Physics-informed machine learning prediction of the martensitic transformation temperature for the design of "NiTi-like" high entropy shape memory alloys

Léo Thiercelin, Laurent Peltier, Fodil Meraghni *

Arts et Métiers Institute of Technology, CNRS, Université de Lorraine, LEM3-UMR 7239, F-57000 Metz, France

A B S T R A C T

The present study proposes a physics-informed machine learning (PIML) algorithm-based approach aimed at predicting the martensitic transformation temperature (Ms) for the design of "NiTi-like" high entropy shape memory alloys (HESMAs). A previously established HESMAs database is enriched and extended to include binary, ternary, quaternary, quinary and senary alloys containing the most employed alloying elements for HEAs design such as Ni equivalents (Fe, Cu, Co, Pd, Pt and Au), Ti equivalents (Zr and Hf), Nb and Ta. The Extremely Randomized Trees algorithm, based on the concept of multiple random decision tree predictions, is adopted as the regression method for Ms temperature prediction. Two strategies for the algorithm inputs have been adopted, discussed, and compared in terms of reliable predictions. The first relies on the composition of the alloying elements, whereas the second exploits a defined set of intrinsic material descriptors. The latter are based on mixing enthalpy, atomic radius, electronegativity, atomic number and number of elements. A high accuracy of the M_S prediction has been reached when considering the material descriptors. In fact, the second strategy induces a mean absolute error that is less than 30 °C for alloys containing up to 4 elements. For more elements there are more discrepancies due to the homogenization state required for HEAs. The validation of the developed approach has been performed using 6 home-made HESMAs prepared specifically for this study. It demonstrated the predictive capabilities of the developed physics-informed machine learning based approach. Finally, a HESMA design tool has been implemented to virtually design new HESMAs with a targeted Ms temperature above 400 °C. It is worth noting that this aspect is one of the most challenging engineering issues for such alloys. An illustrative case applied to the $(NiCuPd)_{50}(TiZr)_{50}$ family of alloys demonstrates the predictive capabilities of the developed approach to design such alloys to achieve a Ms temperature in the range of 300 °C to 700 °C.

1. Introduction

For over three thousand years, the human history has been faced to several revolutions closely linked to metallurgy such as the Bronze Age followed by the Iron Age and the Steel Age [1]. These revolutions led to the development of alloys to improve and/or provide new properties compared to pure elements. Indeed, the beginning of the 20th century was accompanied by the discovery of the "shape memory effect" on Au-Cd alloys by Ölander [2] and by Greninger Mooradian on Cu-Zn alloy few years later [3]. This effect, initially described by Chang and Read [4] was correlated to the thermomechanical induced martensitic transformation. Then, the binary NiTi alloy, incidentally discovered by William Buehler in 1963 [5,6], has begun the most widely used Shape Memory Alloy (SMA). The followed decades were dedicated to improve the transformation temperatures ranged from −70 to 70 °C for binary NiTi [7] by incorporating other alloying elements [8]. In fact, in the review by Ma et al. [9] dealing with High Temperature SMAs (HTSMA), it has been revealed that the addition of Pd, Au, Pt, Zr or Hf significantly increases the transformation temperatures for NiTi-like ternary alloys. However, this addition can lead to a decrease of the mechanical performance of the ternary NiTi-like, notably its functional fatigue durability [10]. Therefore, the new trend is the design of Complex Composition Alloys (CCA), which consists in the addition of more than one element to stabilize the material properties and actuation responses at high temperature. Indeed, as shown by Peltier et al. [11], the thermal fatigue performance of NiTiZr ternary alloy has been improved four

* Corresponding author.
E-mail address: fodil.meraghni@ensam.eu (F. Meraghni).

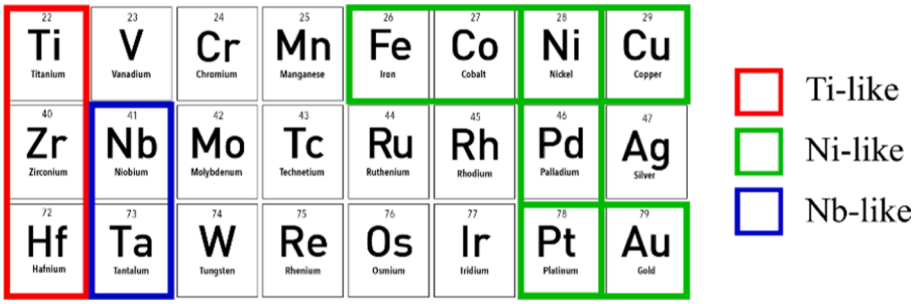


Fig. 1. Periodic element table showing the the Ni-like, Ti-like and Nb-like elements used in this study.

times when considering NiTiZr-CuHf quinary alloy.

