Materials Research
Express

INCLUSIVE PUBLISHING
TRUSTED SCIENCE

PURPOSE-LED
PUBLISHING™

PAPER • **OPEN ACCESS**

Leakage-controlled pre-sintering property prediction in A7075/B_4C/MXene functionally graded aluminum matrix composites

To cite this article: Musa Yilmaz *et al* 2026 *Mater. Res. Express* **13** 216502

View the article online for updates and enhancements.

You may also like

- Dissipative flow of hybrid nanoliquid (H_2O-aluminum alloy nanoparticles) with thermal radiation
Ikram Ullah, Tasawar Hayat, Ahmed Alsaedi et al.

- Imaging of three-dimensional crack open/closed distribution by nonlinear ultrasonic phased array based on fundamental wave amplitude difference
Yoshikazu Ohara, Hiromichi Nakajima, Sylvain Haupert et al.

- Studies on microstructure and mechanical behaviour of A7075- Flyash/SiC hybrid metal matrix composites
V Venkata Reddy, M. Gopi Krishna, K. Praveen Kumar et al.

This content was downloaded from IP address 194.27.210.81 on 09/11/2026 at 21:04

Materials Research Express

Check for updates

OPEN ACCESS

RECEIVED
10 July 2026

REVISED
8 October 2026

ACCEPTED FOR PUBLICATION
26 October 2026

PUBLISHED
3 November 2026

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

PAPER

Leakage-controlled pre-sintering property prediction in A7075/B_4C/MXene functionally graded aluminum matrix composites

Musa Yilmaz[1], **Mürsel Ekrem**[2] and **Hasmet Cagri Sezgen**[2,*]

[1] Yozgat Bozok University, Ahmet Efendi Mahallesi, TOKİ Konutları Yanı, 3500. Cadde No: 4, 66700 Sorgun Yozgat, Türkiye
[2] Mechanical Engineering Department, Faculty of Engineering and Architecture, Necmettin Erbakan University, Köyceğiz Mah. Demeç Sok. No: 44/1, Meram, Konya 42090, Türkiye
[*] Author to whom any correspondence should be addressed.

E-mail: hasmetcagri.sezgen@erbakan.edu.tr

Keywords: A7075 aluminum alloy, B_4C, MXene, functionally graded composites, powder metallurgy, machine learning, nested cross-validation

Supplementary material for this article is available online

Abstract

Designing A7075/B_4C/MXene functionally graded aluminum matrix composites requires process, composition and architecture decisions before costly sintering and mechanical testing are completed. This study evaluates whether descriptors available before sintering can support property-oriented experiment prioritization within a defined powder-metallurgy process window. A traceable dataset of 43 specimen codes was assembled from a Taguchi L27 screening backbone, a concentrated MXene-bearing follow-up branch and one-hour milling checks. Separate regression models were developed for Leeb-derived HBW-equivalent hardness, ultimate compressive strength, specific compressive energy absorption and sintered density. Regularized linear and constrained tree-based learners were evaluated using leakage-controlled nested cross-validation, with preprocessing, feature selection, hyperparameter tuning and model selection confined to the corresponding training partitions. Across the three reported seeds, mean outer-fold R^2 ranged from 0.678 to 0.875 for ultimate compressive strength, 0.480 to 0.861 for specific compressive energy absorption, 0.526 to 0.560 for HBW-equivalent hardness and 0.164 to 0.426 for sintered density. Empirical prediction intervals, residual diagnostics, SHAP-based feature attribution and partial-dependence profiles were used to examine model behavior beyond aggregate accuracy values. These interpretation outputs are treated as model-level associations rather than causal metallurgical evidence. Within the investigated A7075/B_4C/MXene processing and architecture window, the workflow provides a leakage-aware basis for prioritizing pre-sintering experiments, while transfer to other processing routes or independent fabrication campaigns requires further validation.

1. Introduction

A7075-based aluminum matrix composites reinforced with B_4C and $Ti_3C_2T_x$ MXene provide a route for combining low density, ceramic load support and interface-sensitive reinforcement effects. In a functionally graded architecture, however, the response cannot be represented by a single reinforcement fraction. Layer count, through-thickness B_4C distribution, MXene content, compaction, mechanical alloying and thermal history jointly influence densification, interlayer integrity, hardness and compressive deformation [1–3]. Recent work on AA7075/B_4C/TiB_2 composites has also used data-driven models to resolve nonlinear reinforcement-temperature-tribological relationships, indicating that machine learning is increasingly being used in AA7075-based ceramic-reinforced systems while preserving the distinction between different processing routes and target responses [4]. The resulting process-structure-property landscape is

© 2026 The Author(s). Published by IOP Publishing Ltd

heterogeneous and coupled rather than monotonic. Broader reviews of aluminum-based B_4C composites and powder-metallurgy aluminum matrix composites likewise emphasize that reinforcement content, particle distribution, compaction and thermal processing must be considered together when interpreting mechanical performance [5, 6]. MXene- and MAX-derived reinforcement routes add further interface and reaction-path complexity to aluminum or B_4C-containing systems [7, 8].

Powder-metallurgy fabrication and mechanical characterization of graded composites are comparatively low-throughput. Decisions must be made before the final density and mechanical responses are known, while improvements in hardness or strength may be accompanied by reductions in deformation capacity or energy absorption. Data-driven prediction from starting-material descriptors and sintering conditions has been used to guide fracture-toughness design in SiC ceramics, illustrating the value of pre-processing decision support in low-throughput material systems [9]. For the present A7075/B_4C/MXene program, the practical question is not whether experiments can be replaced. It is whether the next experiments can be prioritized with less reliance on trial-and-error fabrication.

Taguchi designs and analysis-of-variance methods remain useful for economical factor screening and main-effect assessment in metal-matrix composite processing [10, 11]. Their role in the present work is to provide the structured experimental backbone. Machine-learning models are introduced as a complementary prediction layer because the merged dataset contains coupled process, composition and graded-architecture descriptors that are difficult to represent through a single low-order response surface. Orthogonal experimental design has likewise been combined with machine learning to reduce the experimental burden of process-window exploration while preserving structured factor variation [12, 13]. Comparative studies on aluminum-matrix surface composites and aluminum-alloy processing have evaluated linear regression, support-vector regression, artificial neural networks and XGBoost, supporting the use of both parsimonious and nonlinear learners under a common process-property problem [14, 15].

In small experimental materials datasets, predictive accuracy depends not only on the selected algorithm but also on how the evaluation boundary is maintained. Scaling, imputation, feature filtering, response transformation and hyperparameter selection can produce optimistic estimates when information from evaluation samples enters model development [16, 17]. A single random split is also highly sensitive to the composition of a small test subset [18]. Nested cross-validation addresses the model-selection boundary by separating inner-loop tuning from outer-fold evaluation, although it does not by itself establish external validity or constitute a complete uncertainty-quantification framework [19–21]. These constraints are especially relevant in materials-informatics studies with tens of records, engineered descriptors and correlated process histories [22–24].

The present study addresses this validation and positioning problem for a data-scarce A7075/B_4C/MXene functionally graded composite system. It uses a published powder-metallurgy campaign that established the A7075/B_4C Taguchi L27 screening backbone, specimen-code structure and mechanical-property characterization route as the experimental basis [25]. The current work consolidates those code-level records with a concentrated MXene-bearing follow-up branch and one-hour milling checks, giving 43 specimen codes in the combined modeling table. Three bounded contributions follow. First, the study links descriptors available before sintering to four measured responses in a traceable A7075/B_4C/MXene graded powder-metallurgy program. Second, it evaluates regularized linear and constrained tree-based models under a leakage-controlled nested-validation workflow in which trainable preprocessing and model-selection operations are restricted to the relevant training partitions. Third, it reports predictive metrics together with residual, empirical-interval and SHAP/PDP-based diagnostics, allowing model behavior to be discussed beyond a single accuracy value. The resulting claims are restricted to experiment prioritization within the investigated process window.

2. Materials and methods

The experimental route links material selection, specimen fabrication, mechanical testing and dataset formation. The fabrication sequence is summarized in figure 1.

2.1. Material system and specimen fabrication

A7075 aluminum alloy was selected as the matrix because the study required a lightweight system with high specific strength and sensitivity to processing history. B_4C was used as the primary ceramic reinforcement, and the graded design combined an aluminum-rich center with harder outer layers.

The starting powders were commercial-purity A7075, B_4C, and Ti_3AlC_2 MAX precursor. Supplier-reported average particle sizes were 70 μm for A7075, 40 μm for B_4C, and 44 μm for Ti_3AlC_2. The corresponding nominal densities were 2.81, 2.52, and 4.2 $g\,cm^{-3}$, respectively. These powders were supplied

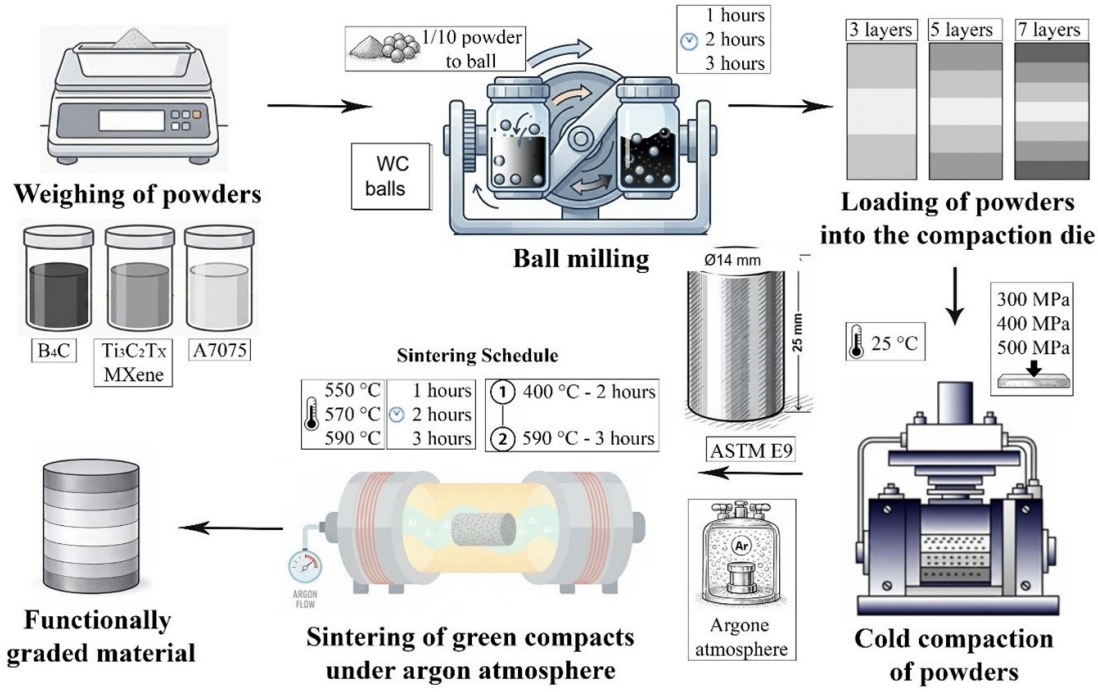


Figure 1. Experimental fabrication route for the A7075/B_4C/MXene specimens, including powder weighing, mechanical alloying, layer-wise die filling, uniaxial cold compaction, argon-atmosphere sintering, and final specimen preparation. The three-, five-, and seven-layer schematics represent the prescribed through-thickness B_4C architectures. The illustrated processing labels should be read together with the detailed route definitions in tables S3–S5.

by ISM Dis Ticaret Pazarlama Ltd Sti., and the MAX precursor was used as the feedstock for subsequent MXene production.

