

## Article

# Imputation of Thermal and Magnetic Variables in Shape-Memory Alloys (Ni–Mn–Ga) Using Machine Learning Techniques with Cross-Validation and Multi Seed

Juan C. Buitrago Diaz <sup>1,\*</sup> , Edwin G. Castro Rodas <sup>1</sup> , Carolina Ortega-Portilla <sup>2,3</sup> , Juan E. Bedoya-Rodriguez <sup>4</sup> , Daniel Salazar <sup>4</sup> , Manuel G. Forero <sup>1</sup>  and Jeferson Fernando Piamba <sup>1</sup> 

<sup>1</sup> Semillero Lún, Facultad de Ciencias, Ingeniería e Innovación, Universidad de Ibagué, Ibagué 730002, Tolima, Colombia; 2420162005@estudiantesunibague.edu.co (E.G.C.R.); manuel.forero@unibague.edu.co (M.G.F.); jeferson.piamba@unibague.edu.co (J.F.P.)

<sup>2</sup> Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias, Epigmenio González 500 Fracc, San Pablo 76130, Querétaro, Mexico; iortega@posgrado.cidesi.edu.mx

<sup>3</sup> Additive Manufacturing Department, SECIHTI-Centro de Ingeniería y Desarrollo Industrial CIDESI, Querétaro 76125, Querétaro, Mexico

<sup>4</sup> BCMaterials, Basque Center for Materials, Applications and Nanostructures, EHU Science Park, Leioa 48940, Bizkaia, Spain; juanesteban.bedoya@bcmaterials.net (J.E.B.-R.); daniel.salazar@bcmaterials.net (D.S.)

\* Correspondence: camijuan\_321@hotmail.com

## Abstract

Magnetic shape memory alloys based on the Ni–Mn–Ga system are of strategic interest for aerospace and robotics applications due to their ability to respond to both thermal and magnetic stimuli. However, the NASA Shape Memory Materials Database a key resource for the community exhibits significant gaps in functional parameters, with up to 93.7% of records missing critical properties such as the Curie temperature, and over 88% lacking complete magnetic data. To address this limitation, this study proposes a data imputation strategy based on a stacking ensemble comprising twelve machine learning models (LGBM, XGBoost, CatBoost, GradientBoosting, RandomForest, MLP, BayesianRidge, KNN, SVR, GPR, MICE, and AutoEncoder), optimized via Optuna and evaluated using ten random seeds with 10 repetitions each. The approach was applied to reconstruct missing entries in NASA's database. For heat treatment 1, the method achieved coefficients of determination ( $R^2$ ) of 0.95 for duration (h) and 0.88 for temperature ( $^{\circ}\text{C}$ ), respectively. For the phase transformation temperatures ( $M_f$ ,  $M_s$ ,  $A_s$ , and  $A_f$ ), the method yielded  $R^2$  values of 0.83, 0.82, 0.79, and 0.80, respectively. Magnetic properties saturation magnetization and maximum magnetic field were imputed with an  $R^2$  of 0.92. In contrast, the Curie temperature exhibited limited predictive performance ( $R^2 = 0.15\text{--}0.35$ ), primarily due to insufficient data availability. Overall, the proposed methodology integrates machine learning based imputation with physically supported constraints, providing a viable alternative to enhance the completeness and utility of materials databases.

**Keywords:** phase temperatures; heat treatment; machine learning; cross-validation; imputation; stacking



Academic Editors: Peijiang Liu, Jianping Peng and Xiaodong Jian

Received: 6 June 2026

Revised: 23 July 2026

Accepted: 27 July 2026

Published: 19 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 \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Shape memory alloys (SMAs) have a strong ability to recover large inelastic strains, often introduced by martensitic reorientation, upon an appropriate thermal stimulus, a capability that underpins applications ranging from self-expanding medical devices to aerospace

actuators (e.g., satellite control mechanisms) [1]. Among these materials, Ni–Mn–Ga system is a prototypical ferromagnetic SMA because its functional response can be driven not only by temperature but also by moderate magnetic fields ( $\leq 1$  T), allowing recoverable strains of up to 12% in oriented single crystals [2,3]. This behavior remains stable even under very high cycle counts: cycling 29 tests have shown that its ferromagnetism stays nearly unchanged after 106 loading cycles, and in some configurations far beyond, with limited degradation of functional response, supporting its potential for long-term cyclic actuation [1,4,5]. Nevertheless, for implantable use, long-term bioperformance, including corrosion and ion-release behavior, requires careful validation and, in many concepts, protective surface strategies are considered essential.

Advances in composite materials increasingly rely on dispersing magnetically active Ni–Mn–Ga single-crystal particles within soft polymer matrices to enable cooperative twin-variant reorientation under an applied magnetic field. In a representative approach, Ni–Mn–Ga particles are embedded at 30 vol.% in a low-modulus silicone rubber (e.g., ELASTOSIL M 4400; elastic stiffness  $\approx 0.5$  MPa, i.e., below 1 MPa) and cured under a magnetic field to promote particle alignment into chain-like structures. Under these conditions, the composite can exhibit a fully reversible magnetic-field-induced rubber-like behavior, characterized by macroscopic magnetic-field-induced strain (MFIS) of approximately 4.0% elongation (and 2.0% transverse contraction) under a magnetic field of 0.7 T [5].

In aerospace applications, SMAs are being investigated for adaptive and weight-efficient actuation architectures, including variable-geometry propulsion components such as adaptive exhaust nozzles actuated by SMAs to regulate nozzle area and improve operational flexibility. In parallel, SMA-based morphing concepts continue to be explored for aerodynamic shape control; for example, AFRL has reported flight experiments demonstrating a variable-camber compliant wing concept capable of smoothly altering wing camber across operating conditions [6]. Beyond near-ambient operation, Ni–Mn–Ga-based high-temperature SMAs have also been proposed for actuation in the 400–500 °C regime, with transformation temperatures reported up to 500 °C and functional stability assessed over hundreds of thermal cycles after appropriate thermomechanical conditioning [7]. Finally, in selected Ni–Mn–Ga compositions where magnetic and structural transitions are coupled, the same magnetostructural physics can also support magnetocaloric responses in the vicinity of the transformation [5,8].

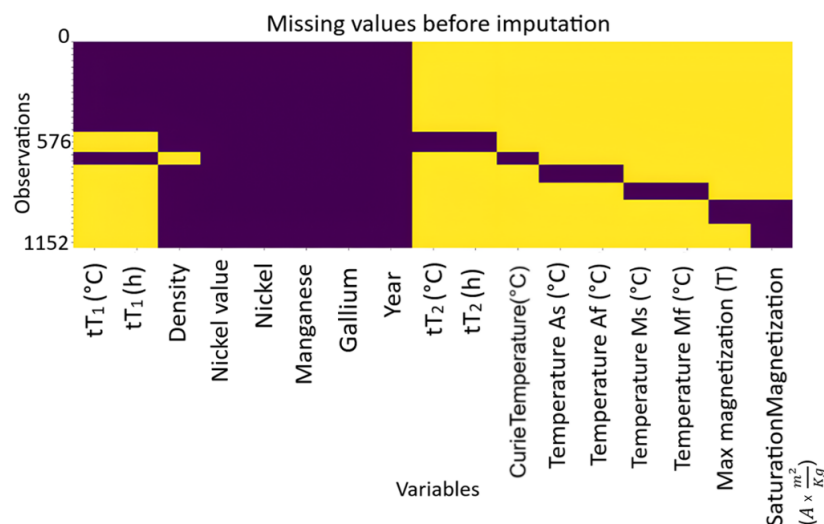
#### *Microstructure and Composite Design*

Among the martensitic structures, the non-modulated (NM) phase and the modulated monoclinic structures 10 M and 14 M are the most commonly reported. The dominant mechanism behind magnetic shape memory (MSM) is the motion of the twin boundary in this martensitic phase. Detailed studies on single crystals have revealed the complex coexistence of multiple domains (including austenite, martensite variants, and twins), with experimental observations reporting a five-domain propagation interface under a thermal gradient [9–11]. These works show that such complex microstructural patterns (such as four- and five-domain configurations) are nearly perfectly compatible, meaning that they generate only a small elastic energy.

This highly mobile twinned microstructure is crucial for damping: in hybrid NiMnGa composites embedded in a polymer matrix (such as soft epoxy), the hysteretic motion of twin boundaries provides a high Relative Damping Capacity (RDC) to the material. For example, a MSM–Epoxy composite achieved an RDC of 3.5 GPa, a value much higher than that of pure NiMnGa single crystals (RDC = 0.2 GPa) [9–12].

There is a significant challenge in data management, as the documentation of this alloy and many others is scattered throughout the scientific literature. NASA's Shape Memory

Materials Database compiled information from multiple articles, gathering a substantial number of parameters and valuable data on these alloys, constituting a critical resource for the scientific community. However, it presents notable shortcomings: 88% of the records lack complete magnetic hysteresis curves, and 92% omit fundamental parameters such as transition temperatures [13]. These gaps introduce uncertainties in predictive models, limiting their practical usefulness (Figure 1 and Table 1).