In 1788, Achard [12] published a precursor book relating the results of research on more than 900 different combinations of alloys, including equiatomic alloys composed of 5 to 7 alloying elements. More recently, Cantor [13] proposed a derived concept of CCA, called High Entropy Alloys (HEA) based on the idea of a mixture of at least five elements whose concentration remains equiatomic and the structure is single phase for some of them. These studies have highlighted a great potential for this new family of materials such as a high mechanical strength due to the crystallographic lattice distortion, a high thermal stability due to the sluggish diffusion and the synergy of several properties of the alloying elements thanks to the cocktail effect [14]. The high entropy SMAs (HESMA) have emerged over the last years and is a promising way to improve at the same time the mechanical strength and other SMAs' features that are kept even at high temperature such as the memory effect [15], the superelasticity [16] or the damping effect [17].

Currently, the challenge facing HESMAs resides in their metallurgical design to achieve specific required properties at high temperature. The use of Machine Learning (ML) techniques, particularly for alloys with five and more elements, constitutes a potential way to address this challenge. The "Machine Learning" Age has started few years ago in metallurgy for the physical properties' prediction such as the phase determination [18–20], the lattice parameters [21,22], the mechanical properties [23–26], or the transformation temperatures, as detailed in the following paragraph. For the HEAs, the input data are generally the chemical composition and the content of each alloying element. In the present work, the chemical inputs are combined with other material physical parameters, namely: the atomic number, the atomic radius, the electronegativity, the valence electron number and the valence electron concentration [26,27]. Furthermore, thermodynamical parameters such as entropy energy, enthalpy energy or the balance are also considered. In addition, the use of thermodynamical calculations, density functional theory or molecular dynamics also enriches the ML algorithm [29–33].

The Ms prediction using ML have been recently implemented for ferromagnetic NiMnSn-SMAs [34], Fe-based SMAs [35], CuAl-based SMAs [36] and more widely for NiTi-like including three elements such as NiTiPd [37], NiTiHf [38,39,40] five elements NiTiCuFePd [26,41] or even six elements TiZrHfNiCoCu alloys [42]. Moreover, several authors extended the approach by including much more alloys containing other alloying elements in their database [43–46]. In the ML approach developed in these, the physical and thermodynamical features are accounted as descriptors and are determined as functions of the alloying element contents. A combination of chemical elements and material descriptors are then selected to predict the transformation temperatures. The number of input parameters for Ms predictions is therefore relatively high compared to the HEAs database. For example, [11] used 11 parameters for only 16 HEAs, [42] used 32 parameters for 50 alloys, while [39] used 48 inputs for 554 entries. Feature importance studies are often used to discriminate which parameters are most influential in prediction. However, the entries generally contain redundant information since they always include the chemical elements with several material descriptors. There are no studies that take into account only material descriptors, nor other works that minimize the number of entries, which can be too large compared to the small size of the databases.

In addition, HESMAs require detailed metallurgical state (homogenized or not, presence of intermetallics or precipitated) and heat treatment information that is currently lacking in established databases. These conditions are mandatories for HESMAs that need to be well homogenized before Differential Scanning Calorimetry (DSC) measurements. Otherwise, the Ms measurements are meaningless and metallurgical state-dependent since two identical chemical composition can lead to different Ms as reported by Piorunek et al. [47,48] and confirmed by Peltier et al. [11]. In fact, these authors have built a first HEAs database containing only homogenized HESMAs where the content of the intermetallics was neglected (less than 5%). They proposed a first linear model to predict the Ms accounting for the alloying elements. Nevertheless, the proposed relationship is unlikely to predict Ms temperature for ternary alloys whose dependency is highly non-linear [9]. Therefore, this model's limitation needs to be addressed to better capture the non-linear dependencies and hence to be suitable for a reliable Ms prediction for ternary, quaternary, quinary, and senary alloys. In addition, the design of HESMAs with Ms temperatures above 400 °C are still limited to some studies [49], which constitutes a challenging issue that could be easily addressed through a "Machine Learning" approach.