$Ti_3C_2T_x$ MXene was synthesized in-house from the Ti_3AlC_2 MAX precursor by selective removal of the Al-containing layers. The chemical etching was performed by immersing the MAX phase powders in an aqueous HF solution at room temperature for 24 h. The resulting suspension was washed repeatedly with deionized water and centrifuged until the pH of the supernatant reached approximately 6. The sediment was then sonicated for 1 h to promote delamination, and the etched $Ti_3C_2T_x$-rich product was collected and dried in a vacuum oven at 80 °C. The structural and phase characteristics of the synthesized powder are presented in figure 2. Relative changes between the Ti_3AlC_2 precursor and etched-product diffraction patterns are consistent with selective Al removal and the formation of a $Ti_3C_2T_x$-rich product. Field-emission scanning electron microscopy images additionally show the characteristic multilayer, accordion-like morphology expected after etching and delamination. This characterization establishes the physical identity and morphology of the MXene-bearing reinforcement used in the second-stage specimens. It does not characterize the final matrix-reinforcement interface after compaction and sintering or provide an isolated measure of the effect of MXene on the composite responses.

Powder preparation was carried out by mechanical alloying in a Retsch PM100 planetary mill. A powder-to-ball ratio of 1:10 was used with 10 mm tungsten-carbide balls under argon. Stearic acid (1.5 wt.%) was added as a process-control agent. Milling was conducted at 300 rpm with a 3 min milling / 3 min pause cycle, and the mechanical alloying duration was varied at 1, 2, and 3 h.

After milling, the powders were manually stacked in the die according to predetermined functionally graded layer sequences. Each layer charge was weighed separately before placement. Intermediate compaction between layers was intentionally avoided; only light vibration was used to level each surface before the next layer was added. For the FGM specimens, the three layer architecture followed a 5/0/5 wt.% B_4C sequence from surface to center to surface. The five layer profile was 10/5/0/5/10 wt.%, and the seven layer profile was 15/10/5/0/5/10/15 wt.%. The single layer reference series comprised unreinforced A7075 and homogeneous 5, 10, and 15 wt.% B_4C states. The MXene notation used later in the specimen codes does not denote a local addition to one layer only. Instead, the stated MXene level was applied uniformly across all layers of the second stage specimen, while the B_4C profile across layers remained fixed by the selected layer count. This arrangement was intended to retain a more compliant center while moving hardness-dominated material to the surface.

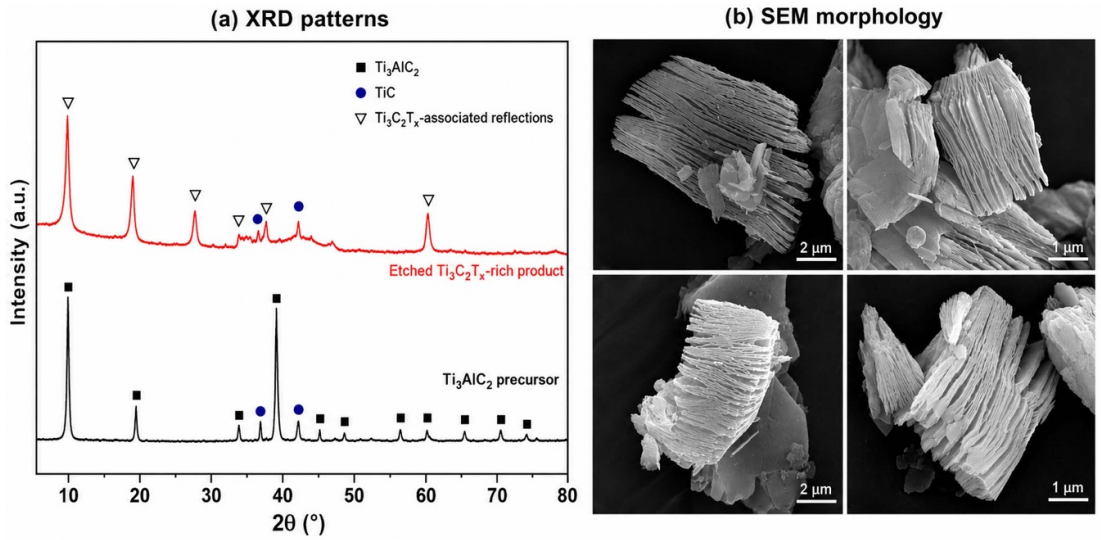


Figure 2. Structural and phase verification of the synthesized $Ti_3C_2T_x$-rich MXene powder used in the second-stage specimens. (a) x-ray diffraction patterns of the Ti_3AlC_2 precursor and the etched product, where the lower trace corresponds to the precursor and the upper trace corresponds to the etched product. The relative changes in the precursor reflections and the MXene-associated low-angle response are consistent with selective Al removal and the formation of a $Ti_3C_2T_x$-rich product; the marked phase assignments are retained for transparency. (b) Field-emission scanning electron microscopy images at the indicated 1 and 2 μm scales, showing the characteristic multilayer, accordion-like morphology of the synthesized powder. These measurements support the inferred phase evolution and document the layered morphology of the reinforcement before its incorporation into the A7075/B_4C graded specimens; they do not independently establish the contribution of MXene to the final composite properties.

Shaped powder compacts were then produced by uniaxial cold pressing in a hydraulic press at 300, 400, or 500 MPa. The green specimens were subsequently sintered in argon in a Protherm PZF-series tube furnace at 550, 570, or 590 °C for 1, 2, or 3 h, using a heating rate of 10 °C min^{-1} and furnace cooling after the hold stage. Across this route, pressing was used to establish green integrity, whereas sintering was used to develop densification, matrix-particle bonding, and interlayer continuity. The overall fabrication sequence is summarized in figure 1.

2.2. Experimental design, process parameters and testing protocols

Because the fabrication space contained several interacting control parameters, the experimental matrix was structured using a Taguchi design rather than by varying one factor at a time. This design serves as the screening backbone for the specimens later used in cross-target modeling.

Five control factors were varied at three levels each: compaction pressure (300, 400, 500 MPa), sintering temperature (550, 570, 590 °C), sintering time (1, 2, 3 h), mechanical alloying time (1, 2, 3 h), and layer count (3, 5, 7). These levels define the main process window considered in the manuscript and are summarized in table S3.

With five factors at three levels, a full factorial study would have required $3^5 = 243$ combinations. The work was therefore organized around a Taguchi L27 orthogonal array. Each run was linked to a specimen code that encodes the main production settings. For example, 13a-1a denotes a specimen milled for 1 h, fabricated with 3 layers, compacted at 300 MPa and sintered for 1 h at 550 °C. The full run-to-code mapping is provided in table S4. In this design, the graded B_4C profile was tied directly to layer count.

In the combined dataset used here, the Taguchi L27 matrix constitutes the inherited A7075/B_4C screening backbone of the fabrication program, whereas MXene-bearing records enter through the concentrated second-stage refinement branch. The merged analysis table therefore combines screening and refinement records.

The Taguchi matrix was used as the screening backbone rather than as the sole source of specimens. Screening and follow-up checks concentrated the second-stage route around 500 MPa compaction and 3 h mechanical alloying, with a revised sintering schedule of 400 °C/2 h followed by 590 °C/3 h and furnace cooling. Under this route, 12 additional FGM codes were produced for the 3, 5 and 7 layer architectures at 0, 1, 3 and 5 wt.% MXene. Together with four single layer reference and pre-test codes, these records are documented in table S5 and define the modeling window for each target.

Compression testing was performed at room temperature in accordance with ASTM 10[9] using an Instron 8801 testing system operated through Bluehill 2 software. Cylindrical specimens with a diameter of 14 mm and a length of 25 mm were loaded at a constant crosshead speed of 1 mm min^{-1}, and three repeats were tested for each specimen type until failure. The code-level compressive strength and specific compressive energy absorption targets were derived by taking the arithmetic mean of these valid repeat tests. The recorded force-displacement data were converted to engineering stress–strain curves using the initial specimen dimensions. Because the layers were stacked sequentially along the compact height during die filling, the compression loading axis was parallel to the through-thickness composition-gradient direction.

Hardness was measured at room temperature using a HARTIP 3000 portable Leeb tester. The non-destructive Leeb rebound method was used to avoid damaging the brittle outer B_4C-rich layers of the structurally graded specimens. The cylindrical compression specimens were machined into 10 × 10 mm coupons with flattened surfaces, and five readings were taken from different points on each specimen surface. The code level hardness target was defined as the arithmetic mean of these five readings. For consistency with the final analysis table, the output was converted to Leeb derived HBW equivalent hardness using the tester's internal conversion algorithm according to the ASTM A956 standard. This hardness target follows the same portable Leeb measurement route used in the published experimental campaign, but is reported here as a consolidated code-level HBW-equivalent response. Sintered density was measured at room temperature by the Archimedes method using the dry mass in air and the suspended mass in water, with the code level value computed as the arithmetic mean of three sequential measurements taken on the retained representative specimen for that code.

Microstructural and phase characterization in the present manuscript focused on the synthesized MXene powder. FE-SEM imaging was performed using a ZEISS field-emission instrument at an accelerating voltage of 2.00 kV to document the layered powder morphology shown in figure 2. Phase analysis was performed using a PANalytical Empyrean diffractometer with Cu-Kα radiation ($\lambda = 1.5406$ Å), a 10–90° scan range, a 0.02° step size, and Bragg–Brentano reflection geometry. Observations concerning reinforcement redistribution, matrix-reinforcement regions, and graded-layer transitions were reported in the published experimental campaign used as the fabrication basis for the present modeling study [25].