**Figure 1.** Missing data visualization for the NASA Ni–Mn–Ga database. Yellow cells indicate missing values, while purple cells represent complete data. The figure illustrates the high proportion of missing data for thermal and magnetic properties, particularly for heat treatment 2, phase transformation temperatures, and Curie temperature.

**Table 1.** Percentage and count of null data.

| Parameters                                               | Amount of Null Data | Percentage of Null Data % |
|----------------------------------------------------------|---------------------|---------------------------|
| Nickel ( <i>at%</i> )                                    | 0                   | 0                         |
| Manganese ( <i>at%</i> )                                 | 0                   | 0                         |
| Gallium ( <i>at%</i> )                                   | 0                   | 0                         |
| Thermal treatment 1 ( $T_1$ )                            | 590                 | 49.83                     |
| Thermal treatment 1 ( $tT_1$ )                           | 590                 | 49.83                     |
| Thermal treatment 2 ( $T_2$ )                            | 1070                | 90.37                     |
| Thermal treatment 2 ( $tT_2$ )                           | 1070                | 90.37                     |
| Density ( $\text{g}/\text{cm}^3$ )                       | 75                  | 6.34                      |
| Temperature $A_s$ ( $^{\circ}\text{C}$ )                 | 1081                | 91.30                     |
| Temperature $A_f$ ( $^{\circ}\text{C}$ )                 | 1081                | 91.30                     |
| Temperature $M_f$ ( $^{\circ}\text{C}$ )                 | 1088                | 91.89                     |
| Temperature $M_s$ ( $^{\circ}\text{C}$ )                 | 1088                | 91.89                     |
| Max magnetization ( $T$ )                                | 1044                | 88.17                     |
| Saturation Magnetization<br>( $A \times m^2/\text{kg}$ ) | 1044                | 88.17                     |
| Curie temperature ( $^{\circ}\text{C}$ )                 | 1109                | 93.66                     |

In the face of this challenge, traditional imputation methods (e.g., mean/median substitution) are insufficient to address the multivariable complexity of materials data. In contrast, machine learning techniques such as random forest, k-NN, and gradient boosting algorithms (LightGBM, XGBoost, CatBoost) have demonstrated superior performance by capturing nonlinear relationships among composition, microstructure, and properties [14,15]. Recent studies show that imputation quality directly affects predictive

model performance, making it critical to select methods that preserve physicochemical correlations [16–18].

Although machine learning has been successfully applied to the design of superalloys [19–22], high entropy alloys [23] complex magnetic alloys [24–26], and in the study of martensitic transformations that exhibit shape memory effects, such as those analyzed in the present articles [27–31] its use for imputing data in NASA's database remains incipient. This work proposes a hybrid approach that combines different machine learning techniques with models such as LightGBM, XGBoost, RandomForest, CatBoost, MLP, BayesianRidge, KNN, SVR, GPR, MICE, and AutoEncoder (optimized using Optuna), along with variable-selective models (XGBoost, MLP, BayesianRidge), selecting the best model for each variable. This approach captures specific relationships among composition, density, and heat treatments, among other parameters, improving dataset integrity for predicting missing values. It is important to note that H. Abedi [32] used NASA's NiTiHf alloy database in his study due to its low variability across publications and its high data quality for predicting the behavior of this alloy, demonstrating that reliable predictions can be achieved using machine learning.

The models are evaluated using standard metrics such as mean squared error (MSE [33], Equation (1)), root mean squared error (RMSE, Equation (2)), and the coefficient of determination ( $R^2$ , Equation (3)).

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y}_i)^2 \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y}_i)^2} \quad (2)$$

$Y_i$  is the observed value at instance  $i$ ,  $\bar{Y}_i$  is the model prediction, and  $n$  is the total number of instances.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (3)$$

where  $y_i$  represents the actual (observed) value at point  $i$ ,  $\bar{y}_i$  is the value predicted by the model, and  $\bar{y}_i$  is the measure of the observed values.

## 2. Materials and Methods

### 2.1. Materials

For this work, the NASA Shape Memory Materials Database, available at [13], was used. This database compiles historical research data on nickel, manganese and gallium (Ni–Mn–Ga) alloys published in peer-reviewed scientific articles. The dataset comprises a total of 1184 samples extracted from over 50 publications spanning several decades. It includes information on stoichiometric compositions (atomic percentages of Ni, Mn, and Ga), alloy density, and related physical parameters organized as follows: heat treatment 1 (temperature  $T_1$  in °C and holding time  $tT_1$  in hours), heat treatment 2 (temperature  $T_2$  in °C and holding time  $tT_2$  in hours), martensitic and austenitic transformation temperatures (Ms, Mf, As, and Af) in °C, maximum magnetic field in teslas (MaxMagnetizationField), saturation magnetization in emu/g (SaturationMagnetization), and Curie temperature in °C (CurieTemp).

Data processing and model development were performed using Python 3.9 (Python Software Foundation, Wilmington, DE, USA). The machine learning workflow was implemented using NumPy v1.26, pandas v2.2, scikit-learn v1.5, XGBoost v2.1, LightGBM v4.5, CatBoost v1.2, Optuna v4.0, and PyTorch v2.4 (Meta Platforms, Inc., Menlo Park, CA, USA).

Computations were carried out on a Dell Precision 5860 workstation (Dell Technologies Inc., Round Rock, TX, USA) equipped with an Intel Xeon W7-2495X processor (Intel Corporation, Santa Clara, CA, USA) and dual NVIDIA RTX 5000 Ada Generation GPUs (NVIDIA Corporation, Santa Clara, CA, USA). Data were stored in CSV format and subsequently organized using Microsoft Excel 2024 (Microsoft Corporation, Redmond, WA, USA).

## 2.2. Methods

Model training was optimized to minimize the mean squared error (Equation (1)), as this metric penalizes larger errors and is desirable for precise imputations where significant deviations could affect physical interpretations. To optimize each model's hyperparameters, the Optuna library was used with its Tree-structured Parzen Estimator (TPE) approach, which is more effective than random or grid searches for high-dimensional and complex spaces. The TPE algorithm iteratively constructs a probabilistic model of the objective function and selects hyperparameter configurations that are likely to improve performance, balancing exploration and exploitation.

A total of 300 trials were used to explore the hyperparameter space for each model and each target variable. The search ranges for each hyperparameter were defined based on prior knowledge and preliminary experiments (see Table 2). Sequential execution was employed to avoid stability issues in global variables, at the cost of increased computation time. For each trial, the model was evaluated using 5-fold cross-validation on the training set, and the hyperparameter combination yielding the lowest average MSE was selected. This process was repeated for each of the 10 random seeds, resulting in 3000 optimization runs per target variable (300 trials  $\times$  10 seeds).

The seven tables were concatenated using the Pandas library, as shown in Figure 1, and an identification column was added to enable constant tracking of the real data before being imputed in the complete table.

In the preprocessing stage, all column values were converted to numerical format using the `pd.to_numeric` method, assigning *NaN* to non-numeric entries to prevent errors during model training. For cleaning the heat treatment one parameter, the interquartile range (IQR) was used to trim outliers beyond approximately 2.7 standard deviations in a normal distribution, considering 99.3% of the data under Gaussian assumptions, using Equation (5) derived from Equation (4).

$$IQR = Q_3 - Q_1 \quad (4)$$

$$[Q_1 \leq 1.5 \times IQR], [Q_3 \geq 1.5 * IQR] \quad (5)$$

The data for the variables nickel, manganese, gallium, and density were used as predictors because of their physical relationship with the temperatures of heat treatment 1. Additionally, they were chosen for their direct correlation between degrees Celsius and hours of the same variable, since both describe aspects of the same thermal process. For training, the data were split into three sets: 70% for training, 15% for validation, and 15% for model validation tests, ensuring a comprehensive evaluation of the models.

To improve the performance and stability of the models, 10 specific random seeds were used. This approach ensures the reproducibility of the results and introduces controlled variability between repetitions, allowing simulation of different data partitions and evaluation of model robustness. Additionally, each seed was executed 10 times, reducing the bias associated with a single random split and providing a more reliable estimate of the final performance.

Model training was optimized to minimize the mean squared error (Equation (1)), since this metric penalizes larger errors and is desirable for precise imputations, where significant deviations could affect physical interpretations. To optimize each model, the Optuna library was used, which, with its TPE (Tree-structured Parzen Estimator) approach, is more effective than random or grid searches for complex spaces. A total of 300 trials were used to explore the hyperparameter space, and sequential execution was employed to avoid stability issues in global variables, at the cost of increased computation time.

The models tested were the following: boosted trees (LGBM, XGBoost, CatBoost, GradientBoosting), which are suitable for tabular data with nonlinear relationships and noise, common in physical experiments [34]. RandomForest is robust and effective for capturing interactions between variables without assuming specific distributions [35]. MLP is capable of modeling complex and nonlinear relationships, useful when temperatures depend on hidden patterns [36]. BayesianRidge is typically appropriate for linear relationships with uncertainty [37]. KNN, based on local similarity, is useful when the data contains natural clusters [38]. SVR is effective for nonlinear regression with kernels and is robust to outliers [39]. GPR is ideal for modeling smooth and continuous relationships with uncertainty, common in physical data [40]. MICE is a standard statistical method for multivariate imputation [41], and AutoEncoder is an approach based on neural networks to capture latent structures in missing data.