The objective of this work is to develop a physics-informed machine learning (PIML) algorithm-based approach to reliably predict the martensitic transformation temperature in HESMAs based on a reduced number of parameters. The developed prediction tool will be used to design and optimize the composition of "NiTi-like" HESMAs with respect to a target Ms temperature. The HESMAs database previously established by Peltier et al. [11], has been enriched and extended to binary, ternary, quaternary, quinary and senary alloys containing the most employed alloying elements for HEAs design such as Ni equivalents (Fe, Cu, Co, Pd, Pt and Au), Ti equivalents (Zr and Hf), Nb and Ta. A metallurgical criterion based on hysteresis of martensitic transformation was proposed to make the database more consistent than previous studies. The Ms temperature predictions will account for a set of an optimized number of intrinsic material descriptors defined through the mixing enthalpy, the atomic radius, the electronegativity, the atomic number and the number of elements. Finally, the predictive capabilities of the PIML approach are demonstrated through the design and the virtual elaboration of HESMAs with Ms temperature above 400 °C.

The paper is outlined as follows: the next section is devoted to the description and the analysis of the "NiTi-like" database extended to include binary, ternary, quaternary, quinary and senary alloys. The intrinsic physical materials descriptors to the chemical composition are established and discussed. The second section details the machine learning algorithm and provides a discussion dealing with the two considered strategies for the data input and the related sensitivity on the Ms temperature predictions. The relevance of the proposed strategy, based on the intrinsic material descriptors, is demonstrated and discussed through an illustrative case study for designing a $(NiCuPd)_{50}(TiZr)_{50}$ HESMA with a target Ms temperature above 400 °C. The

Table 1
Percentage occurrence of the different pairs of elements (%). The gray cases directly represent the percentage of occurrence of each chemical element, since Ni and Ti are always present for each composition. The database consists of 396 alloys, including 240 binary and ternary alloys, 113 quaternary alloys and 43 HEAs.

	Ni	Cu	Co	Fe	Pd	Pt	Au	Ti	Zr	Hf	Nb	Ta
Ni		38	9	4	15	2	2	100	23	33	12	15
Cu			5	0	5	0	0	38	12	17	2	0,3
Co				0	0	0	0	9	4	4	0	0,3
Fe					0	0	0	4	0	0	0,5	0,3
Pd						0	0	15	2	3	0	1
Pt							0	2	0	0	0	0
Au								2	0	0	0	0
Ti									23	33	12	15
Zr										15	1	0
Hf											2	2
Nb												1
Ta												

concluding remarks of the present work are summarized in the last section, which draws some pertinent perspectives as next steps for the developed approach.

2. Data Availability: Database of NiTi-like SMAs

2.1. A general overview of the database

This section details the establishment and the development of the database, which focuses on “NiTi-like" alloys containing 1, 2, 3, and more of the most employed elements that are thermodynamically closest to Ni and Ti. Indeed, Peltier et al. [15–17] classified the elements as Ni-like (Ni, Fe, Co, Cu, Pd, and Pt), Ti-like (Ti, Zr and Hf) and Nb-like (Nb and Ta), see Fig. 1. For the data availability, the database and the related references can be found at:

https://data.mendeley.com/datasets/z6j6g3b9yf/draft?a = 00b9d4d9-3e19-4552-b670-cf763375cc25, an open-source online data repository hosted at Mendeley Data (Thiercelin et al., 2023).[‡] https://data.mendeley.com/datasets/z6j6g3b9yf/draft?a = 00b9d4d9-3e19-4552-b670-cf763375cc25.

It is worth noticing that only the compositions well homogenized and, in some cases, followed by heat treatments at high temperature above 800 °C have been included in the database. This constraint is mandatory for the HESMAs elaboration otherwise there could be many inconsistencies in the Ms temperature estimation. Indeed, it is commonly stated that the Ms is strongly dependent on metallurgical state: several non-homogenized alloys with the same chemical composition can exhibit a Ms with large discrepancies [11,47,48]. A criterion of homogenization of the alloys has been adopted based on the DSC curves. If the endothermic and exothermic peaks are straight enough, then the alloy is included in the database. Otherwise, the alloy is omitted and withdrawn from the database. A difference between the start and the end of the martensitic transformation suitable is set to be below 50 °C (Ms - Mf < 50 °C).

The database contains 408 compositions including 59% binary and ternary alloys, 28% quaternary alloys and the remaining 13% are high entropy alloys (more than 5 elements). Table 1 shows the distribution of each alloying element in the database. It should be noted that Cu, Hf and Zr are the most employed elements in the elaboration of NiTi-like with a proportion of 38, 33 and 23%, respectively. In addition, those three elements are often combined to elaborate quaternary and high entropy alloys since Cu-Hf, Zr-Hf and Cu-Zr represent 17, 15 and 12% respectively of the database (Table 1). In addition, 6 other NiTi-like HEAs have been designed. These fabricated HESMAs alloys have then been heat treated to be well homogenized in term of microstructures as shown in Appendix A. DSC measurements confirmed the homogenized state of each material since the martensitic transformation extends to at least 50 °C. Table 2 summarizes the chemical composition of the 6 alloys with the transformation temperatures measured with the DSC curves. After the database filtering process, the number of “allowed” alloys matching the homogenization condition decreased to 364 where the percentage of HEAs went to 7%. Therefore, the new proportion of binary and ternary and quaternary slightly increased to 63% and 30%, respectively.