Selected post-sintering BSE-SEM images for fabricated specimens are shown in Supplementary figure S6. Associated report-level EDS observations provide elemental context for the selected records. Because the images and reports do not resolve all specimen interfaces or establish surviving $Ti_3C_2T_x$ phase, they are interpreted as contextual evidence rather than phase-resolved interface verification.

Specific compressive energy absorption was defined as the area under the compressive stress–strain response per unit volume,

$$E_c = \int_0^{\varepsilon_f} \sigma \, d\varepsilon \tag{1}$$

where ε_f is the last valid strain value at test termination. In this manuscript, specific compressive energy absorption is reported in MJ m^{-3}.

This quantity is not equivalent to fracture toughness (K_{IC}); the distinction is important because standard K_{IC} protocols are not directly applicable to the layered FGM geometry used here. This target corresponds to the same underlying compression energy quantity used in the published experimental campaign, but is expressed here in an explicit code level form and manuscript notation for machine learning analysis.

During route refinement, S/N ratios and preliminary ANOVA were used to identify influential factors and to narrow the workable process window prior to multivariate modeling. For that reason, detailed ANOVA Tables are not repeated here. The coupled relationships between processing, composition, and structural layers were analyzed with the multivariate machine learning pipeline used in this study. In the broader publication sequence, Taguchi/SN/ANOVA served as the screening and route-refinement logic of the experimental backbone, whereas the present study treats the same fabrication program as the basis for leakage-controlled multivariate prediction across four outputs.

2.3. Dataset construction, variable definitions and quality controls

For data driven analysis, the experimental results were reformulated into a single matrix. Each retained row corresponds to one traceable specimen code and its consolidated measurement set. Consequently, the modeling dataset is a code level consolidated table ($n = 43$), not a coupon level raw log. The final analysis set contained 43 code level records: 27 L27 screening runs, 12 second stage FGM follow up codes (3 K-0X, 5 K-0X, 7 K-0X and the corresponding 1, 3, and 5 wt.% MXene variants), and four

Table 1. Descriptive summary of the principal continuous process inputs and code-level target responses in the merged modeling dataset ($n = 43$ specimen codes).

Variable	Role (Unit)	N	Mean	Std. Dev.	Min	Median	Max
Compaction pressure	Input (MPa)	43	437.21	81.72	300.00	500.00	500.00
Sintering temperature	Input (°C)	43	577.44	16.34	550.00	590.00	590.00
Sintering time	Input (min)	43	142.33	49.03	60.00	180.00	180.00
MA time	Input (h)	43	2.37	0.82	1.00	3.00	3.00
HBW-equivalent hardness	Output (HBW-equivalent)	43	116.45	19.74	81.40	113.40	170.00
Ultimate compressive strength	Output (MPa)	43	238.47	87.95	71.64	219.59	421.43
Specific compressive energy absorption[a]	Output (MJ m^{-3})	43	49.64	35.05	3.08	44.39	106.99
Sintered density	Output (g cm^{-3})	43	2.54	0.08	2.27	2.53	2.70

[a] Specific compressive energy absorption denotes the integrated compressive stress–strain response per unit volume and is reported in MJ m^{-3}. The target values are consolidated code-level responses rather than individual replicate measurements. Sintering time refers to the final high-temperature hold and is reported in minutes.

single layer reference and pre test codes (1 K-SAF, 1 K-5B, 1 K-10B, and 1 K-15B). Table S4 therefore reports only the L27 backbone, whereas table S5 gives the full 43-code master map.

These 16 additional records document refinement around the experimentally workable route. Together, the L27 backbone and the targeted follow up specimens combine screening coverage with concentrated follow up around the workable route for the present FGM system.

The retained variables were separated into process/design inputs and measured outputs. Table 1 reports the key process inputs used throughout the manuscript: compaction pressure, sintering temperature, sintering time, and mechanical alloying time. The four outputs carried forward to the modeling stage were ultimate compressive strength, hardness, sintered density, and specific compressive energy absorption. Table 1 summarizes only the key continuous process variables and response variables. Structural and compositional descriptors such as layer count, structure type, B_4C profile across layers, composition gradient descriptors, and MXene fraction are defined in table 2 and visualized in supplementary figure S2.

Several variables were derived deterministically from the specimen design records rather than entered as free text descriptors. These code derived fields included layer count, graded versus homogeneous structure, B_4C contents by layer, outer, center, and mean B_4C composition, composition gradient measures, and MXene fraction. Building these fields from design metadata preserved traceability and kept information measured after sintering performance information out of the input space.

Data type and unit harmonization were applied before modeling. Percentage expressions were converted to numeric wt.% values, and numeric fields were normalized to a consistent decimal format. In the analysis matrix, 'sintering time' denotes the final high temperature hold only, whereas the hour based route history, including the 400 °C/2 h hold used in the concentrated second stage branch, was preserved separately in table S5 for traceability. Structural categories and continuous measurements were stored separately so that categorical design information could be encoded without disturbing the physical units of the measured variables.

The density feature implementation includes Green_Density_g_cm3 as a pre-sintering candidate. For the reviewer-directed sensitivity check, green compact density was recovered from the supplied raw density workbook and joined by specimen code; it was not reconstructed from a post-sintering target. In that sensitivity check, Route_2step was encoded as 0 for the single-step L27 route and 1 for the 400 °C/2 h + 590 °C/3 h branch, while the final high-temperature hold remained a numeric process descriptor. The originally reported seed-specific table 3 metrics are retained and are not relabeled as outputs of this added sensitivity configuration.

Measurement consolidation was handled at the specimen code level. Compression testing was performed in triplicate for each code, and the stored strength and specific compressive energy absorption targets correspond to the code level values derived from those repeat stress–strain responses rather than to any single curve. Hardness was stored as the arithmetic mean of five Leeb readings taken on the prepared coupon surface and reported as HBW equivalent hardness. Sintered density was stored as the arithmetic mean of three sequential Archimedes measurements taken on the retained representative specimen for that code. The dataset therefore contains 43 consolidated code level records, not 43 isolated physical tests.

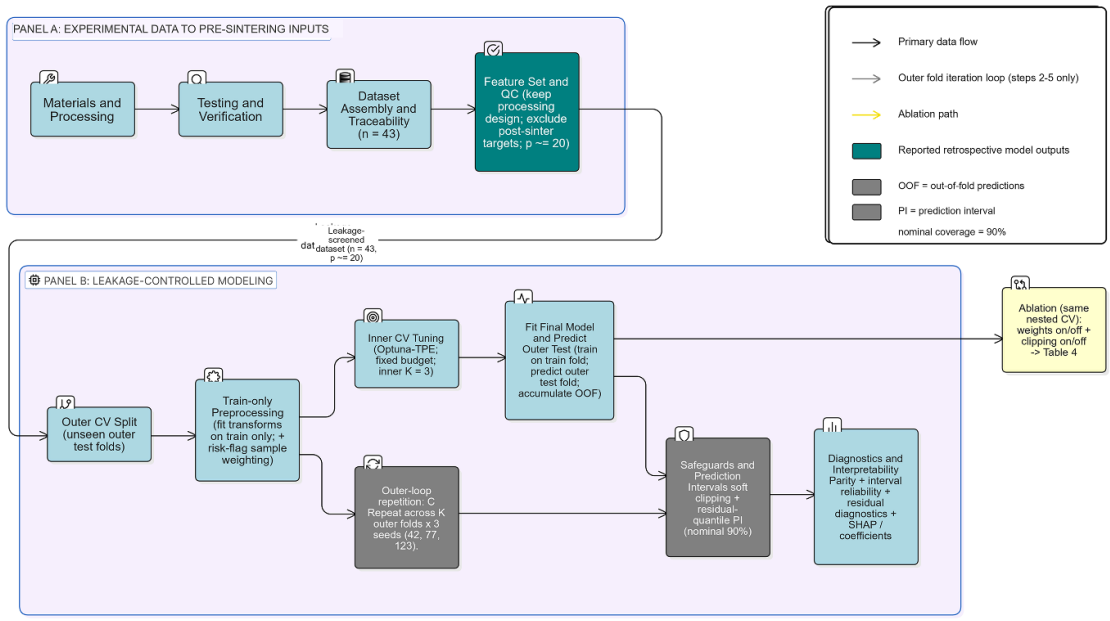


Figure 3. Dataset assembly and leakage-controlled nested-validation workflow applied separately to each target. Trainable preprocessing, feature selection, response transformation, hyperparameter tuning, and model selection were fitted within the relevant training partitions before evaluation on the corresponding outer-test fold. The diagram represents the implemented retrospective evaluation workflow and does not imply external deployment or automated process selection.

Domain-informed quality controls were then applied to the assembled dataset. Non-physical response values, out-of-window process settings, missing specimen identifiers, and coding inconsistencies between fabrication notes and derived descriptors were screened before the modeling stage. Only complete records were retained for the four outputs reported in table 1, which is why all retained variables have $N = 43$ in the final summary.

Integrity checks before analysis were also used to guard against target leakage. Duplicate specimen codes were checked to enforce one row per retained specimen, targets measured after sintering were kept separate from the input dictionary used before sintering, and the pre-sintering input boundary was locked before any machine learning runs were started. As a result, measured outputs such as hardness, compressive strength, sintered density, and specific compressive energy absorption were treated strictly as targets rather than recycled predictors.

Together, tables 1 and 2 and supplementary tables S3 to S5 document the retained variables, process window, and specimen code mapping used for modeling. This traceability is important because the later modeling section relies on a relatively small but fully code level dataset.

3. Machine learning methodology

The modeling stage was designed as a decision-support workflow for use before executing the selected sintering route, with three fixed conditions: pre-sintering input scope, fold-isolated trainable transformations, and traceable evaluation reporting. Under $n = 43$, this structure was used to keep model development aligned with leakage-controlled assessment and bounded process-window interpretation. The end-to-end modeling and evaluation logic used for each target in the study is summarized in figure 3.

3.1. Feature engineering and model inputs

Each response variable, sintered density, HBW-equivalent hardness, ultimate compressive strength, and specific compressive energy absorption, was treated as a separate supervised regression problem. The input space was restricted to process, composition, and structural descriptors that are known before completion of the selected sintering and testing route. In this context, pre-sintering prediction denotes prediction before executing the selected sintering schedule; the prescribed temperature and hold time are therefore known process inputs. Measurements taken after sintering were excluded from the predictor set by rule.

Table 2. Pre-sintering feature groups used to represent process conditions, material composition, and graded architecture. All predictors were defined from information available before completion of the selected sintering and testing route. Several derived composition and architecture descriptors originate from the same prescribed layer profile and were therefore interpreted as correlated engineered representations rather than independent physical variables.