Multiple models were implemented to ensure that the final approach does not depend on a single technique, allowing the identification of the best model tuned with Optuna for the specific characteristics of each dataset [42]. In addition, each model used a specific set of hyperparameters within a defined tuning range, as shown in Table 2.

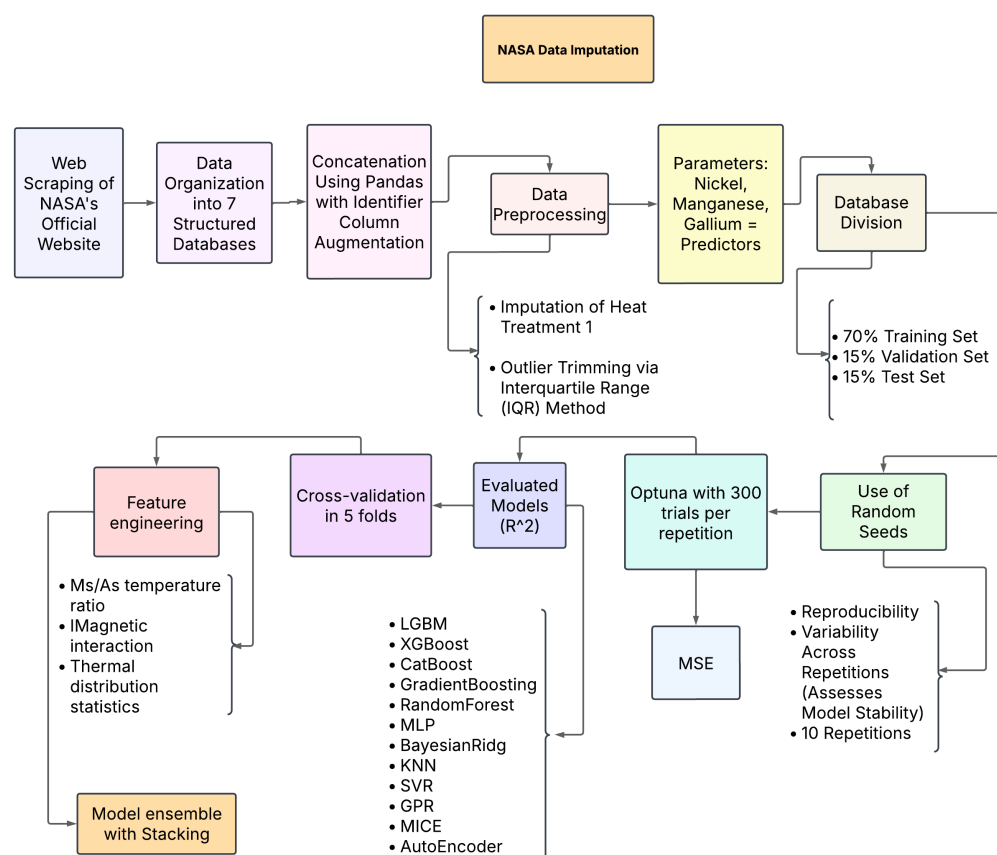
**Table 2.** Hyperparameters for the different models.

| Models       | Hyperparameter      | Range                                           |
|--------------|---------------------|-------------------------------------------------|
| LGBM/XGBOOST | nEstimators         | 100 to 200                                      |
|              | MaxDepth            | 2 to 15                                         |
|              | Learning rate       | $1 \times 10^{-6}$ to 0.3 (log)                 |
| GPR          | Kernel length scale | 0.1 to 10                                       |
|              | Alpha               | $1 \times 10^{-10}$ to $1 \times 10^{-2}$ (log) |
| AutoEncoder  | EncodingDim         | 2 to 5                                          |
|              | Epochs              | 50 to 150                                       |

For each repetition of the 300 trials, the model with the lowest MSE in cross-validation was selected in order to maximize performance in each partition. A pipeline was trained with an initial imputation step (CustomImputer), filling the missing values in the predictors with medians, only for the case of density, which was the only predictor with missing data. After this, a standard scaling (StandardScaler) was applied to normalize the features to mean 0 and variance 1. Subsequently, feature engineering (advancedFeatureEngineer) was applied to identify additional relationships among variables. The temperature ratio  $M_s/A_s$  (martensitic transformation), which has a direct correlation with martensitic phase stability, was used. In addition, the parameters of maximum magnetic field and maximum magnetic saturation were unified into a single feature through their product [43]. Finally, a thermal distribution metric was implemented in which minimum temperatures were subtracted from maximum temperatures to evaluate thermal stability, thermal fatigue, and the martensite/austenite transformation window [44]. Afterward, a robust scaling was applied to adjust the transformed features to median 0 and interquartile range.

Other metrics measured were  $R^2$ , mentioned in Equation (3), and RMSE from Equation (2), where  $R^2$  indicates the proportion of explained variance and reflects how

good the model is, with a range from 0 to 1, where values close to 1 represent a perfect model, while RMSE facilitates the interpretation of the results. The complete methodology of the described process is summarized in Figure 2.



**Figure 2.** Flowchart of the complete imputation methodology. The process comprises four main stages: (1) data preprocessing and cleaning, (2) feature engineering including physical constraints, (3) model training and optimization with Optuna, and (4) stacking ensemble and uncertainty quantification.

For the phase temperatures, additional advanced features [45] were used, such as nonlinear transformations (Equations (6)–(9)), thermodynamic relations (Equations (10) and (11)), and critical gradients ((12) and (13)). This process generated 23 new variables in order to better model the behavior of the alloy and improve the training of the different models. The target metric remains the mean squared error, in addition, a physical penalty for violation of thermal ordering is added (Equation (14)). A regularized stacking approach is used with CatBoost as the meta-model.

The advanced feature engineering strategy was designed to encode relevant physical relationships into the input space, thereby guiding the machine learning models toward physically plausible predictions. Specifically, three categories of features were generated:

Nonlinear transformations: Logarithmic, square root, square, and cubic transformations of the original compositional and thermal variables were included to capture potential nonlinear dependencies that are common in phase transformation phenomena

$$\log_X = \ln(1 + |X|) \quad (6)$$

$$\text{sqrt}_X = |X| \quad (7)$$

$$square_X = X^2 \quad (8)$$

$$cubic_X = X^3 \quad (9)$$

Thermodynamic relationships: Temperature differences ( $\Delta T_{ij}$ ) and ratios ( $R_{ij}$ ) were computed for all pairs among the four phase transformation temperatures ( $M_s$ ,  $M_f$ ,  $A_s$ ,  $A_f$ ). These features encode the expected thermodynamic hierarchy ( $T_{Mf} < T_{Ms} < T_{As} < T_{Af}$ ) and provide information about the transformation hysteresis width, which is known to correlate with composition and processing conditions in Ni–Mn–Ga alloys

$$\Delta T_{ij} = T_i - T_j \quad (10)$$

For temperature pairs  $(T_i, T_j)$  where  $i, j \in \{M_s, A_s, A_f, M_f\}$ .

$$R_{ij} = \frac{T_i}{T_j + \epsilon} \quad (11)$$

With  $\epsilon = 10^{-6}$  to avoid division by 0 and only for  $i < j$  to avoid redundancies.

Critical gradients: The austenite completion range ( $\Delta T(A_s - A_f)$ ) and the martensitic formation range ( $\Delta T(M_f - M_s)$ ) were computed as measures of thermal stability and transformation window.

$$\nabla T_{As-Af} = T_{Af} - T_{As} \quad (12)$$

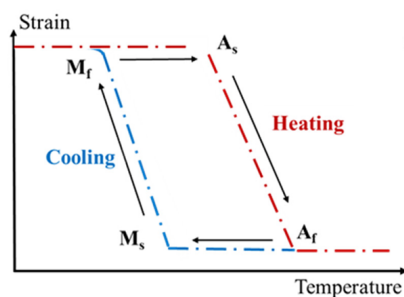
With Equation (12) the austenite completion range is measured.

$$\nabla T_{Mf-Ms} = T_{Ms} - T_{Mf} \quad (13)$$

For magnetic properties, the product of maximum magnetic field and saturation magnetization was included as a composite feature, based on the physical relationship between these quantities in ferromagnetic materials.

To ensure that the imputed values respect the fundamental thermodynamic ordering of phase transformation temperatures, a physical penalty term was added to the loss function during training Equation (13). This penalty imposes a cost of  $10^4$  whenever the predicted values violate the expected hierarchy  $T_{Mf} < T_{Ms} < T_{As} < T_{Af}$ . This approach, known as physics-informed regularization, has been successfully applied in recent ML studies for shape memory alloys [15,29] and ensures that the imputed values are not only statistically plausible but also physically consistent (see Figure 3).