Table 2
Transformation temperatures of the 6 homemade HESMAs elaborated. See the DSC curves and the metallurgical state in Appendix A.

Chemical Composition	Mf [°C]	Ms [°C]	Af [°C]	As [°C]	Gap = |Ms-Mf|[°C]
$(Ni_{25}Cu_{20}Co_5)(Ti_{30}Zr_{20})$	−48	−20	35	0	28
$(Ni_{25}Cu_{25})(Ti_{25}Zr_{15}Hf_{10})$	−92	−80	−56	−72	12
$(Ni_{25}Cu_{17.5}Co_{7.5})(Ti_{16.7}Zr_{16.7}Hf_{16.7})$	128	165	187	168	37
$(Ni_{24.8}Cu_{15.5}Co_{8.5})(Ti_{16.5}Zr_{15.5}Hf_{13.5})Ta_{5.7}$	80	126	210	164	46
$(Ni_{24.7}Cu_{18.3}Co_{7.4})(Ti_{17.9}Zr_{16.4}Hf_{15.3})$	65	112	132	106	47
$(Ni_{23.8}Cu_{26.4})(Ti_{18.1}Zr_{16.5}Hf_{15.2})$	46	82	108	75	36

Table 3
Physical material descriptors chosen for the prediction of Ms temperature. Note that $e_v^i, Z_i, r_i and \chi_i$ correspond to the properties of the pure alloying elements i given in the Table B.1 of the Appendix B. H_{mix}^{ij} refers to the mixing enthalpy between a couple of alloying elements i and j whose values are given in Table B.2 of the Appendix B.

Material descriptors	Equations
Number of elements	n
Proportion of Ni-like, Ti-like, and Nb-like element	$\bar{f}_{Ni} = f_{Ni} + f_{Cu} + f_{Co} + f_{Fe} + f_{Pd} + f_{Au} + f_{Pt}$ $\bar{f}_{Ti} = f_{Ti} + f_{Zr} + f_{Hf}$ $\bar{f}_{Nb} = f_{Nb} + f_{Ta}$
Valence Electron Number	$ven = \sum_i f_i e_v^i$
Valence Electron Concentration	$vec = \frac{ven}{\bar{Z}}$ with $\bar{Z} = \sum_i f_i Z_i$
Mixing enthalpy energy	$\Delta H_{mix} = \sum_i \sum_{j>i} 4 f_i f_j H_{mix}^{ij}$
Atomic radius (mean $\bar{r}$ and variation coefficient δr)	$\bar{r} = \sum_i f_i r_i$ and $\delta r = \sqrt{\sum_i f_i \left(1 - \frac{r_i}{\bar{r}}\right)^2}$
Electronegativity (mean $\bar{\chi}$ and variation coefficient $\delta\chi$)	$\bar{\chi} = \sum_i f_i \chi_i$ and $\delta\chi = \sqrt{\sum_i f_i \left(1 - \frac{\chi_i}{\bar{\chi}}\right)^2}$

2.2. Physical material descriptors “driving” the Ms temperature

In this section, the relevance of 11 physico-chemical and thermodynamic parameters assumed to be driving the martensitic trans-

[‡] Thiercelin, Léo; Peltier, Laurent; MERAGHNI, Fodil (2023), “Database providing Ms Temperature for NiTi-Like alloys”, Mendeley Data, v1. http://dx.https://doi.org/10.17632/z6j6g3b9yf.1.