Feature group	Variables included	Unit / type	Why included
Process conditions and measured green state	Compaction pressure, sintering temperature, sintering time, mechanical alloying time, and green compact density in the density feature implementation; the two-step-route indicator is retained for the reviewer-directed route sensitivity check	MPa, °C, min, h, g cm^{-3}, binary	Primary controllable settings plus the measured pre-sintering compact state; route history is retained separately and was encoded only in the reviewer-directed sensitivity check.
Layer architecture	Number of layers, graded/homogeneous indicator	integer, binary	Represents the prescribed structural layout.
Composition descriptors	MXene fraction, outer/center/mean B_4C content, B_4C gradient	wt%, derived scalar	Summarizes composition level and gradient.
Composition profile across layers	B_4C content across Layer 1 through Layer 7	wt%	Preserves through-thickness composition distribution.
Engineered process-interaction descriptors	Pressure x temperature, thermal budget, layer complexity	derived continuous terms (scaled before fitting)	Represents coupled process effects.
Deviation-from-reference terms	Linear and quadratic deviations for temperature, pressure, and time	derived continuous terms (train-fold centered/scaled)	Encodes distance from a reference process window.
Regime and coupling terms	High-MXene regime flag and its pressure/layer/gradient couplings	binary + derived continuous terms	Provides descriptors for branch-associated regime and interaction patterns.
Higher-order engineered proxies	See table S6 for canonical engineered terms	mixed derived units; standard-scaled before fitting	Provides additional bounded process-architecture descriptors.

Note: Only process and design variables available before sintering are included. Targets measured after sintering are excluded. Final pools and retained p values for each target are listed in table S2.

A separate model was preferred to a multioutput model because the dataset is small ($n = 43$), the four responses exhibit different validation behavior, and the transformations, clipping rules, and weighting schemes differ by target. Under these conditions, separate models for each target were easier to audit and easier to keep aligned with the pre-sintering input boundary.

The raw input dictionary was organized into process conditions, layer architecture, composition descriptors, and explicit B_4C profiles across layers, as summarized in table 2. This structure allowed each specimen to be represented both at the global level and across the through-thickness gradient. That distinction matters in functionally graded materials because specimens with similar average composition may still differ substantially in local reinforcement distribution and therefore in their mechanical response.

The second-stage 400 °C/2 h hold was retained as route-history metadata in table S5. In the reviewer-directed route sensitivity check it was encoded as Route_2step = 0 for the single-step branch and 1 for the two-step branch; the final temperature and time descriptors remained numeric predictors. The originally reported target-specific feature pools in table S6 were otherwise retained, and no route-on/off sensitivity result is used as a primary result.

Engineered interaction terms were added to represent physically plausible couplings rather than leaving them to be inferred from a small sample alone. Pressure-temperature and temperature-time terms approximated coupled compaction and thermal exposure effects, a sintering-energy proxy linked hold time with furnace temperature, and layer-complexity and gradient descriptors captured the additional heterogeneity introduced by graded architectures. The aim was not feature inflation for its own sake, but a more faithful representation of the coupled process and architecture descriptors associated with densification and compressive behavior. These engineered terms were not introduced in isolation from the fabrication evidence. The published experimental campaign in the same A7075/B_4C graded system had already shown that compaction pressure, mechanical alloying history, graded reinforcement placement, and densification related effects jointly shape hardness, strength, and density behavior, which motivated the retained interaction and proxy terms used here.

In the manuscript form, the most-used engineered terms are defined as follows: thermal budget $= T \times t$, pressure x temperature $= P \times T$, densification driving force $= P \times T \times \ln(1+t)$, pressure-to-temperature ratio $= P/(T+\epsilon)$, B_4C loading efficiency $= \overline{B_4C} \times$ layer count, gradient-compaction coupling $= \nabla B_4C \times P$, outer-center B_4C contrast $= B_4C_{outer} - B_4C_{center}$, MXene-pressure coupling $=$ MXene $\times P$, gradient x sintering energy $= \nabla B_4C \times T \times t$, MXene-energy absorption index $=$ MXene $\times$ ∇B_4C, and thermal ductility ratio $= T/(P+\epsilon)$. Supplementary table S6 provides the frozen candidate vocabulary.

A second engineered block described the distance of each process setting from a reference operating window. Linear and quadratic deviation terms were computed for pressure, temperature, and time, but the reference values themselves were estimated only on the training partition inside each fold. This restriction prevented the center of the process window from drifting toward information contained in held-out observations and therefore kept these deviation features compatible with leakage-controlled validation.

The final candidate space was kept deliberately bounded. Rather than exposing the algorithms to every possible constructed term, two candidate pools were defined: a composition-enhanced pool (CE), which emphasized higher-order engineered proxies and densification-oriented interactions, and a layer-vector-enhanced pool (LVE), which preserved the through-thickness B_4C distribution more explicitly. Across the four targets, the locked final settings retained $p = 15$ to $p = 20$ variables, as reported in table S2.

In the retained analysis record, feature selection is reported at the level directly supported by table 2 and supplementary table S6. One active pool (CE or LVE) was fixed for each target, and a reduced subset of size p was then selected inside each training partition by a filter-based F-test for regression implemented through SelectKBest. Supplementary table S6 reports the frozen candidate vocabulary, while table S2 reports the retained dimensionality p and final pipeline settings. The exact random states, fold definitions, scaling scope and software-version record are consolidated in supplementary table S9.

Variables unavailable at prediction time were excluded. Hardness, compressive strength, sintered density, and specific compressive energy absorption were treated as outputs rather than predictors.

3.2. Validation strategy and data leakage prevention

Model evaluation was based on nested cross-validation. In this manuscript, a fold denotes one partition of the dataset, the outer test fold denotes the held out partition used only for final evaluation in that split, and the training fold denotes the remaining data used for fitting and inner-loop tuning. Mean outer-fold R^2 denotes the arithmetic mean of the outer-fold test R^2 values, whereas pooled out-of-fold R^2 is computed from the single prediction vector obtained by concatenating all outer-fold test predictions. These two summaries are not interchangeable in a small dataset because they weight folds differently. Throughout the main text and comparative tables, R^2 is reported with its protocol label.

All analyses were conducted on the merged dataset of $n = 43$, which combines the Taguchi experiments with the follow up pre test records defined in the materials and methods workflow.

The outer loop was used for model evaluation and the inner loop for model selection and tuning. Inner K was fixed at 3 for all targets, and split stochasticity was examined with the seed set (42, 77, 123). The final choice was Outer $K = 5$ for sintered density and Outer $K = 3$ for hardness, strength, and specific compressive energy absorption. Table 3 reports the originally reported seed-specific final configuration for each target, while table S1 preserves the full three-seed fold- and seed-level context. The nested design reduces model-selection bias from tuning and preprocessing, but it does not replace independent external validation or provide a complete uncertainty quantification framework.

All trainable preprocessing steps were learned on training partitions only. Imputation, scaling, fold-internal subset reduction, and any response transformation were fitted inside the outer loop training data and then transferred unchanged to the corresponding outer test fold. When a specimen contained fewer than seven layers, the non-existent layer positions were preserved as structurally missing slots in the raw layer-profile encoding, while layer count and graded/homogeneous indicators explicitly carried the absence information. Any numeric imputation required by the estimator was then fitted only on the active training partition and applied unchanged to the outer test fold. No fictitious physical layer composition was injected into the specimen definition, and no statistic from the held out observations was allowed to influence feature construction, transformation parameters, or hyperparameter choice.

Sample reweighting was activated for targets that showed stronger sensitivity to training set outliers. The weights were computed within each fold from QC flags defined on the active outer training partition and were used during training. In the final configurations, weighting remained active for sintered density and specific compressive energy absorption and was effectively inactive for hardness and ultimate compressive strength.

All reported tables, captions, and discussion paragraphs were aligned to the same modeling workflow and protocol labels. The naming used in the manuscript tables remains the primary reference.

Because the engineered descriptors overlap algebraically with their base variables, a fold-internal VIF and condition-number audit was added for the density sensitivity configuration (table S8). The observed redundancy motivates Ridge regularization for density and grouped interpretation of the engineered feature families; feature coefficients and attributions are therefore treated as conditional model quantities rather than isolated physical effects.

3.3. Model family, hyperparameter search and selection criteria

Two model families were evaluated to represent different response structures: Ridge regression as a regularized low-variance linear learner and XGBoost as a constrained tree-based learner for nonlinear and interaction-sensitive responses. The final family was chosen separately for each target on the basis of cross-validated performance and stability. For sintered density, Ridge with strong L2 regularization was selected. For HBW-equivalent hardness, ultimate compressive strength, and specific compressive energy absorption, XGBoost represented nonlinear process-composition patterns more effectively within the observed descriptor support and became the final family. The resulting patterns describe model response structure within that support. Neural-network models were not prioritized because the study design required a small-data comparison between parsimonious linear and constrained tree-based learners.

Hyperparameter search was embedded in the inner loop by using Optuna with a TPE sampler. Optuna optimized the mean inner fold validation R^2 after any response transformation had been inverted back to the native response scale and after the active clipping rule had been applied. Manuscript-level model selection was then based on comparison of mean outer-fold R^2, pooled out-of-fold R^2, RMSE, and stability across folds and seeds. As documented in table S2, the search space was strictly limited to model family hyperparameters (e.g. alpha for Ridge; tree depth and learning rate for XGBoost). The search budget was fixed at 1000 trials for all four targets, without early stopping. Pool selection, dimension p, response transform, clipping switches, and weighting rules were fixed pipeline settings. These architectural alternatives were explored and frozen before the continuous Optuna loops.

In table S2, the explicit search bounds correspond to tunable model family hyperparameters explored inside the inner loop. The reported pool identity, retained dimensionality p, response transform, clipping switch, and weighting floor are the final pipeline settings carried into nested evaluation.

Response transformation was treated as part of the modeling specification rather than as an afterthought. Box–Cox was used for sintered density, Yeo–Johnson was used for strength and specific compressive energy absorption, and no transform was retained for hardness. Each transform was fitted only on the training partition inside the active outer fold, and predictions were back-transformed before metrics were computed on outer test observations.

Selection was not reduced to a single number. Mean outer-fold R^2, pooled out-of-fold R^2, RMSE, and fold/seed dispersion were treated as complementary generalization summaries rather than as interchangeable or externally validated scores. The final choice for each target was made by combining central performance with stability across folds and seeds, then freezing the selected feature pool, feature count, transform, and hyperparameters. Table S2 records the search bounds and final settings, including the CE/LVE pool assignment and the values of p, Outer K, and trial budget for each target.