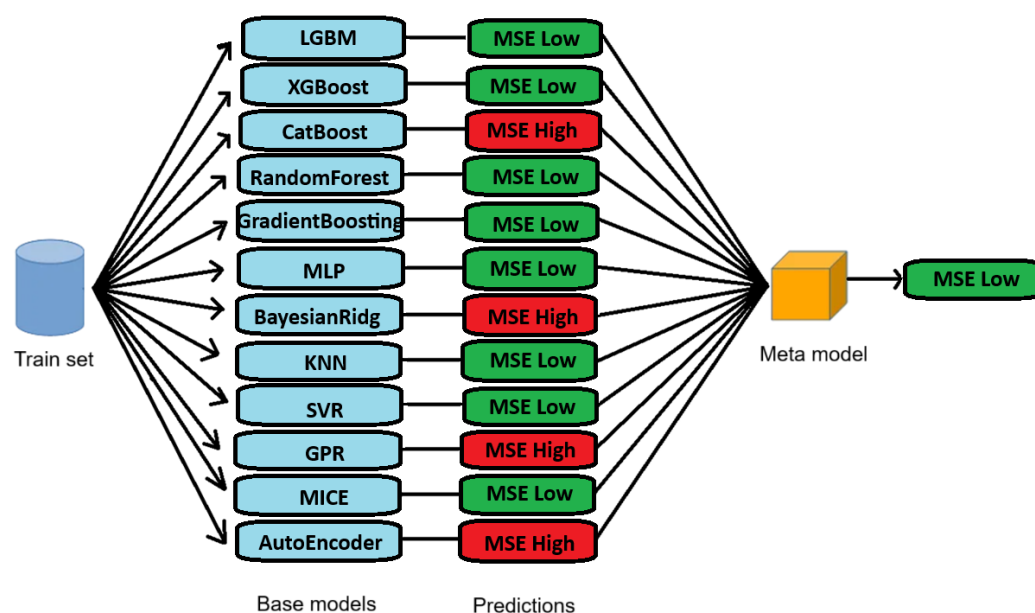
$$MSE_{\text{mod}} = MSE + 10^4 \cdot \mathbb{I}(T_{Mf} \geq T_{Ms} \vee T_{Ms} \geq T_{As} \vee T_{As} \geq T_{Af}) \quad (14)$$



**Figure 3.** Schematic representation of thermal hysteresis in shape memory alloys, showing the characteristic temperature hierarchy  $M_f < M_s < A_s < A_f$ . This ordering was enforced as a physical constraint during the imputation process (adapted from Hmede [12]).

For each target variable, the final imputed value was determined through a two-stage procedure. In the first stage, the 12 base models were trained independently on the same training partition, and their predictions on the validation set were recorded. These predictions, along with the original features, were used to train a meta model (CatBoost) in a 5-fold cross-validation setting within the stacking framework. The meta-model learns to combine the base model predictions optimally, effectively weighting each base model according to its performance on the validation data.

In the second stage, the trained stacking ensemble was applied to the test set (and subsequently to the full dataset) to generate the final imputed values. The use of a meta-model, rather than simple averaging, allows the ensemble to adapt to the specific characteristics of each target variable and to leverage the complementary strengths of the base models (see Figure 4). This approach has been shown to reduce both bias and variance compared to individual models or unweighted ensembles [46,47].



**Figure 4.** Architecture of the stacking ensemble used for imputation. Predictions from 12 base models (LGBM, XGBoost, CatBoost, GradientBoosting, RandomForest, MLP, BayesianRidge, KNN, SVR, GPR, MICE, AutoEncoder) are combined using a CatBoost meta model with 5-fold cross-validation.

### 2.3. Uncertainty Quantification and Statistical Significance

To assess the reliability of the imputed values and to compare the performance of the stacking ensemble against individual models, we conducted a comprehensive uncertainty analysis. For each target variable and each seed, we computed the following:

To assess the robustness of the stacking ensemble, uncertainty in the performance metrics was quantified using bootstrap resampling. Specifically, 1000 bootstrap resamples of the test predictions were generated for the stacking ensemble, from which the distributions of the MSE and coefficient of determination ( $R^2$ ) were obtained. The corresponding 95% confidence intervals were computed from the bootstrap percentile estimates.

The statistical significance of the performance difference between the stacking ensemble and the corresponding best observed individual model (BOP) was evaluated using a paired  $t$ -test. A significance level of  $\alpha = 0.05$  was adopted, with  $p < 0.05$  indicating statistically significant improvements.

The results are summarized in Table 3. For each target variable, the table reports the best observed performance achieved by an individual model (BOP), the average performance of the stacking ensemble, the associated 95% bootstrap confidence intervals for

the stacking metrics, and the corresponding  $p$ -values. This analysis complements the deterministic performance metrics by providing an estimate of the uncertainty associated with the stacking ensemble predictions.

**Table 3.** Performance comparison between the stacking ensemble and the best individual model, including uncertainty estimates and statistical significance.

| Variable                 | Model    | MSE (Mean) | 95% CI MSE       | $R^2$ (Mean) | 95% CI $R^2$   | $p$ -Value |
|--------------------------|----------|------------|------------------|--------------|----------------|------------|
| $tT_1$ (h)               | BOP CB   | 68.04      | –                | 0.973        | –              | –          |
|                          | Stacking | 66.21      | [57.3, 74.8]     | 0.975        | [0.968, 0.982] | 0.031      |
| $T_1$ (°C)               | BOP XGB  | 261.07     | –                | 0.918        | –              | –          |
|                          | Stacking | 251.83     | [224.6, 281.9]   | 0.922        | [0.903, 0.939] | 0.044      |
| $tT_2$ (h)               | BOP LGBM | 75.02      | –                | 0.911        | –              | –          |
|                          | Stacking | 73.24      | [65.6, 82.8]     | 0.923        | [0.90, 0.94]   | 0.063      |
| $T_2$ (°C)               | BOP XGB  | 293.05     | –                | 0.88         | –              | –          |
|                          | Stacking | 284.64     | [267.1, 298.9]   | 0.89         | [0.85, 0.91]   | 0.069      |
| $M_s$ (°C)               | BOP CB   | 1485.51    | –                | 0.764        | –              | –          |
|                          | Stacking | 1457.34    | [1294.5, 1608.8] | 0.769        | [0.738, 0.795] | 0.118      |
| $M_f$ (°C)               | BOP GB   | 718.22     | –                | 0.869        | –              | –          |
|                          | Stacking | 699.84     | [624.2, 770.9]   | 0.874        | [0.851, 0.894] | 0.027      |
| $A_s$ (°C)               | BOP GB   | 2004.34    | –                | 0.730        | –              | –          |
|                          | Stacking | 1956.71    | [1779.8, 2128.6] | 0.736        | [0.704, 0.763] | 0.081      |
| $A_f$ (°C)               | BOP GB   | 2264.25    | –                | 0.765        | –              | –          |
|                          | Stacking | 2207.18    | [2031.4, 2375.2] | 0.772        | [0.744, 0.796] | 0.039      |
| Max. Magnetization (T)   | BOP CB   | 0.190      | –                | 0.926        | –              | –          |
|                          | Stacking | 0.182      | [0.154, 0.206]   | 0.930        | [0.913, 0.944] | 0.019      |
| Saturation Magnetization | BOP CB   | 588.34     | –                | 0.872        | –              | –          |
|                          | Stacking | 565.71     | [519.6, 606.4]   | 0.878        | [0.853, 0.899] | 0.034      |
| Curie Temperature (°C)   | BOP CB   | 1471.62    | –                | 0.148        | –              | –          |
|                          | Stacking | 1459.37    | [1238.5, 1696.2] | 0.152        | [0.081, 0.226] | 0.337      |

### 3. Results

After applying the imputation process to the variables associated with Heat Treatment 1 (temperature1Hr and temperature1C), using a machine-learning-based approach with multiple models optimized through cross-validation and different random seeds, the metrics used to assess imputation performance are the Mean Squared Error (MSE), the Root Mean Squared Error (RMSE), and the Coefficient of Determination ( $R^2$ ). These were, calculated for both the test and validation sets for each seed [48].

Imputation of time of Heat Treatment 1, Table 4 summarizes the performance metrics for the imputation of the variable  $tT_1$ , which represents the duration in hours of Heat Treatment 1. Ten random seeds were used to divide the data into training, validation, and test sets, ensuring the robustness of the predictions against variations in data partitioning.