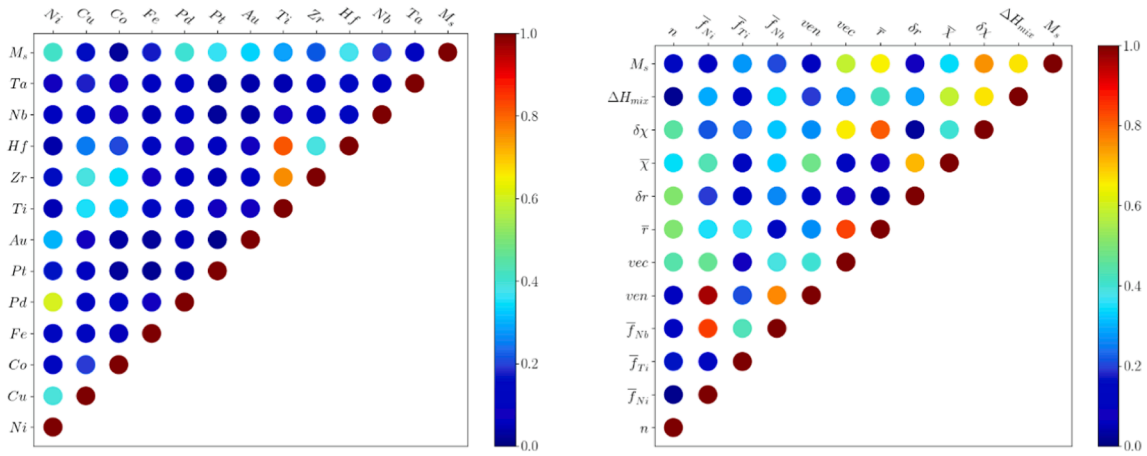


Fig. 2. Pearson's correlation matrix between the Ms temperature and the input parameters that are a) the alloying elements and b) the physical materials descriptors.

formation temperature is discussed. These descriptors are summarized in Table 3. Firstly, the number of elements n in the composition is explicitly considered since for more than 4 elements, the so called "cocktail effect" in high entropy alloys must be important. To reduce the number of inputs, the proportion of each element is replaced by the proportion of Ni-like, Ti-like and Nb-like, noted $\overline{f}_{Ni}$, $\overline{f}_{Ti}$ and $\overline{f}_{Nb}$, respectively. In addition, the valence electron and the atomic number appear to have a strong effect on the Ms temperature electron number as explained by Zarinejad et al. [50,51]. Therefore, the valence electron number *ven* and the valence electron concentration *vec* are also considered. Based on the ML models on the elaboration of HEAs, several parameters are also integrated to describe the mismatches between each element in terms of mixing enthalpy, namely ΔH_{mix}, the average atomic radius (average $\overline{r}$ and the variation coefficient δr) and the electronegativity (average $\overline{\chi}$ and the variation coefficient $\delta\chi$).

Fig. 2 displays the Pearson's correlation matrix (absolute values) between the input parameters and the Ms temperature. A value close to 1 between two parameters means that the latter are correlated and can be expressed through linear relationship. A value of 0 reflects a large scatter between these two parameters and no linear dependency can be found between them. The first row of the Pearson's correlation matrix (Fig. 2a) expresses the correlation between each alloying element and the Ms temperature. On the one hand, one can notice that the Ms temperature is more correlated to the elements Ni, Pd, Pt and Hf with an average correlation coefficient close to 0.4. On the other hand, Cu, Co, Nb and Ta exhibit a lower correlation with Ms (with an average correlation coefficient up to 0). However, the Pearson's correlation matrix (Fig. 2a) should be carefully interpreted since it is estimated based on the occurrence percentage of each alloying element in NiTi-like alloys (gray cases in Table 1).

A solution to the previous problem lies in the use of unified physical material descriptors that can be calculated for all NiTi-like alloys considered, regardless of the alloying elements. In fact, Fig. 2b shows a more reliable correlation Pearson's matrix between the physical materials descriptors, on one hand, and with the transformation temperature Ms on the other hand. One can notice that the correlation with Ms is more pronounced with this set of parameters for values that can reach 0.8. The latter corresponds to twice the best correlation reached with the alloying elements. In fact, the parameters $\delta\chi$, ΔH_{mix}, *vec* and $\overline{r}$ exhibit a high correlation with Ms, which reflects a certain sensitivity of Ms to these descriptors. Conversely, *ven* and Ms appear to be less correlated, as the correlation value is about 0.

In addition, based on the Pearson's correlation matrix, the Ms of all the considered alloys in the database are plotted against the most

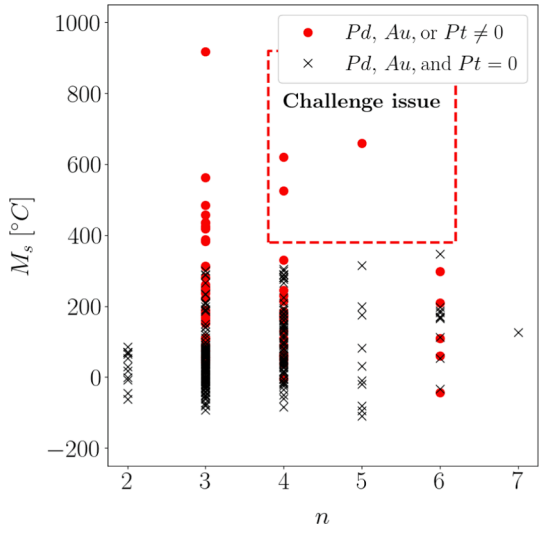


Fig. 3. Effect of the number of alloying elements on Ms temperature. The chemical elements Pd, Au and Pt allow reaching a Ms temperature exceeding 400 °C which is a challenge for the design of SMAs with at least 4 elements.