A limited clipping layer was retained for targets that required upper-end control. Global train-range guardrails were defined within each fold for clipped targets, and additional upper caps based on inputs were retained for sintered density and specific compressive energy absorption within the frozen pipelines.

For the density pipeline, the exact input-based rule was $L = \max(2.20, \min(y_{\text{train}}) - 0.03)$ and $U_0 = \min(3.10, \max(y_{\text{train}}) + 0.08)$. When the theoretical-density proxy p_i (the row-level proxy defined in table S6) was available for a held-out row, $u_i = \text{clip}(p_i + 0.05, L + 0.05, 3.10)$ and $U_i = \max(u_i, U_0)$; otherwise $U_i = U_0$. The returned prediction was $\hat{y}_i = \min(\max(\hat{y}_{i,\text{raw}}, L), U_i)$. The boundaries were computed from the active outer-training partition and applied unchanged to that fold's test predictions. Clipping is only an output constraint: it does not create extrapolation skill or validate predictions at the extremes of the process window.

3.4. Interpretability and diagnostic analyses

Interpretability analyses were used to examine how the final models expressed feature-response patterns within the studied process window and graded composite design. Explainable machine-learning approaches have been used to relate material descriptors to predicted transition behavior in materials systems [26]. In the present work, SHAP and partial-dependence outputs are therefore treated as model diagnostics in engineered feature space rather than as causal metallurgical evidence.

Global importance was examined with SHAP based analyses fitted to a full-data refit of the final pipelines on all $n = 43$ retained records. For the Ridge density model, a linear SHAP explainer was used. For the XGBoost models used for hardness, strength, and specific compressive energy absorption, exact tree SHAP values were computed. Features were ranked by mean absolute SHAP value, and the top 20 labels are reported in figure 6.

Effect direction and interaction structure were then examined with partial dependence plots. One dimensional profiles were generated by sweeping the focal feature from the 5th to the 95th percentile of its observed range while holding the remaining features at their training data medians, and two dimensional surfaces were generated on cross-product grids of the two most influential features. Because the engineered descriptors are correlated, these profiles are interpreted at the level of model response and grouped descriptor behavior, not as isolated physical effects.

Reliability was assessed jointly rather than through a single plot. Parity, calibration, residual-versus-fitted, and normal Q–Q diagnostics were interpreted together so that systematic bias, heteroskedastic behavior, and isolated outlier influence could be separated. Prediction intervals were constructed with a symmetric empirical residual-quantile rule: within each outer fold, $r_{\mathrm{train,OOF}}$ denotes the vector of inner loop out of fold residuals computed on the outer-training partition after any response transformation had been inverted back to the native response scale. The quantity $q = \mathrm{Quantile}(|r_{\mathrm{train,OOF}}|, \alpha)$ was therefore calibrated in native response units only, and the interval $[\hat{y} - q, \hat{y} + q]$ was then applied to the outer-test predictions. This empirical interval construction reports retrospective coverage under the implemented resampling protocol and should not be interpreted as a distribution-free guarantee.

For cross-target comparability, the same interval protocol was used for all four responses, with nominal coverage fixed at 90% ($\alpha = 0.90$). The calibration diagram in figure 5 reports observed versus nominal coverage across four equal-frequency response bins. Only the principal reliability and SHAP summaries are retained in the main manuscript body, while detailed metrics, bounds, and diagnostics are moved to Supplementary Information. Table 3 reports performance results for each target, while table S1 provides the full three-seed context.

4. Results and discussion

This section summarizes the final predictive results together with the key reliability and interpretability diagnostics used to contextualize them. The discussion proceeds from performance summary to configuration progression, model behavior, literature positioning, and study scope.

4.1. Main predictive performance results

Primary results are reported under the evaluation protocol, combining the chosen Outer K for each target with fixed Inner $K = 3$. Mean outer-fold R^2, pooled out-of-fold R^2, and RMSE provide complementary views of fold-balanced behavior, pooled predictive behavior, and error scale. Table 3 summarizes the originally reported seed-specific final configurations, and table S1 provides the full fold- and seed-level context for seeds 42, 77, and 123. The displayed anchors are retained for continuity with the original report, while pooled OOF R^2 remains a complementary summary rather than a replacement.

As table 3 shows, ultimate compressive strength delivered the strongest overall generalization profile among the four targets. Specific compressive energy absorption and HBW-equivalent hardness followed with credible but more conservative predictive behavior, whereas sintered density remained the lowest-scoring task in absolute terms but still retained a usable positive signal under the more stable Ridge-based setting. Because the final protocols differ slightly across responses, the results are most appropriately interpreted as performance summaries for each target within the final evaluation framework. Taken together, the evidence indicates the strongest predictive behavior for ultimate compressive strength, useful predictive behavior for specific compressive energy absorption and HBW-equivalent hardness, and a more modest but positive signal for sintered density. These findings are further supported by the parity patterns in figure 4 and the interval reliability results in figure 5.

The relative difficulty pattern remains visible across the seed contexts reported in table S1. Strength is the most stable target, density is the most difficult target, and specific compressive energy absorption is more sensitive to fold composition than strength. Hardness lies between these cases, with moderate predictive behavior and greater variability than strength.

From a materials perspective, the ranking of target difficulty is also plausible. Ultimate compressive strength is the most coherently predicted response because it reflects the combined outcome of compaction quality, thermal exposure, and graded reinforcement architecture in a comparatively integrated

Table 3. Performance of the reported seed-specific final configurations for the four target responses under the stated nested-validation settings. Mean outer-fold R^2 and pooled out-of-fold R^2 describe different summaries of the same evaluation protocol and should not be interpreted interchangeably. The listed values are representative seed-specific anchors rather than averages across seeds; complete fold- and seed-level results are provided in table S1.

Target	Mean outer-fold R^2	Pooled OOF R^2	RMSE	Model	Outer K	Inner K	Reported seed
HBW-equivalent hardness	0.560	0.601	13.64	XGBoost	3	3	42
Ultimate compressive strength (MPa)	0.875	0.878	30.36	XGBoost	3	3	42
Specific compressive energy absorption (MJ m^{-3})	0.861	0.869	13.35	XGBoost	3	3	123
Sintered density (g cm^{-3})	0.426	0.448	0.06	Ridge	5	3	77

Outer K: Sintered density is reported under Outer $K = 5$, while the other three targets are reported under Outer $K = 3$. RMSE is reported in each target's native unit.

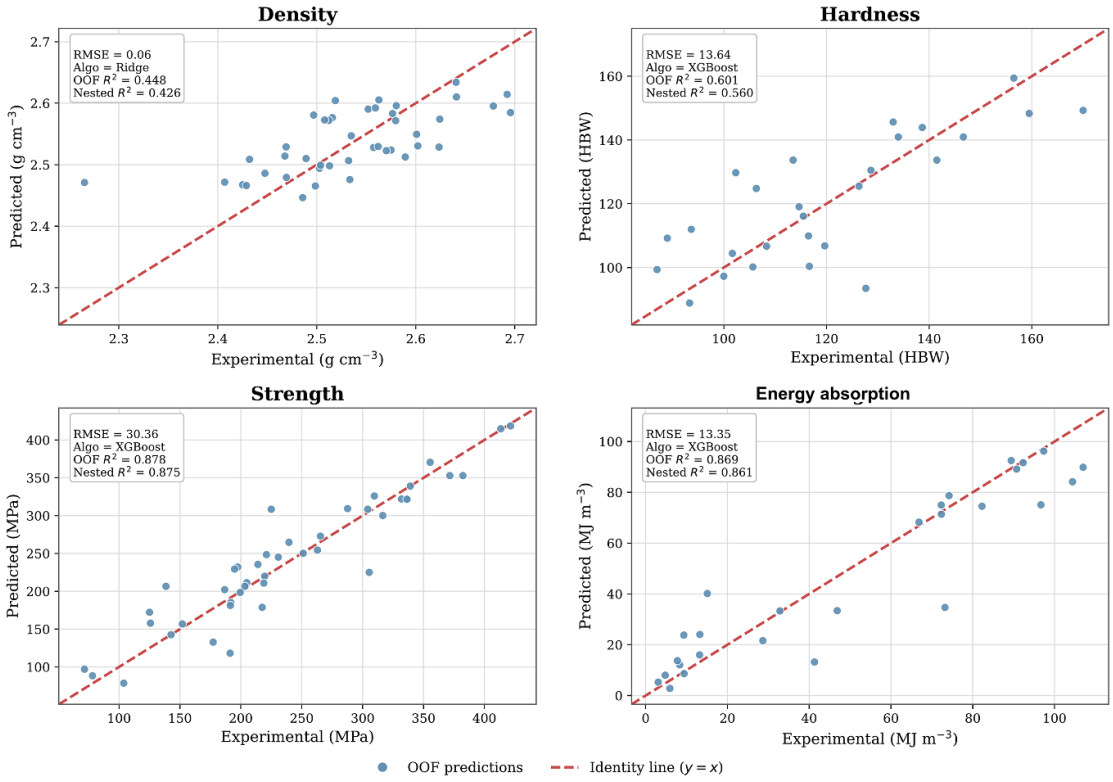


Figure 4. Actual-versus-predicted outer-fold results for the four target properties based on concatenated outer-fold predictions (pooled out-of-fold vectors). In-panel annotations report RMSE, model family, pooled out-of-fold R^2 and mean outer-fold R^2 for the displayed seed-specific configurations. In the hardness panel, HBW denotes the same Leeb-derived HBW-equivalent hardness target defined in section 2.2.

way. Sintered density is more difficult because it is highly sensitive to porosity, local packing heterogeneity, and route history effects that are only partially captured by descriptors available before sintering. HBW-equivalent hardness remains predictable but is influenced not only by reinforcement placement and alloying history, but also by the Leeb-to-HBW conversion used for reporting, which adds an additional layer of measurement abstraction relative to direct static indentation. Specific compressive energy absorption is design relevant but also sensitive to tradeoffs, because the same graded architecture that increases hardening and load-bearing capacity can also reduce deformation capacity or intensify interlayer stress concentration in the more ceramic-rich zones.

The density results are consistent with a response shaped by local porosity, packing heterogeneity, route history, and threshold-like densification behavior. Within the present data, these factors define the target's predictive difficulty rather than separately identified mechanisms. Ridge was retained as the lower-variance model under the collinear feature structure, providing a stable density baseline for experiment prioritization within the observed process window.