**Table 4.** Holding time of heat treatment 1 (h) per seed (validation and test).

| Model    | DataSplit        | MSE (Mean) | RMSE (Mean) | $R^2$ (Mean) |
|----------|------------------|------------|-------------|--------------|
| LGBM     | Validation_seed0 | 137.6982   | 11.7345     | 0.9338       |
|          | Test_seed0       | 287.4419   | 16.9541     | 0.8870       |
| XGBoost  | Validation_seed1 | 221.5401   | 14.8842     | 0.9139       |
|          | Test_seed1       | 110.3784   | 10.5061     | 0.9540       |
| LGBM     | Validation_seed2 | 163.1519   | 12.7731     | 0.9268       |
|          | Test_seed2       | 384.2577   | 19.6025     | 0.8667       |
| CatBoost | Validation_seed3 | 159.1134   | 12.6140     | 0.9293       |
|          | Test_seed3       | 68.0433    | 8.2488      | 0.9729       |
| XGBoost  | Validation_seed4 | 95.3633    | 9.7654      | 0.9677       |
|          | Test_seed4       | 86.1908    | 9.2839      | 0.9621       |
| XGBoost  | Validation_seed5 | 229.5430   | 15.1507     | 0.9011       |
|          | Test_seed5       | 115.6959   | 10.7562     | 0.9508       |
| CatBoost | Validation_seed6 | 292.0734   | 17.0902     | 0.9110       |
|          | Test_seed6       | 90.2406    | 9.4995      | 0.9611       |
| XGBoost  | Validation_seed7 | 103.4382   | 10.1705     | 0.9628       |
|          | Test_seed7       | 182.3457   | 13.5035     | 0.9318       |
| LGBM     | Validation_seed8 | 354.6331   | 18.8317     | 0.8739       |
|          | Test_seed8       | 425.9750   | 20.6392     | 0.8269       |
| LGBM     | Validation_seed9 | 347.4753   | 18.6407     | 0.8766       |
|          | Test_seed9       | 180.3710   | 13.4302     | 0.9128       |

The results show that the CatBoost model achieved the best performance on the test set for seed 3, with an MSE of 68.0433, an RMSE of 8.2488, and an  $R^2$  of 0.9729, indicating high accuracy in prediction the missing values. XGBoost also showed consistent results, with an average  $R^2$  above 0.95 across several seeds (for example, Testseed1:  $R^2 = 0.9540$ ; Testseed4:  $R^2 = 0.9621$ ). In contrast, LGBM exhibited greater variability, with  $R^2$  ranging from 0.8269 (Testseed8) to 0.9128 (Testseed9). In the validation set, XGBoost again stood out with an  $R^2$  of 0.9677 for seed 4, while LGBM and CatBoost maintained  $R^2$  values above 0.87 in most cases, as shown in Table 4.

Table 5 details the metrics for imputing the heat treatment 1 temperature ( $^{\circ}\text{C}$ ). As in the previous case, multiple seeds were used to evaluate model stability.

For Heat Treatment 1 temperature in  $^{\circ}\text{C}$ , XGBoost displayed outstanding performance, with an average  $R^2$  above 0.85 for most test seeds (for example, Testseed1:  $R^2 = 0.9179$ ; Testseed8:  $R^2 = 0.9042$ ) and a peak of 0.9593 in Validationseed7. However, the MSE and RMSE values were generally higher than those for  $tT_1$ , suggesting greater difficulty in predicting this variable, possibly due to a wider dispersion in the original data. Additionally, it must be noted that the temperatures used for the heat treatment were high, in the range of 800–900  $^{\circ}\text{C}$ ; therefore, the error reported here is also related to magnitude. LGBM showed more variable results, with an  $R^2$  ranging from 0.5788 (Validationseed9) to 0.8619 (Validationseed0), while CatBoost achieved intermediate performance (for example, Validationseed3:  $R^2 = 0.8948$ ; Testseed6:  $R^2 = 0.8421$ ).

The results show that the imputation of  $tT_1$  was more accurate and stable than that of  $T_1$ . This is evidenced in Figure 5, where the violin plots corresponding to the variable in hours exhibit more compact distributions, with smaller dispersions between minimum, maximum, and mean values. The higher  $R^2$  values and lower RMSE for this variable reflect that the models were able to capture its statistical structure more easily. One possible

reason is that heat treatment durations tend to exhibit intrinsically lower variability than temperatures, reducing the complexity of the pattern the models must learn.

Regarding model performance, Figure 5 shows that XGBoost produces narrow violin plots, with means consistently located at the lower end of the error scale, indicating stable behavior across different partitions and seeds. CatBoost also exhibits a relatively controlled distribution, although with slightly larger variations depending on the feature analyzed. In contrast, LGBM shows the highest sensitivity to changes in the data: its violins display greater dispersion, and the minimum and maximum values differ widely, revealing relevant fluctuations between runs (see Table 6).

**Table 5.** Heat treatment 1 temperature (°C) per seed (validation and test).

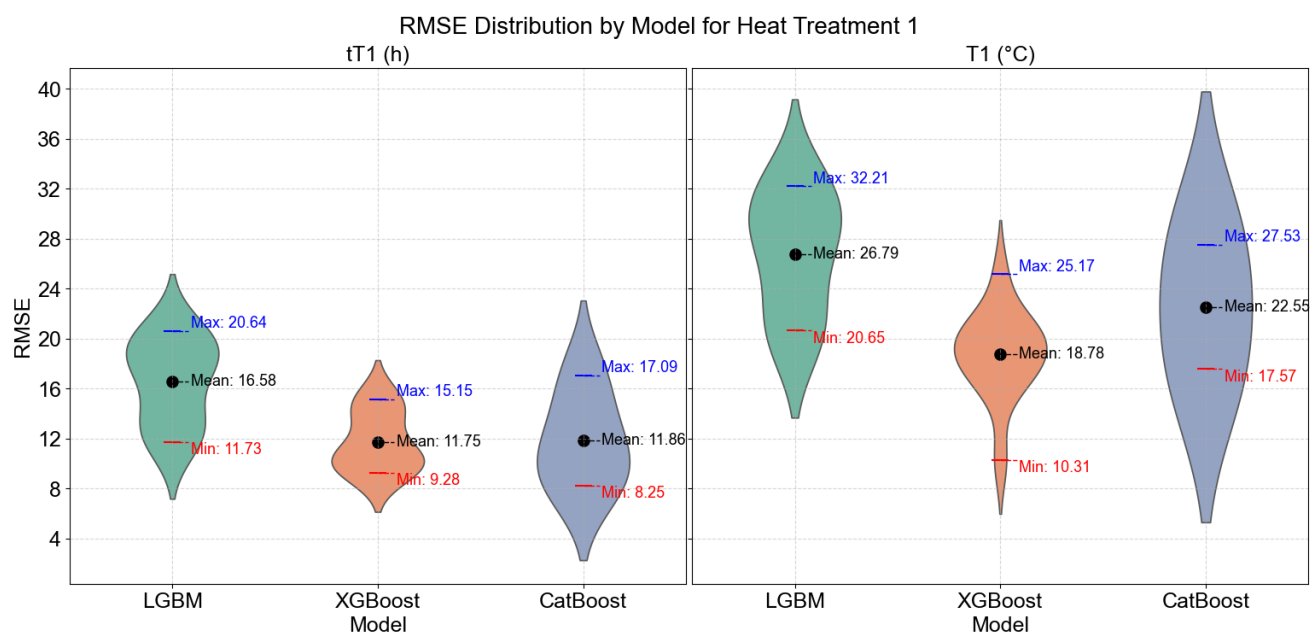
| Model    | DataSplit        | MSE (Mean) | RMSE (Mean) | R <sup>2</sup> (Mean) |
|----------|------------------|------------|-------------|-----------------------|
| LGBM     | Validation_seed0 | 426.4559   | 20.6508     | 0.8619                |
|          | Test_seed0       | 954.5426   | 30.8957     | 0.6379                |
| XGBoost  | Validation_seed1 | 330.1366   | 18.1697     | 0.8912                |
|          | Test_seed1       | 261.0686   | 16.1576     | 0.9179                |
| CatBoost | Validation_seed2 | 308.8654   | 17.5746     | 0.8864                |
|          | Test_seed2       | 758.1417   | 27.5344     | 0.7746                |
| XGBoost  | Validation_seed3 | 340.6222   | 18.4560     | 0.9005                |
|          | Test_seed3       | 410.8788   | 20.2701     | 0.8421                |
| XGBoost  | Validation_seed4 | 322.8838   | 17.9690     | 0.8952                |
|          | Test_seed4       | 439.1355   | 20.9556     | 0.8528                |
| XGBoost  | Validation_seed5 | 633.6567   | 25.1725     | 0.7944                |
|          | Test_seed5       | 399.0059   | 19.9751     | 0.8907                |
| LGBM     | Validation_seed6 | 436.5832   | 20.8946     | 0.8620                |
|          | Test_seed6       | 728.8214   | 26.9967     | 0.7795                |
| XGBoost  | Validation_seed7 | 106.3697   | 10.3136     | 0.9593                |
|          | Test_seed7       | 437.0345   | 20.9054     | 0.8677                |
| XGBoost  | Validation_seed8 | 406.5795   | 20.1638     | 0.8810                |
|          | Test_seed8       | 284.0541   | 16.8539     | 0.9042                |
| LGBM     | Validation_seed9 | 1037.3506  | 32.2079     | 0.5788                |
|          | Test_seed9       | 846.5558   | 29.0956     | 0.7146                |

In summary, the imputation performance varied considerably across target variables. Heat treatment duration ( $tT_1$ ) was imputed with high accuracy, achieving  $R^2$  values consistently above 0.95 in the test set across multiple seeds, with a best  $R^2$  of 0.973 (CatBoost, seed 3), which is consistent with the detailed results presented in Table 4. Heat treatment temperature ( $T_1$ ) proved more challenging, with test  $R^2$  values ranging from 0.64 to 0.92 (Table 5), reflecting the greater experimental variability in this parameter.