relevant parameters for the prediction of the Ms. Firstly, Fig. 3 shows the transformation temperature classified by the number of elements in each alloy. Most of the analyzed composition have a Ms ranging between −110 °C and 300 °C. Above 300 °C, there are only few alloys that contain valuable elements such as Au, Pd and Pt. Furthermore, it must be pointed out that the design of HESMAs does not necessarily lead to high transformation temperatures since only one homogenized composition has Ms above 400 °C [49].

In Fig. 4a and Fig. 4b, Ms temperature is represented against *ven* and *vec* for ternary alloys to investigate the influence of the substitution alloying element. Fig. 4a shows that the most of alloys have a *ven* close to 7, which corresponds to the *ven* of the binary alloys $Ni_{50}Ti_{50}$. Moreover, it is observed that the highest values of Ms are reached for alloys having a *ven* close to 7. This is explained by the fact that the increase of Ms can be achieved by replacing Ni and/or Ti by their respective equivalent element with the same valence number, *i.e.,* in the same column of the Mendeleev table (Fig. 1), namely: Pd and Pt (to substitute Ni) and, Zr and Hf (to replace Ti). Conversely, as reported in literature [50,51], it can be claimed that the further away the *ven* is from 7, the lower the Ms. In fact, it is worth noticing that a huge proportion of Nb-like elements

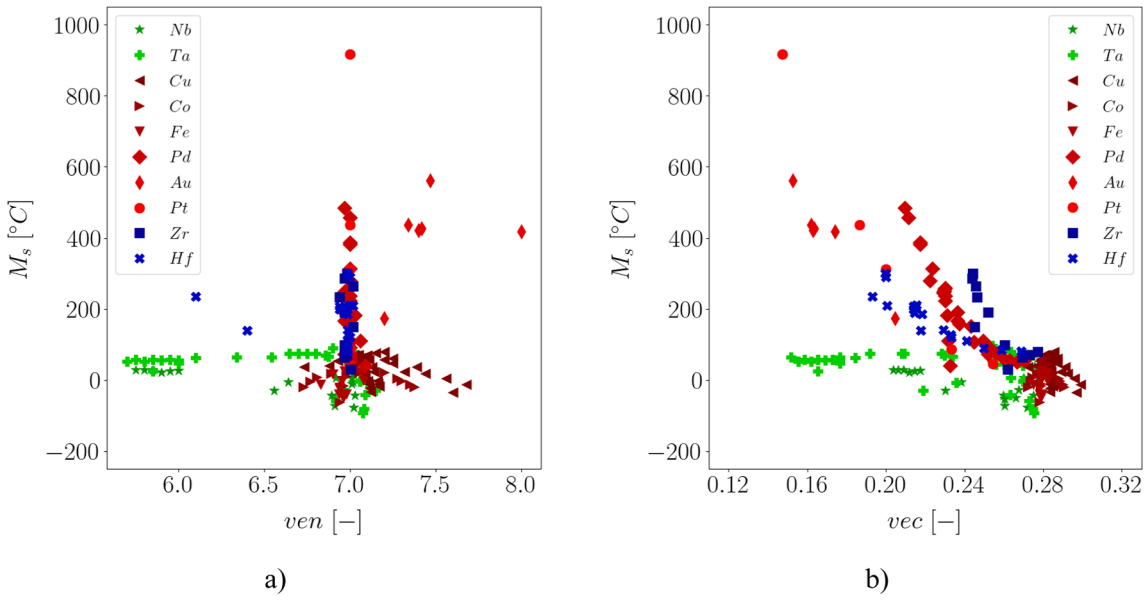


Fig. 4. Effect of the alloying element, for ternary alloys, on the Ms temperature plotted a) against ven and b) vec. The substitution element to Ni and Ti (ven close to 7) combined with the lowest vec allow to maximize the Ms temperature. Conversely, a huge proportion of Nb or Ta elements bring about a saturation of the Ms value around 50 °C.