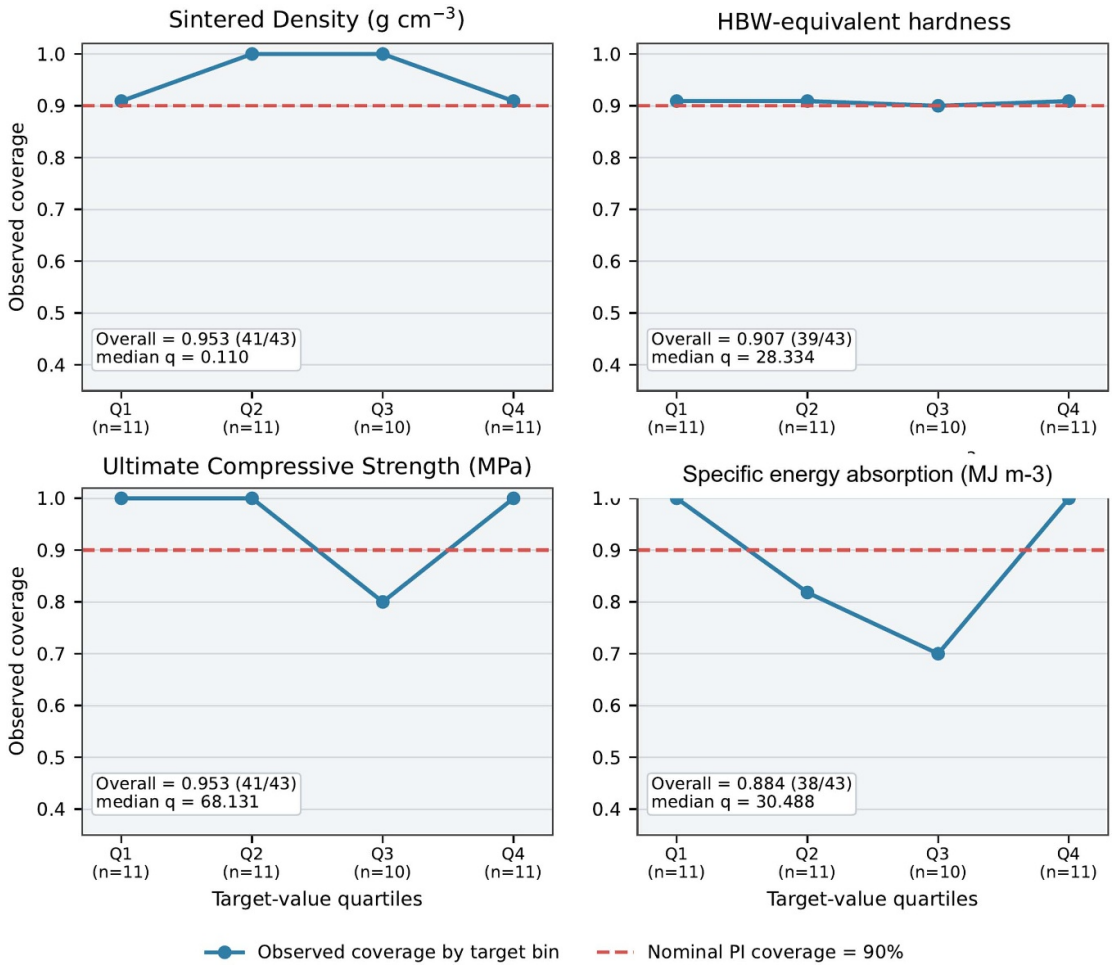


Figure 5. Empirical interval reliability by target-value quartile under nominal 90% prediction intervals. Here, q denotes the absolute-residual quantile estimated from inner-loop out-of-fold residuals within the corresponding outer-training partition after back-transformation to native response units. The plot reports retrospective coverage under the implemented nested-evaluation protocol, not an external or distribution-free guarantee. In the hardness panel, HBW denotes the same Leeb-derived HBW-equivalent hardness target defined in section 2.2.

4.2. Configuration progression from baseline to final model

Table 4 summarizes how the final configurations improve over simpler starting points. The stage-wise gains reflect the jointly refined effect of feature-pool choice, dimensionality p, and hyperparameter settings.

The density line shows a cumulative rather than abrupt improvement pattern. Its gain from the weak baseline to the candidate configuration is already substantial, and the final Ridge-based setting adds a further step without abandoning the same CE-oriented engineered descriptor logic. This behavior is consistent with the role of regularized linear structure, clipping, and training-side QC weighting in a small-sample target where excessive nonlinearity was penalized by instability across folds and seeds rather than rewarded.

Specific compressive energy absorption exhibits the largest stepwise improvement from baseline to final configuration. This gain is attributed to the combined effect of richer LVE descriptors, target-specific transformation, clipping, and active QC weighting in a target that remains more sensitive to fold composition than ultimate compressive strength.

Strength and hardness improve in different ways. Ultimate compressive strength reaches a high performance level already at the candidate stage, with the final model adding a marginal refinement. For HBW-equivalent hardness, the final configuration remains close to the candidate peak while preserving protocol stability.

Across the four targets, the most defensible general gain sources are not isolated one by one in the main text. What table 4 does support is a broader interpretation: leakage controlled validation, domain-informed feature design, target specific transformation policy, retained clipping, and training side QC

Table 4. Recorded progression from the baseline pipeline to the reported final configuration for each target. The differences summarize cumulative changes across the stored modeling configurations and do not isolate the causal contribution of an individual feature pool, transformation, clipping rule, weighting choice, or hyperparameter.

Target	Baseline model Mean outer-fold R^2	Candidate configuration Mean outer-fold R^2	Reported final configuration Mean outer-fold R^2	ΔR^2 (final−baseline)	Final configuration	Selection note
HBW-equivalent hardness	0.267	0.578	0.560	+0.293	XGBoost, LVE feature pool, feature selection enabled, clipping based on inputs applied; sample weighting disabled ($w_{\min} = 1.00$)	Selected after considering the reported outer-fold performance and predefined pipeline criteria.
Ultimate compressive strength (MPa)	0.702	0.801	0.875	+0.173	XGBoost, CE feature pool, Yeo–Johnson response transformation, clipping based on inputs applied; sample weighting disabled ($w_{\min} = 1$)	Selected after considering the reported outer-fold performance and predefined pipeline criteria.
Specific compressive energy absorption ($MJ\,m^{-3}$)	0.311	0.613	0.861	+0.550	XGBoost, LVE feature pool, Yeo–Johnson response transformation, clipping based on inputs applied; sample weighting enabled ($w_{\min} = 0.40$)	Selected after considering the reported outer-fold performance and predefined pipeline criteria.
Sintered density ($g\,cm^{-3}$)	0.107	0.357	0.426	+0.319	Ridge, CE feature pool, Box–Cox response transformation, clipping based on inputs applied; sample weighting enabled ($w_{\min} = 0.15$)	Ridge retained because of lower variability across folds and seeds.

ΔR^2 shows the net change from the baseline pipeline to the reported final configuration. These gains reflect the combined refinement of feature pool, retained dimensionality, transformation, clipping, weighting, and hyperparameters.

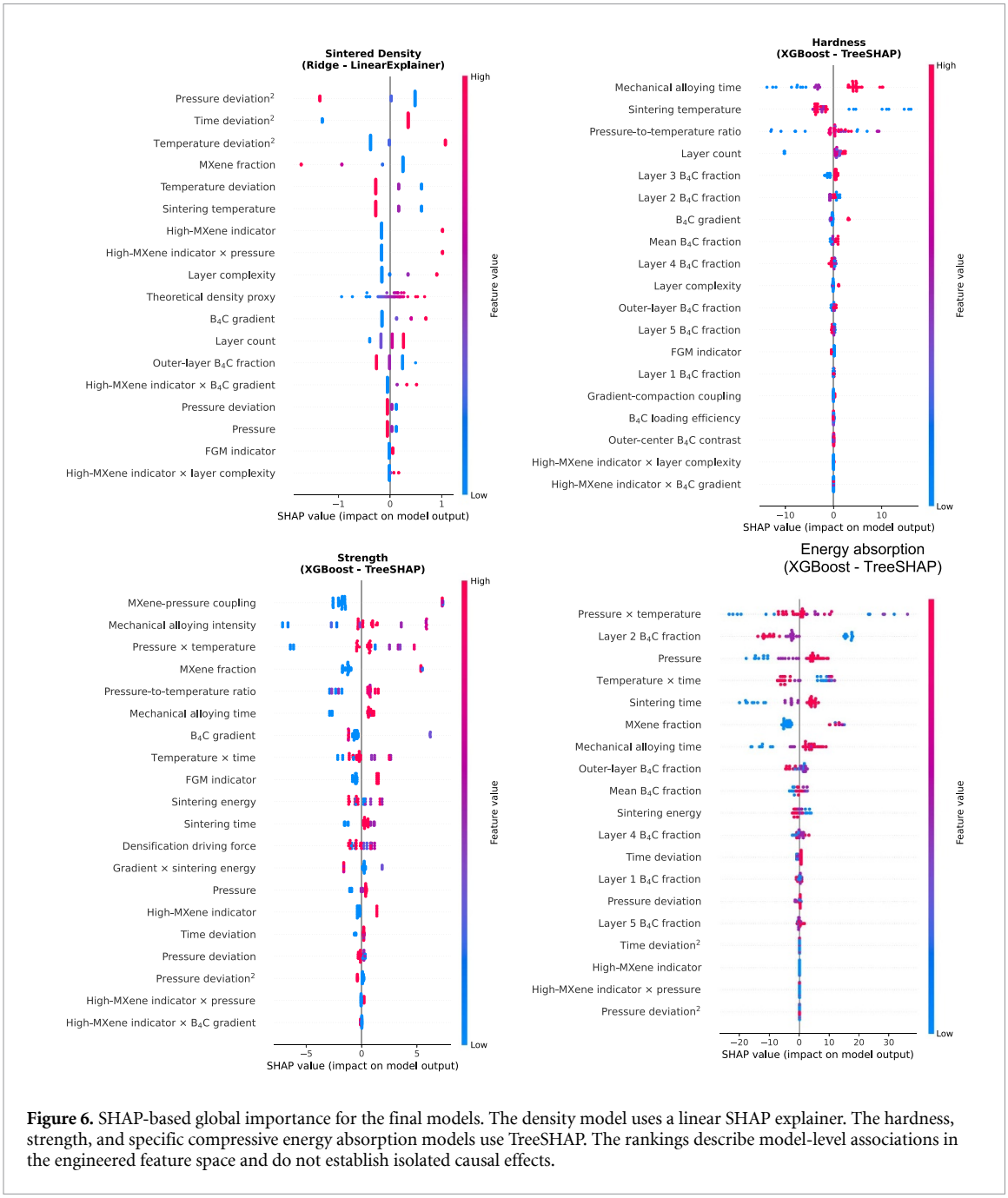


Figure 6. SHAP-based global importance for the final models. The density model uses a linear SHAP explainer. The hardness, strength, and specific compressive energy absorption models use TreeSHAP. The rankings describe model-level associations in the engineered feature space and do not establish isolated causal effects.

weighting where needed all contribute to the final package. The final configurations for density and specific compressive energy absorption retain weighting, whereas hardness and strength remain effectively unweighted, which also reinforces the point that different targets required different QC controls.