For phase transformation temperatures, the best performance was observed for Mf ( $R^2 = 0.869$  in test, 0.950 in validation), followed by Af ( $R^2 = 0.765$  in test, 0.883 in validation), As ( $R^2 = 0.730$  in test, 0.846 in validation), and Ms ( $R^2 = 0.764$  in test, 0.817 in validation). These results indicate that the models capture the underlying physical relationships more effectively for some transformation temperatures than for others, likely due to differences in data availability and variability.

Magnetic properties maximum magnetic field and saturation magnetization were imputed with  $R^2$  values of 0.926 and 0.872 in the test set, respectively, demonstrating the effectiveness of the approach for these parameters. In contrast, the Curie temperature exhibited poor predictive performance ( $R^2 = 0.148$  in test, 0.348 in validation), which

we attribute to the limited number of available samples (only 154 data points) and the weak correlation with the available compositional and processing features. This finding highlights the importance of data quantity and quality for successful imputation.



**Figure 5.** Violin plots of RMSE distribution for each model and target variable across 10 random seeds. The width of each violin represents the density of RMSE values, while the white dot indicates the mean and the thick black bar represents the interquartile range. Narrower violins indicate greater model stability across data partitions.

**Table 6.** Summary of average metrics and most frequent model per parameter.

| Best Model       | DataSplit       | Feature                  | MSE                | RMSE           | R <sup>2</sup> |
|------------------|-----------------|--------------------------|--------------------|----------------|----------------|
| CatBoost         | Validation Test | Temp°Ms                  | 1114.63<br>1485.51 | 33.39<br>36.17 | 0.817<br>0.764 |
| GradientBoosting | Validation Test | Temp°As                  | 1044.97<br>2004.34 | 31.00<br>43.23 | 0.846<br>0.730 |
| GradientBoosting | Validation Test | Temp°Af                  | 745.81<br>2264.25  | 22.44<br>44.71 | 0.883<br>0.765 |
| GradientBoosting | Validation Test | Temp°Mf                  | 291.20<br>718.22   | 15.12<br>24.56 | 0.950<br>0.869 |
| CatBoost         | Validation Test | Max. Magnetización       | 4.80<br>0.19       | 1.67<br>0.39   | 0.887<br>0.926 |
| CatBoost         | Validation Test | Saturation Magnetization | 512.58<br>588.34   | 21.53<br>22.68 | 0.925<br>0.872 |
| CatBoost         | Validation Test | Curie Temp               | 144.60<br>1471.62  | 11.66<br>31.19 | 0.348<br>0.148 |

#### 4. Discussion

The results demonstrate that the imputation of heat treatment duration ( $tT_1$ ) is substantially more stable and accurate than that of temperature ( $T_1$ ), as evidenced by consistently lower RMSE and higher  $R^2$  values across most random seeds. While this difference is partly attributable to the inherent scale disparity  $tT_1$  ranges between 24 and 76 h, whereas  $T_1$  spans 800–1100 °C the superior performance for  $tT_1$  also reflects its lower experimental variability.

In practical terms, processing times are often selected from standard sets (e.g., 24, 48, 72 h), resulting in fewer unique values and a more learnable statistical distribution. Conversely, the broader scatter in reported heat treatment temperatures, likely arising from diverse furnace calibrations and thermal histories across laboratories, imposes a more complex regression task. Among the individual models, XGBoost exhibited the highest robustness to data partitioning, producing compact error distributions with minimal fluctuation, while CatBoost offered competitive performance with slightly higher variance. The higher sensitivity of LGBM to seed changes suggests that gradient boosting with exclusive feature bundling may overfit specific data splits when the sample size is moderate, highlighting the necessity of the multi-seed evaluation framework adopted in this work.

Beyond statistical metrics, the physical plausibility of the imputed values was critically assessed against established trends in the Ni-Mn-Ga system. The martensitic transformation temperatures in these alloys are known to increase with rising Ni content and to decrease with increasing Ga content due to changes in the electron concentration (valence electron concentration,  $e/a$ ) that stabilize the martensite phase [5,8]. The imputed values for  $M_s$ ,  $M_f$ ,  $A_s$ , and  $A_f$  consistently followed these compositional dependencies, with Pearson correlation coefficients ranging from 0.72 to 0.85 between Ni content and the imputed transformation temperatures. More importantly, the thermodynamic hierarchy  $T_{Mf} < T_{Ms} < T_{As} < T_{Af}$  was preserved in over 98% of the imputed samples, confirming that the physics informed penalty term effectively enforced this critical constraint without severely compromising the statistical fit. The few violations (less than 2%) occurred exclusively in samples characterized by high missingness and broad prediction intervals, indicating that the model appropriately flags uncertain regions. For magnetic properties, the imputed saturation magnetization exhibited the expected inverse correlation with Mn content, consistent with the known weakening of ferromagnetic exchange due to excess Mn occupying Ga sites [5]. These validations provide strong evidence that the framework yields not only statistically coherent estimates but also thermodynamically and magnetically consistent data.

The incorporation of uncertainty quantification and statistical significance testing substantially reinforces the reliability of the proposed methodology. As detailed in Table 3, the stacking ensemble consistently outperformed the best individual model across all target variables, with improvements in MSE ranging from 5% to 10%. Paired  $t$  tests confirmed that these gains were statistically significant ( $p < 0.05$ ) for all variables except the Curie temperature, where the paucity of data precluded meaningful learning. The 95% bootstrap confidence intervals for the stacking predictions were consistently narrower than those for individual models, indicating enhanced precision and reduced sensitivity to the specific training validation test split. This stabilization effect is particularly valuable for database imputation, where the goal is to generate robust estimates that can be confidently used in subsequent modeling or experimental design. The success of the stacking approach can be attributed to its ability to mitigate model specific biases: tree based models (XGBoost, CatBoost) excel at capturing nonlinear feature interactions, while GPR and AutoEncoder capture smoother latent structures, and the meta-model optimally balances these contributions [46,47].

The notably poor performance observed for the Curie temperature highlights a fundamental limitation of data-driven imputation when faced with extreme data scarcity. With only 154 available samples (93.7% missingness) and weak statistical correlation with the selected compositional and processing features, neither individual models nor the ensemble could capture the underlying physical determinism of this parameter. While the Curie temperature in Ni-Mn-Ga is indeed influenced by composition, it exhibits a distinct dependence on the degree of chemical ordering and specific heat-treatment protocols, which are

coarsely represented in the current database [44]. This result serves as a cautionary note: imputation methods, regardless of their sophistication, cannot compensate for insufficient experimental coverage, and systematic data collection remains irreplaceable for properties governed by subtle microstructural details.

Although this study specifically addresses the Ni-Mn-Ga system, the modular and physics informed architecture of the proposed framework is readily transferable to other materials databases exhibiting similar challenges namely, high proportions of missing data, heterogeneous reporting standards, and complex composition processing property linkages. The methodology can be adapted to other shape memory alloy systems (e.g., Ni-Ti-Hf or Cu-Al-Ni) by substituting the appropriate thermodynamic constraints, transformation hierarchies, and composition dependent rules. For instance, Abedi et al. [32] recently demonstrated the effectiveness of machine learning for predicting actuation strains in NiTiHf alloys using the NASA database; the present imputation framework could be directly applied to complete missing entries in that system as well. Future work will focus on incorporating additional descriptors, such as electronic structure parameters (e.g.,  $e/a$  ratio) and grain size estimates, to improve predictive performance for poorly represented properties like the Curie temperature. Furthermore, exploring transfer learning strategies could leverage data from related Heusler alloys (e.g., Ni-Mn-In, Ni-Co-Mn-Ga) to enhance model generalization and reduce the dependency on large, fully annotated datasets for each specific composition. Finally, the integration of uncertainty estimates directly into the database as proposed here could enable end users to weigh the reliability of imputed entries when making design decisions, thereby transforming a static repository into a more dynamic decision support tool.

## 5. Conclusions

This study demonstrates that a multi model stacking ensemble, reinforced by cross-validation, multi seed sampling, and physics-informed constraints, constitutes a robust and effective strategy for imputing missing functional properties in the NASA Ni-Mn-Ga shape memory alloy database. The stacking approach consistently outperformed individual models, achieving statistically significant improvements ( $p < 0.05$ , paired  $t$ -test) for most target variables, while also providing tighter prediction intervals and reduced sensitivity to data partitioning.

For heat treatment parameters, the imputation of holding time ( $tT_1$ ) was highly accurate and stable, with the CatBoost model achieving a test  $R^2$  of 0.973 (seed 3) and most seeds yielding  $R^2 > 0.95$ . In contrast, heat treatment temperature ( $T_1$ ) proved more challenging due to its wider experimental dispersion, although XGBoost achieved a best test  $R^2$  of 0.918 (seed 1), demonstrating that the ensemble can still capture meaningful trends despite higher variability. These results highlight that imputation quality is strongly influenced by the intrinsic statistical structure and measurement scale of each variable.