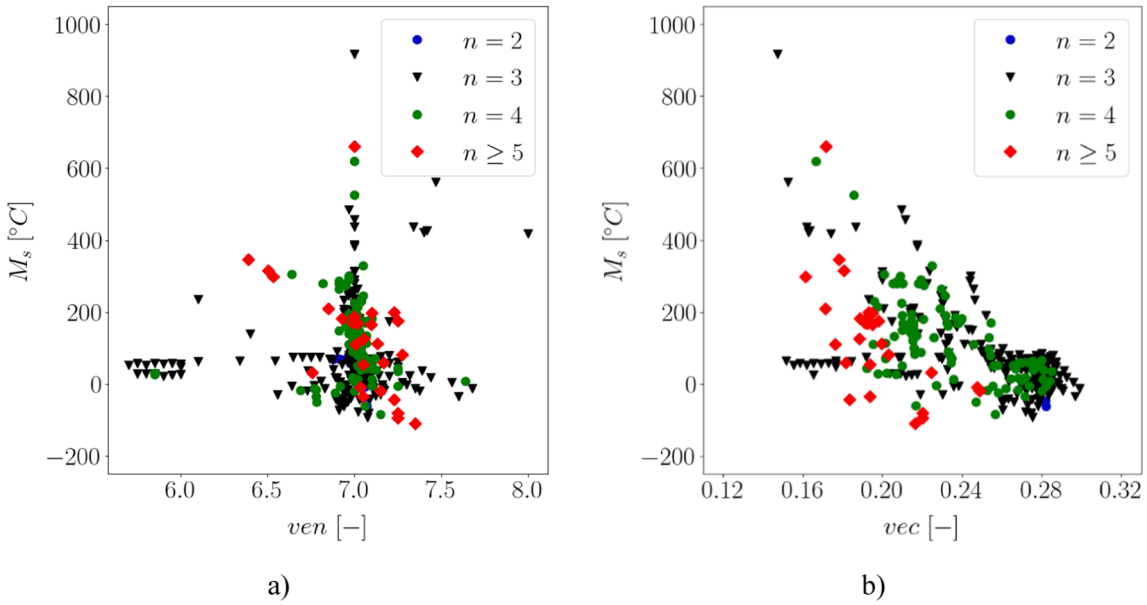


Fig. 5. Effect of the number of elements combined to a) ven and b) vec on the Ms temperature. Conversely to ternary alloys, the effect of vec for more than 4 elements is less pronounced since an alloy with 5 elements can lead to a Ms less than 0 °C even for a low vec.

saturate Ms to around 50 °C. Fig. 4b shows the effect of the *vec*, defined as the ratio of *ven* and the average atomic number for ternary alloys. It is observed that the less *vec* is the higher Ms is. To sum up, for ternary alloys containing Ni or Ti equivalent elements, it can be concluded that their Ms increases with the atomic number. Therefore, Ni-equivalent elements with a close atomic number like Fe, Cu or Co will not induce a high Ms compared to that reached with Pd, Au or Pt.

The influence of the number of elements combined to *ven* and *vec* is investigated as depicted in Fig. 5a and Fig. 5b. Conversely to ternary alloys, the effect of *vec* for more than 4 elements is less significant since other parameters must be taken into account for the Ms prediction. Indeed, an alloy with 5 elements can lead to Ms less than 0 °C even for a low *vec* of 0.18. Conversely, for such a *vec*, a ternary alloy reached Ms above 400 °C, as displayed in Fig. 4b.

Fig. 6 represents the effect of electronegativity on the Ms temperature in terms of the average value and the coefficient of variation. For the case of binary, ternary, and quaternary alloys, the Ms temperature plotted against the average electronegativity has a parabolic variation trend (Fig. 6a). The minimal Ms value of -70 °C is reached for an electronegativity of around 1.7 that corresponds to the binary alloy $Ni_{51.2}Ti_{48.8}$. Most of the alloys have an electronegativity below 1.7 since they contain elements with electronegativity lower than those of Ti such as Zr and Hf. For an electronegativity greater than 1.7, the Ms value increases strongly, thus reflecting the effect of alloying elements with higher electronegativities, such as Pd and Pt. Finally, for the case of more than 5 elements, there is again no clear tendencies anymore. Fig. 6b, shows the sensitivity of the Ms to coefficient of variation of electronegativity. For less than 4 elements, an increase of $\delta\chi$ will