From an application perspective, the final configurations are preferable because they preserve stronger overall diagnostic balance under the final evaluation protocol.

4.3. Model behavior, reliability and materials consistency

Figure 6 shows both shared structure and differences among the final models. Interaction terms recur among the higher-contributing descriptors for strength and specific compressive energy absorption, especially pressure *x* temperature, temperature *x* time, MXene-pressure coupling, and related gradient-sensitive terms. Hardness retains a stronger imprint of mechanical alloying time, sintering temperature, layer count, and B_4C placement across layers. The density model, in contrast, places greater model weight on deviation-from-window terms, layer complexity, and densification-oriented surrogates than on one single raw processing variable.

The MXene-related descriptors in the fitted models should also be read together with the experimental characterization in figure 2. The x-ray diffraction and electron-microscopy results establish that the

nominal MXene additions used in the follow-up specimens correspond to a synthesized $Ti_3C_2T_x$-rich powder with a distinct layered morphology. This provides a physical basis for retaining MXene fraction, high-MXene regime indicators, MXene-pressure coupling, and high-MXene pressure or gradient terms in the pre-sintering descriptor space. These terms appear among the influential descriptors for sintered density, ultimate compressive strength, and specific compressive energy absorption. The agreement represents material traceability and model plausibility, not mechanistic validation. Because the MXene-bearing specimens belong to the concentrated follow-up branch, their MXene levels are partly coupled with pressure, mechanical-alloying history, and the revised two-stage sintering route. The SHAP contributions should therefore be interpreted as MXene-route-associated predictive patterns rather than isolated effects of the layered morphology.

Thermal exposure near 550–590 °C can alter the retained $Ti_3C_2T_x$ phase, particularly in the presence of residual oxygen. The nominal MXene fraction is therefore an initial composition descriptor, while the selected post-fabrication records provide local evidence without resolving the phase assemblage or interface chemistry for all specimens. MXene-related SHAP/PDP terms are consequently interpreted as route-associated patterns within the observed process window, and the unmeasured post-sintering state is treated as a source of response variability rather than as an isolated material effect.

These SHAP and PDP outputs describe directional model trends, threshold-like behavior, and feature couplings within the modeled process window. They are most defensible when interpreted as grouped descriptor responses because several engineered variables combine the same underlying pressure, temperature, time, layer and reinforcement information.

The VIF/condition audit supports interpreting SHAP rankings as conditional, grouped model contributions rather than isolated physical effects, while the correlated-feature PDPs remain restricted to the observed descriptor support.

These model based patterns are also consistent with the published experimental campaign in the related A7075/B_4C graded system, where mechanical alloying time, compaction pressure, and graded reinforcement placement were already observed to influence densification, hardness, and compressive response, and where higher outer layer B_4C loading was associated with increased porosity in the more ceramic rich zones. The SHAP and PDP trends reported here therefore remain aligned with the broader fabrication and characterization evidence base. A similar materials first logic has been emphasized in recent studies where cure history, thermal gradients, and process induced deformation were explicitly linked to residual stress development and final mechanical behavior, reinforcing the broader principle that variables describing the processing path must remain physically interpretable in predictive materials frameworks [27, 28].

For hardness, the persistent influence of mechanical alloying time and B_4C placement across layers is compatible with the earlier microstructural evidence of reinforcement redistribution and milling induced structural refinement reported for the same A7075/B_4C fabrication backbone.

For density and strength, the greater model contribution of pressure and densification oriented descriptors is likewise consistent with the previously observed sensitivity of compact integrity and porosity to compaction and thermal exposure in the underlying powder metallurgy campaign.

Taken together, the interpretability results support physically reasonable feature-response patterns within the studied process window. Here, SHAP and PDP analyses are used to clarify model response tendencies and to check whether the fitted pipelines remain consistent with the fabrication context.

The reliability and residual diagnostics in figure 5 and supplementary figure S1 provide an assessment of performance beyond score values. Sintered density and ultimate compressive strength align well with nominal 90% coverage, while hardness and specific compressive energy absorption remain close enough to the nominal level to support cautious use as retrospective diagnostics. Combined with the parity patterns in figure 4, the evidence supports model behavior that is plausible within the studied process window. The final models are useful as pre-sintering decision-support tools for experiment prioritization, and their interpretability outputs clarify directional patterns and feature couplings within the resulting pipelines without assigning isolated material effects to individual reinforcements.

4.4. Comparison with the literature

The literature positioning below concerns external methodological comparators. The published A7075/B_4C experimental campaign from which the present screening backbone is inherited is treated separately as the fabrication basis of this ML extension [25].

Table 5 positions the present study methodologically relative to related materials ML work.

As summarized in table 5, the closest external comparators address related but not identical materials design problems. Recent studies have used data-augmentation-assisted machine learning for alloy design

Table 5. Methodological positioning of the present study relative to related materials ML work.

Author Refrences	Material system	*n*	Input scope	Model(s)	Validation regime	Reported outputs and diagnostics	Match to present study	Direct numeric comparison
Maurya *et al* [29]	Al2024 with B_4C powder metallurgy composite	33	Inputs available before sintering for process and composition variables such as milling time, reinforcement ratio, and reinforcement size	ANN	Hold out split (25/8). No nested outer loop	Density, hardness, tensile strength, Pearson's *r*, adjusted R^2, and MAE	Partial	No
Manohar [30]	Al with B_4C powder metallurgy composite	n.r.	Inputs available before sintering for process and composition variables such as B_4C content, pressure, temperature, and time	RF, ANN	Validation reported. No nested outer loop stated	Tensile strength, R^2, MAE, RMSE, residual diagnostics, and feature importance	Partial	No
Varol *et al* [31]	Al–Cu–Mg with B_4C powder metallurgy composite	n.r.	Inputs available before sintering for process and composition variables such as reinforcement size, amount, milling time, and compaction pressure	ANN	Split between training and test data. No nested outer loop reported	Densification with density and porosity, plus MAPE and RMSE	Partial	No
Alam *et al* [32]	Sintered Al with SiC powder metallurgy composite	13†	Inputs available before sintering for process and composition variables	ANN, RSM	Small CCD based workflow for training, test, and validation. No nested CV	Hardness focused modeling with density context, plus R^2, RMSE, and MAPE	Partial	No
Elkatatny *et al* [33]	Al-4Cu with Al_2O_3 powder metallurgy composite	24	Inputs available before sintering for process, geometry, and composition variables such as Al_2O_3 wt.%, H/D ratio, and compaction pressure	LR, RT, RF, GPR, ANN	Random split between training and test data. No nested CV	Relative density, hardness, yield strength, compressive strength, fracture strain, R^2, and RMSE	Partial	No
Mudakavi and Adinarayanappa [34]	SS316-IN718 additively manufactured FGM	n.r.	Composition and location enriched inputs. Not a powder metallurgy system with inputs available before sintering	GB, RF, XGB, SVR	Five fold CV	Microhardness, R^2, MAE, and residual style diagnostics	Partial	No

under competing performance objectives [35], explainable AI and symbolic regression for descriptor-driven composite design [36], and predictive modeling for mechanically evolving composite responses under repeated deformation [37]. Materials Research Express studies have also combined experimental design or manufacturing variables with machine-learning prediction in settings such as friction-stir-processed aluminum composites, laser-welded aluminum alloy joints, wire-electrical-discharge machining and orthogonal-design-guided steel processing [12–15]. These studies confirm the growing role of interpretable and data-driven design strategies in advanced materials research, but they address different material systems, response definitions, and process boundaries from the problem considered here for A7075/B_4C/MXene FGMs before sintering.

The broader comparison separates related but different problem definitions. Some studies emphasize proof-of-concept prediction, some prioritize process history or cure-dependent response modeling, and others target descriptor discovery in high-dimensional design spaces. Work on uncertainty-aware prediction and SHAP analysis in cementitious materials further illustrates why prediction intervals and interpretation outputs should be reported as model diagnostics rather than as external guarantees [38]. The present work focuses on prediction before sintering for a graded powder metallurgy system with reliability, residual, and interpretability diagnostics.

Overall, the literature comparison places the present work within the subset of materials ML studies that combine inputs available before sintering with explicit validation control and diagnostic reporting. In that context, the contribution of the study is methodological as well as predictive for the present FGM powder metallurgy window.

4.5. Study scope and limitations

The study is based on a compact code-level dataset of $n = 43$. The final dataset combines the L27 screening backbone with a denser follow-up branch around the experimentally workable route, so route-history effects remain embedded in the retained code-level table.

MXene level, campaign membership, compaction pressure, mechanical-alloying history, and the revised 400 °C/2 h + 590 °C/3 h thermal route are partly coupled in the follow-up branch. Consequently, MXene-related model terms represent branch-associated predictive patterns and cannot be interpreted as isolated causal material effects. The characterization in figure 2 supports reinforcement traceability, but it does not separate MXene morphology from the accompanying route changes in the final composite responses.

The selected BSE-SEM images in Supplementary figure S6, together with the associated report-level EDS observations, extend post-fabrication traceability for the retained specimens. They provide local evidence but do not establish a specimen-wide, phase-resolved interface map; the nominal MXene descriptor therefore remains an initial composition variable.

Measurement-side variation also remains part of the retained responses: strength and specific compressive energy absorption are code-level averages over repeated compression tests, hardness is reported as a Leeb-derived HBW-equivalent quantity, and density is based on the representative-specimen Archimedes workflow described in the materials and methods section. Training-side QC weighting, where retained, was used as part of the modeling workflow and should not be read as a substitute for additional experimental replication.

The VIF and condition-number audit confirms strong collinearity created by the engineered descriptor expansion. Ridge regularization stabilizes the density prediction, while the interpretation uses grouped, conditional model quantities. The reported metrics and diagnostics define a retrospective operating envelope for the studied process window; transfer to a new campaign is addressed through independent validation and recalibration.