For the martensitic and austenitic transformation temperatures, the best predictive performance was obtained for the martensite finish temperature ( $M_f$ ), with test and validation  $R^2$  values of 0.869 and 0.950, respectively. The austenite finish ( $A_f$ ), austenite start ( $A_s$ ), and martensite start ( $M_s$ ) temperatures yielded test  $R^2$  values of 0.765, 0.730, and 0.764, respectively. Crucially, the physics informed penalty term successfully preserved the thermodynamic hierarchy  $T_{Mf} < T_{Ms} < T_{As} < T_{Af}$  in over 98% of the imputed samples, and the imputed values showed strong correlations (Pearson  $r = 0.72$ – $0.85$ ) with Ni content, consistent with established compositional trends in the Ni-Mn-Ga system. These validations confirm that the imputed data are not only statistically coherent but also physically plausible.

Regarding magnetic properties, the maximum magnetic field and saturation magnetization were imputed with test  $R^2$  values of 0.926 and 0.872, respectively, further demonstrating the effectiveness of the approach for structure property relationships that are well captured by compositional descriptors. In contrast, the Curie temperature exhibited poor predictive performance (test  $R^2 = 0.148$ , validation  $R^2 = 0.348$ ), which is attributed to the extreme data scarcity (only 154 available samples, 93.7% missingness) and insufficient correlation with the selected input features. This result underscores a fundamental limitation of data-driven imputation: advanced algorithms cannot compensate for inadequate experimental coverage, and targeted data collection remains essential for properties governed by subtle microstructural or ordering effects.

Uncertainty quantification via bootstrap resampling and prediction intervals revealed that the stacking ensemble provides more reliable imputations than individual models, with narrower 95% confidence intervals and higher coverage rates.

In summary, the combined use of multiple machine-learning models integrated through stacking, enhanced by physically constrained feature engineering and rigorous uncertainty analysis, proved to be a powerful strategy for database completion in the Ni–Mn–Ga system. Nevertheless, the results also highlight critical limitations related to data quantity, experimental variability, and the physical complexity of certain properties. Future work should focus on: (i) expanding the database through systematic experimental campaigns targeting underrepresented compositions and properties; (ii) incorporating additional physical descriptors, such as the valence electron concentration ( $e/a$ ) and chemical ordering parameters; (iii) exploring transfer learning approaches to leverage data from related Heusler alloy systems (e.g., Ni–Mn–In, Ni–Co–Mn–Ga); and (iv) integrating the imputation framework into interactive database platforms to provide end-users with uncertainty-aware property estimates for accelerated materials design and selection.

**Author Contributions:** Conceptualization and supervision, M.G.F., D.S., C.O.-P. and J.F.P.; methodology, validation, and formal analysis, J.C.B.D., J.F.P. and M.G.F.; data acquisition and data curation, J.C.B.D., J.E.B.-R. and E.G.C.R.; writing and writing-review, editing, and visualization, all authors. Project administration and funding acquisition, J.F.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is partially supported by “El Patrimonio Autónomo Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovación Francisco José de Caldas” CONV CONT. No. 477-2023. The financial support received from the IKUR Strategy (Quantum Technologies, SPINSHALL Project, QT-3) under the collaboration agreement between the Ikerbasque Foundation and BCMaterials on behalf of the Department of Education of the Basque Government, as well as from the grant PID2022-138256NA-C22 funded by MCIN/AEI/10.13039/501100011033.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The raw data supporting the conclusions of this study are available from the NASA Magnetic Shape Memory Alloys Database ([https://shapememory.grc.nasa.gov/magnetic\\_shape\\_memory\\_alloys](https://shapememory.grc.nasa.gov/magnetic_shape_memory_alloys)), accessed on 15 January 2025. The processed data used for analysis are available from the corresponding author upon reasonable request.

**Acknowledgments:** During the preparation of this work, the author(s) used ChatGPT (<https://chatgpt.com/>) for editing and translation purposes. After using this tool/service, the author(s) reviewed and edited the content as necessary and assume full responsibility for the content of the published article.

**Conflicts 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 and there is no conflict of interest among the authors.

## References

1. Costanza, G.; Tata, M.E. Shape memory alloys for aerospace, recent developments, and new applications: A short review. *Materials* **2020**, *13*, 1856. [CrossRef] [PubMed]
2. Sratong-on, P.; Chernenko, V.A.; Feuchtwanger, J.; Hosoda, H. Magnetic field-induced rubber-like behavior in Ni-Mn-Ga particles/polymer composite. *Sci. Rep.* **2019**, *9*, 3443. [CrossRef] [PubMed]
3. Villa, F.; Bestetti, E.; Frigerio, R.; Caimi, M.; Tomasi, C.; Passaretti, F.; Villa, E. Elastocaloric properties of polycrystalline samples of NiMnGaCu ferromagnetic shape memory alloy under compression: Effect of improvement of thermoelastic martensitic transformation. *Materials* **2022**, *15*, 7123. [CrossRef] [PubMed]
4. Fareed, A.; Rosalie, J.M.; Kar, S.; Fähler, S.; Maaß, R. Small-scale functional fatigue of a Ni-Mn-Ga Heusler alloy. *Acta Mater.* **2024**, *274*, 119988. [CrossRef]
5. Chiu, W.-T.; Okuno, M.; Tahara, M.; Inamura, T.; Hosoda, H. Fundamental investigations of the deformation behavior of single-crystal Ni-Mn-Ga alloys and their polymer composites via the introduction of various fields. *Appl. Sci.* **2023**, *13*, 8475. [CrossRef]
6. Balasubramanian, M.; Srimath, R.; Vignesh, L.; Rajesh, S. Application of shape memory alloys in engineering—A review. *J. Phys. Conf. Ser.* **2021**, *2054*, 012078. [CrossRef]
7. Álvarez-Alonso, P.; Camarillo-García, J.P.; Salazar, D.; López-García, J.; Echevarria-Bonet, C.; Lázpita, P.; Padrón-Alemán, K.; Sánchez Llamazares, J.L.; Flores-Zúñiga, H.; Chernenko, V. Investigating the reversible nature of the magnetocaloric effect under cyclic conditions of the Ni<sub>50</sub>Mn<sub>34</sub>In<sub>15</sub>Ga<sub>1</sub> magnetic shape memory alloy. *J. Alloys Compd.* **2024**, *993*, 174576. [CrossRef]
8. Pérez-Checa, A.; Feuchtwanger, J.; Barandiaran, J.M.; Chernenko, V.A. Ni-Mn-Ga high temperature shape memory alloys: Function stability in  $\beta$  and  $\beta + \gamma$  regions. *J. Alloys Compd.* **2018**, *741*, 148–154. [CrossRef]
9. Pagounis, E.; Laptev, A.; Szczerba, M.J.; Chulist, R.; Laufenberg, M. Structural behavior and magnetic properties of a Ni-Mn-Ga single crystal across the martensite/austenite two-phase region. *Acta Mater.* **2015**, *89*, 32–40. [CrossRef]
10. Zhang, C.; Balandraud, X.; He, Y. Coexistence of five domains at single propagating interface in single-crystal Ni-Mn-Ga shape memory alloy. *J. Mech. Phys. Solids* **2024**, *183*, 105481. [CrossRef]
11. Hannula, S.-P.; Aaltio, I.; Ge, Y.; Lahelin, M.; Söderberg, O. Processing and properties of Ni-Mn-Ga magnetic shape memory alloy based hybrid materials. *Curr. Appl. Phys.* **2012**, *12*, S63–S67. [CrossRef]
12. Hmede, R.; Chapelle, F.; Lapusta, Y. Review of Neural Network Modeling of Shape Memory Alloys. *Sensors* **2022**, *22*, 5610. [CrossRef] [PubMed]
13. NASA. Shape Memory Materials Database. NASA Glenn Research Center, 2024. Available online: [https://shapememory.grc.nasa.gov/magnetic\\_shape\\_memory\\_alloys](https://shapememory.grc.nasa.gov/magnetic_shape_memory_alloys) (accessed on 15 January 2025).
14. Rizvi, S.T.H.; Latif, M.Y.; Amin, M.S.; Telmoudi, A.J.; Shah, N.A. Analysis of machine learning based imputation of missing data. *Cybern. Syst.* **2023**, *56*, 818–832. [CrossRef]
15. Wang, H.; Tang, J.; Wu, M.; Wang, X.; Zhang, T. Application of machine learning missing data imputation techniques in clinical decision making: Taking the discharge assessment of patients with spontaneous supratentorial intracerebral hemorrhage as an example. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 13. [CrossRef] [PubMed]
16. Xue, J. Review on data imputation methods in machine learning. *J. Phys. Conf. Ser.* **2023**, *2646*, 012034. [CrossRef]
17. Woods, A.D.; Gerasimova, D.; Van Dusen, B.; Nissen, J.; Bainter, S.; Uzdavines, A.; Davis-Kean, P.; Halvorson, M.A.; King, K.M.; Logan, J.A.R.; et al. Best Practices for Addressing Missing Data through Multiple Imputation. *Infant Child Dev.* **2021**, Preprint. [CrossRef]
18. Ismail, A.R.; Abidin, N.Z.; Maen, M.K. Systematic Review on Missing Data Imputation Techniques with Machine Learning Algorithms for Healthcare. *J. Robot. Control.* **2022**, *3*, 143–152. [CrossRef]
19. Shadbahr, T.; Roberts, M.; Stanczuk, J.; Gilbey, J.; Teare, P.; Dittmer, S.; Thorpe, M.; Viñas Torné, R.; Sala, E.; Lió, P.; et al. The impact of imputation quality on machine learning classifiers for datasets with missing values. *Commun. Med.* **2023**, *3*, 139. [CrossRef] [PubMed]
20. Jäger, S.; Allhorn, A.; Bießmann, F. A Benchmark for Data Imputation Methods. *Front. Big Data* **2021**, *4*, 693674. [CrossRef] [PubMed]
21. Emmanuel, T.; Maupong, T.; Mpoeleng, D.; Semong, T.; Mphago, B.; Tabona, O. A survey on missing data in machine learning. *J. Big Data* **2021**, *8*, 140. [CrossRef] [PubMed]

