossRef]
36. Xiong, Y.; Wen, D.; Zheng, Z.; Sun, C.; Xie, J.; Li, J. Effect of heat treatment on microstructure and mechanical properties of directed energy deposition-Arc 300M steel. *Mater. Charact.* **2023**, *198*, 112756. [CrossRef]
37. Choi, D.-K. Data-Driven Materials Modeling with XGBoost Algorithm and Statistical Inference Analysis for Prediction of Fatigue Strength of Steels. *Int. J. Precis. Eng. Manuf.* **2019**, *20*, 129–138. [CrossRef]
38. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: New York, NY, USA, 2016; pp. 785–794. [CrossRef]
39. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Available online: <https://github.com/Microsoft/LightGBM> (accessed on 1 November 2025).
40. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased Boosting with Categorical Features. Available online: <https://github.com/catboost/catboost> (accessed on 1 November 2025).
41. Grinsztajn, L.; Oyallon, E.; Varoquaux, G. Why do tree-based models still outperform deep learning on typical tabular data? *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 507–520. [CrossRef]
42. Gorishniy, Y.; Rubachev, I.; Khrulkov, V.; Babenko, A. Revisiting Deep Learning Models for Tabular Data. Available online: <https://github.com/yandex-research/rtdl> (accessed on 1 November 2025).

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