SHAP values are derived from full-data refits of the final pipelines, and the PDP panels are interpreted as descriptive summaries in engineered feature space. The empirical prediction intervals report coverage under the present resampling protocol. The models are specific to the studied material system, process window, graded architecture, specimen geometry, and target definitions. Transfer to different alloys, routes, or response definitions would require new calibration and independent validation.

Natural extensions of the present workflow include larger and more balanced datasets, broader route-history coverage, repeated independent fabrication runs, and prospective testing of the experiment-prioritization decisions suggested by the models.

5. Conclusions

This study shows that descriptors available before sintering can support property-oriented experiment prioritization for A7075/B_4C/MXene functionally graded aluminum matrix composites within the investigated powder metallurgy process window. Under leakage-controlled nested validation, ultimate compressive strength showed the strongest predictive behavior, HBW-equivalent hardness and specific compressive energy absorption retained useful predictive signals, and sintered density remained the most difficult response while still showing a positive signal under the selected Ridge-based configuration.

The main technical contribution is the combined workflow: code-level specimen tracking, pre-sintering descriptor construction, fold-confined preprocessing and model selection, empirical interval assessment, and SHAP/PDP diagnostics reported with their validation boundaries. The resulting model diagnostics support experiment prioritization within the studied process window and provide a traceable basis for selecting the next experiments before costly sintering and mechanical testing are completed.

The framework remains tied to the present A7075/B_4C/MXene powder-metallurgy system, specimen definition, target definitions, and process window. Transfer to different alloys, fabrication routes, geometries, or response definitions would require recalibration and independent validation.

Acknowledgment

The authors have no acknowledgements to declare.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Supplementary information available at https://doi.org/10.1088/2053-1591/aeb9b5/data1.

Code-level specimen dataset available at https://doi.org/10.1088/2053-1591/aeb9b5/data2.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics statement

Not applicable. This study involved materials fabrication and characterization only and did not involve human participants, animals, human tissue, or identifiable personal data.

Author contributions

Musa Yilmaz
Investigation (lead), Resources (lead), Validation (supporting)

Mürsel Ekrem 0000-0001-5324-7929
Conceptualization (lead), Methodology (equal), Project administration (lead), Resources (equal), Supervision (lead)

Hasmet Cagri Sezgen 0000-0002-6265-4259
Data curation (lead), Formal analysis (lead), Software (lead), Validation (equal), Visualization (equal), Writing – original draft (lead), Writing – review & editing (lead)

References

[1] Rajan T P D and Pai B C 2014 Developments in processing of functionally gradient metals and metal-ceramic composites: a review *Acta Metall. Sin. (Engl. Lett.)* **27** 825–38
[2] Wyatt B C, Nemani S K and Anasori B 2021 2D transition metal carbides (MXenes) in metal and ceramic matrix composites *Nano Convergence* **8** 16
[3] Wang Z J, Zheng Z and Fu M W 2024 Aluminum matrix composites: Structural design and microstructure evolution in the deformation process *J. Mater. Res. Technol.* **30** 3724–54
[4] Kumaravel A and Chakravarthi P 2025 Experimental and triboinformatics approach for wear and friction prediction of aa7075/b4c/tib2 composites at elevated temperatures using machine learning techniques *Mater. Res. Express* **12** 116509
[5] Chen C, Li B, Wang Y, Bian M, Kang X and Yang X 2025 Recent advances on aluminum-based B4C composites: performance, fabrication and applications *Materials* **18** 5469
[6] Akgümüş Gök D, Bayraktar C and Hoşkun M 2024 A review on processing, mechanical and wear properties of Al matrix composites reinforced with Al_2O_3, SiC, B_4C and MgO by powder metallurgy method *J. Mater. Res. Technol.* **31** 1132–50
[7] Li Z, Li Q, You J, Li F, Yu G, Zhang W and Liang Z 2025 Preparation of aluminum matrix composites reinforced with hybrid MAX-MXene particles for enhancing mechanical properties and tribological performance *J. Compos. Sci.* **9** 552
[8] Chandran N C B S, Thakur A, Nemani S K, Moharana N, Scott C P H, Kumar K C H, Kumar R and Anasori B 2025 MXene-derived TiB2 formation in B4C at high temperatures *Adv. Compos. Hybrid Mater.* **8** 308
[9] Xu J, Zhu X, Liu Q, Du S and Zhang Y 2024 Unlocking the strength: the predictions of silicon carbide fracture toughness revealed through data-driven approach *Mater. Res. Express* **11** 025601
[10] Babu K A and Jeyapaul R 2022 An investigation into the wear behaviour of a hybrid metal matrix composite under dry sliding conditions using Taguchi and ANOVA methods *J. Bio- Tribo-Corros.* **8** 15
[11] Daba R D, Sinha D K, Beri H, Kumar A and Yadav A 2024 Multi-response optimization of physical, mechanical and tribological properties of alumina and graphite-reinforced aluminium matrix composites *JMST Adv.* **6** 283–96
[12] Dai N, Li J, Qin H, He G, Li P, Wu Z and Wang S 2024 Process optimization of quenching and partitioning by machine learning aided with orthogonal experimental design *Mater. Res. Express* **11** 016519
[13] Mohanraj A and Kumaravel A 2026 Optimization of WEDM parameters for stir-cast Al6082–TiB2–Gr–Mg hybrid composite using taguchi and machine learning approaches *Mater. Res. Express* **13** 026507
[14] Adiga K, Herbert M A, Rao S S, Shettigar A K and Shrivathsa T V 2024 Development of machine learning regression models for the prediction of tensile strength of friction stir processed AA8090/SiC surface composites *Mater. Res. Express* **11** 076517
[15] Kumar H, Chakravarty S, Kuamr N and Kumar N 2025 Prediction of laser welding qualities of al alloys using regression and machine learning techniques *Mater. Res. Express* **12** 066501
[16] Cawley G C and Talbot N L C 2010 On over-fitting in model selection and subsequent selection bias in performance evaluation *J. Mach. Learn. Res.* **11** 2079–107
[17] Rosenblatt M, Tejavibulya L, Jiang R, Noble S and Scheinost D 2024 Data leakage inflates prediction performance in connectome-based machine learning models *Nat. Commun.* **15** 1829
[18] Varoquaux G 2018 Cross-validation failure: small sample sizes lead to large error bars *NeuroImage* **180** 68–77
[19] Varma S and Simon R 2006 Bias in error estimation when using cross-validation for model selection *BMC Bioinform.* **7** 91
[20] Arlot S and Celisse A 2010 A survey of cross-validation procedures for model selection *Stat. Surv.* **4** 40–79
[21] Wong T T and Yeh P Y 2020 Reliable accuracy estimates from k-fold cross validation *IEEE Trans. Knowl. Data Eng.* **32** 1586–94
[22] Xu P, Ji X, Li M and Lu W 2023 Small data machine learning in materials science *npj Comput. Mater.* **9** 42
[23] Zhang Y and Ling C 2018 A strategy to apply machine learning to small datasets in materials science *npj Comput. Mater.* **4** 25
[24] Zong B, Li J, Yuan T, Wang J and Yuan R 2025 Recent progress on machine learning with limited materials data: Using tools from data science and domain knowledge *J. Materiomics* **11** 100916
[25] Anonymous 2026 Anonymized prior work on functionally graded aluminum matrix materials by powder metallurgy *Anonymized for double-anonymous review* Details withheld for double-anonymous review
[26] Ahmedabadi P M 2024 Comparison of mathematical and supervised machine-learning models for ductile-to-brittle transition in bcc alloys *Mater. Res. Express* **11** 076506
[27] Kim D-H, Kim S-W and Kim T S 2026 Prediction of process-induced deformation in CFRP multispar flaps using sequentially coupled thermomechanical analysis considering fabric patterns and stacking sequences *Adv. Compos. Hybrid Mater.* **9** 169
[28] Lee I J, Je S, Choi J H, Lee H, Lee J and Kim S Y 2026 A cure-kinetics-based framework linking residual stress evolution to strength degradation in carbon/cyanate ester laminates *Adv. Compos. Hybrid Mater.* **9** 165
[29] Maurya A K, Narayana P L, Wang X-S and Reddy N S 2024 Machine learning modeling of the mechanical properties of Al2024-B4C composites *J. Powder Mater.* **31** 382–9
[30] Manohar G 2025 Machine learning approaches for tensile strength prediction in Al/B4C metal matrix composites: a comparative analysis of neural networks and ensemble methods *Int. J. Interact. Des. Manuf.* **19** 8267–83
[31] Varol T, Canakci A and Ozsahin S 2018 Modeling of the prediction of densification behavior of powder metallurgy al–cu–mg/b4c composites using artificial neural networks *Acta Metall. Sin. (Engl. Lett.)* **28** 182–95
[32] Alam M A, Ya H H, Azeem M, Hussain P B, bin Salit M S, Khan R, Arif S and Ansari A H 2020 Modelling and optimisation of hardness behaviour of sintered Al/SiC composites using RSM and ANN: a comparative study *J. Mater. Res. Technol.* **9** 14036–50
[33] Elkatatny S, Alsharekh M F, Alateyah A I, El-Sanabary S, Nassef A, Kamel M, Alawad M O, BaQais A, El-Garaihy W H and Kouta H 2023 Optimizing the powder metallurgy parameters to enhance the mechanical properties of Al-4Cu/Al_2O_3 composites using machine learning and response surface approaches *Appl. Sci.* **13** 7483
[34] Mudakavi D and Adinarayanappa S M 2026 Machine learning-based microhardness prediction and microstructural analysis of stainless steel 316-inconel 718 functionally graded material through metal additive manufacturing *Rapid Prototyping J.* **32** 142–61
[35] Bu W, Wang Y, Zhang Y, Hao J, Wang R, Hu Z and Liang X 2026 Data augmentation-assisted machine learning for accelerated discovery of high-strength lightweight alloys with enhanced solidification cracking resistance *Adv. Compos. Hybrid Mater.* **9** 40

[36] Li T, Li L, Xu K, Wang Y, Zhang C and Huang X 2026 Leveraging explainable ai in the design of carbon composites with high thermal conductivity for thermal management *Adv. Compos. Hybrid Mater.* **9** 156
[37] Lee D-K, Hur O, Kim E, Kang B-H, Kang S H, Min K and Park S-H 2025 Unveiling unexpected mechanical softening/stiffening in carbon nanotube composites under cyclic deformation: experiments and predictive modeling *Adv. Compos. Hybrid Mater.* **8** 194
[38] Zhou Z, Chakma J, Hoque Md A, Chakma V and Ahmed A 2025 Prediction of uhpc mechanical properties using optimized hybrid machine learning model with robust sensitivity and uncertainty analysis *Mater. Res. Express* **12** 085703