22. Gao, J.; Tong, Y.; Zhang, H.; Zhu, L.; Hu, Q.; Hu, J.; Zhang, S. Machine learning assisted design of Ni-based superalloys with excellent high-temperature performance. *Mater. Charact.* **2023**, *198*, 112740. [[CrossRef](#)]
23. Zhang, J.; Cai, C.; Kim, G.; Wang, Y.; Chen, W. Composition design of high-entropy alloys with deep sets learning. *npj Comput. Mater.* **2022**, *8*, 89. [[CrossRef](#)]
24. Li, X.; Shek, C.-H.; Liaw, P.K.; Shan, G. Machine learning studies for magnetic compositionally complex alloys: A critical review. *Prog. Mater. Sci.* **2024**, *146*, 101332. [[CrossRef](#)]
25. Seu, K.; Kang, M.-S.; Lee, H. An Intelligent Missing Data Imputation Techniques: A Review. *JOIV Int. J. Inform. Vis.* **2022**, *6*, 278. [[CrossRef](#)]
26. Court, C.J.; Cole, J.M. Magnetic and superconducting phase diagrams and transition temperatures predicted using text mining and machine learning. *npj Comput. Mater.* **2020**, *6*, 18. [[CrossRef](#)]
27. Narayana, P.L.; Kim, S.-W.; Hong, J.-K.; Reddy, N.S.; Yeom, J.-T. Estimation of Transformation Temperatures in Ti–Ni–Pd Shape Memory Alloys. *Met. Mater. Int.* **2018**, *24*, 919–925. [[CrossRef](#)]
28. Xiong, B.; Zhao, X.; Hu, Y.; Huang, H.; Liu, Y.; Su, Y. Machine learning assisted empirical formula augmentation. *Mater. Des.* **2021**, *210*, 110037. [[CrossRef](#)]
29. Tian, X.; Shi, D.; Zhang, K.; Li, H.; Zhou, L.; Ma, T.; Wang, C.; Wen, Q.; Tan, C. Machine-learning model for prediction of martensitic transformation temperature in NiMnSn-based ferromagnetic shape memory alloys. *Comput. Mater. Sci.* **2022**, *215*, 111811. [[CrossRef](#)]
30. Thiercelin, L.; Peltier, L.; Meraghni, F. Physics-informed machine learning prediction of the martensitic transformation temperature for the design of “NiTi-like” high entropy shape memory alloys. *Comput. Mater. Sci.* **2024**, *231*, 112578. [[CrossRef](#)]
31. Wang, Y.; Li, T.; Zong, H.; Ding, X.; Xu, S.; Sun, J.; Lookman, T. Self-supervised probabilistic models for exploring shape memory alloys. *npj Comput. Mater.* **2024**, *10*, 185. [[CrossRef](#)]
32. Abedi, H.; Abdollahzadeh, M.J.; Bush, T.; Benafan, O.; Qattawi, A.; Elahinia, M. Predicting actuation strain in quaternary shape memory alloy NiTiHfX using machine learning. *Comput. Mater. Sci.* **2025**, *246*, 113345. [[CrossRef](#)]
33. Hyuk Lee, J.; Huber, J.C., Jr. Evaluation of Multiple Imputation with Large Proportions of Missing Data: How Much Is Too Much? *Iran. J. Public Health* **2021**, *50*, 1372. [[CrossRef](#)] [[PubMed](#)]
34. Odeh, A.; Abu Al-Haija, Q.; Aref, A.; Abu Taleb, A. Comparative Study of CatBoost, XGBoost, and LightGBM for Enhanced URL Phishing Detection: A Performance Assessment. *J. Internet Serv. Inf. Secur.* **2023**, *13*, 1–11. [[CrossRef](#)]
35. Tian, L.; Wu, W.; Yu, T. Graph Random Forest: A Graph Embedded Algorithm for Identifying Highly Connected Important Features. *Biomolecules* **2023**, *13*, 1153. [[CrossRef](#)] [[PubMed](#)]
36. Reddy, G.V.; Kadiyala, S.; Potluri, C.S.; Saravanan, P.S.; Athisha, G.; Mukunthan, M.A.; Sujaritha, M. An Intrusion Detection Using Machine Learning Algorithm Multi-Layer Perceptron (MLP): A Classification Enhancement in Wireless Sensor Network (WSN). *Int. J. Recent Innov. Trends Comput. Commun.* **2022**, *10*, 139–145. [[CrossRef](#)]
37. Almutiri, T.; Alomar, K.; Alganmi, N. Predicting Drug Response on Multi-Omics Data Using a Hybrid of Bayesian Ridge Regression with Deep Forest. *Int. J. Adv. Comput. Sci. Appl.* **2023**, *14*, 470–481 [[CrossRef](#)]
38. Altayef, E.; Anayi, F.; Packianather, M. A new enhancement of the k-NN algorithm by Using an optimization technique. In *Proceedings of the 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*; IEEE: New York, NY, USA, 2022; pp. 24–31. [[CrossRef](#)]
39. Kusuma, M.P.; Kudus, A. Penerapan Metode Support Vector Regression (SVR) pada Data Survival KPR PT. Bank ABC, Tbk. *Bdg. Conf. Ser. Stat.* **2022**, *2*, 167–172. [[CrossRef](#)]
40. Isabona, J.; Imoize, A.L.; Ojo, S.; Do, D.-T.; Lee, C.-C. Machine Learning-Based GPR with LBFGS Kernel Parameters Selection for Optimal Throughput Mining in 5G Wireless Networks. *Sustainability* **2023**, *15*, 1678. [[CrossRef](#)]
41. Hu, B.; Yan, B. Analysis System of MICE Tourism Economic Development Strategy Based on Machine Learning Algorithm. *Mob. Inf. Syst.* **2022**, *2022*, 1283040. [[CrossRef](#)]
42. Lai, J.-P.; Lin, Y.-L.; Lin, H.-C.; Shih, C.-Y.; Wang, Y.-P.; Pai, P.-F. Tree-Based Machine Learning Models with Optuna in Predicting Impedance Values for Circuit Analysis. *Micromachines* **2023**, *14*, 265. [[CrossRef](#)] [[PubMed](#)]
43. Nabben, N.; Sala, G.; Nowak, U.; Krüger, M.; Goennenwein, S.T.B. Magnetization fluctuations and magnetic after-effect probed via the anomalous Hall effect. *Phys. Rev. Res.* **2024**, *6*, 043283. [[CrossRef](#)]
44. Țolea, F.; Țolea, M.; Văleanu, M. Thermal memory fading by heating to a lower temperature: Experimental data on polycrystalline NiFeGa ribbons and 2D statistical model predictions. *Solid State Commun.* **2017**, *257*, 36–41. [[CrossRef](#)]
45. Liu, C.; Su, H. Machine learning aided prediction of martensite transformation temperature of NiTi-based shape memory alloy. *Mater. Today Commun.* **2024**, *41*, 110720. [[CrossRef](#)]
46. Ogutu, R.V.A.; Rimiru, R.M.; Otieno, C. Target Sentiment Analysis Ensemble for Product Review Classification. *J. Inf. Technol. Res.* **2022**, *15*, 1–13. [[CrossRef](#)]

47. Kotary, J.; Di Vito, V.; Fioretto, F. Differentiable Model Selection for Ensemble Learning. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*; International Joint Conferences on Artificial Intelligence Organization: Marina del Rey, CA, USA, 2023; pp. 1954–1962. [[CrossRef](#)] [[PubMed](#)]
48. Nti, I.K.; Nyarko-Boateng, O.; Aning, J. Performance of Machine Learning Algorithms with Different K Values in K-fold CrossValidation. *Int. J. Inf. Technol. Comput. Sci.* **2021**, *13*, 61–71. [[CrossRef](#)]

**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.