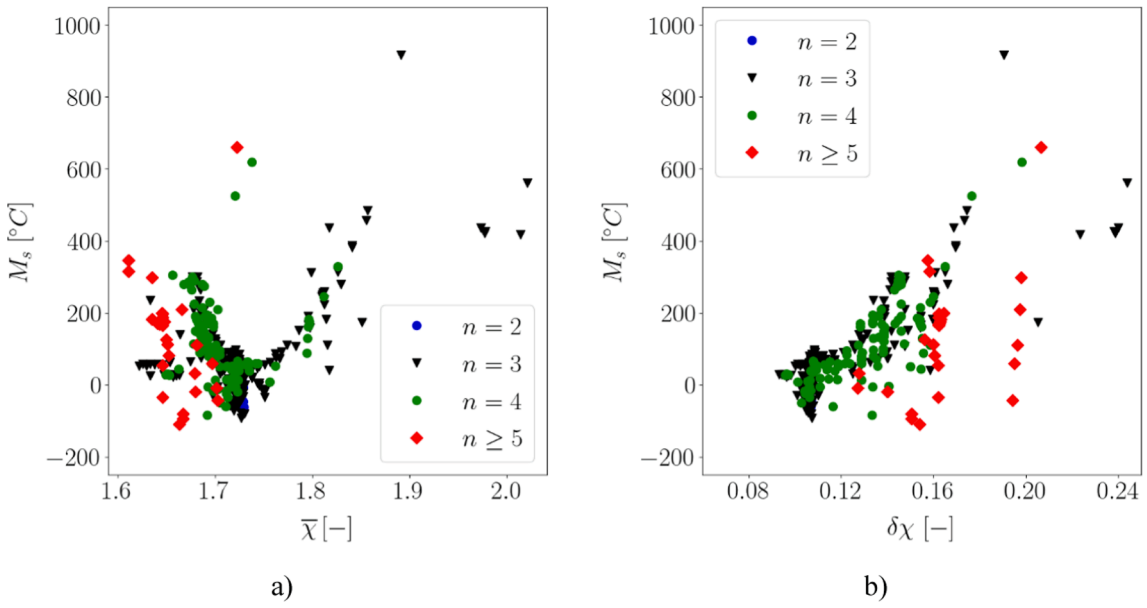


Fig. 6. Ms vs electronegativity a) average $\overline{\chi}$ and b) its variation coefficient $\delta\chi$. The average electronegativity $\overline{\chi}$ a) has a parabolic variation trend with a minimal value of 1.7 corresponding to the binary alloy $Ni_{51.2}Ti_{48.8}$. The variation coefficient of the electronegativity $\delta\chi$ has more pronounced effect on the Ms temperature in the case of high entropy alloys ($n \geq 5$) than ternary and quaternary alloys.

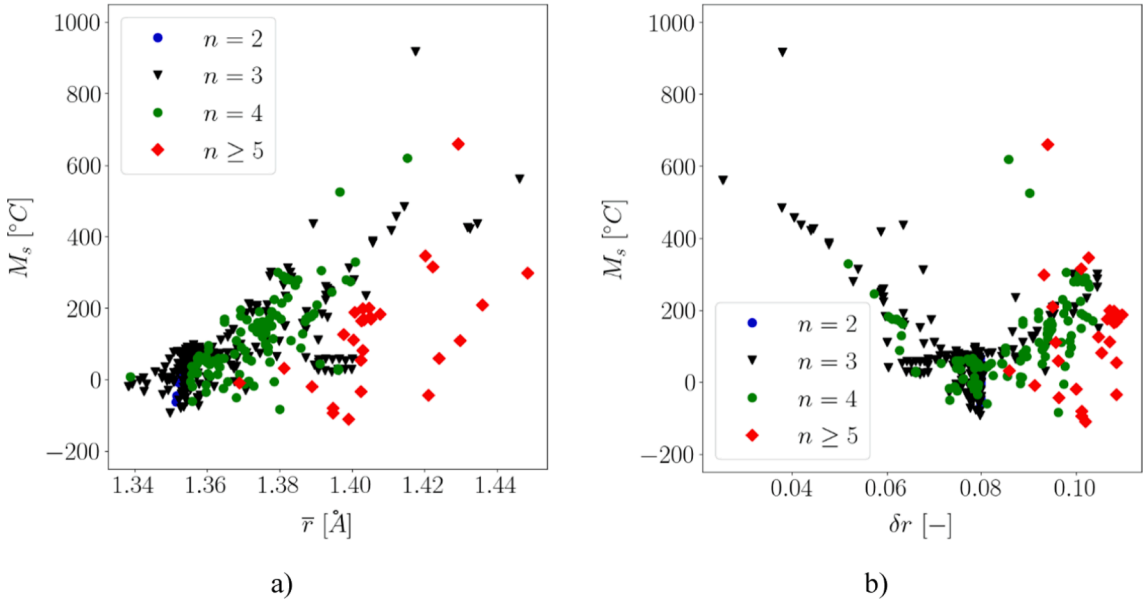


Fig. 7. Ms vs atomic radius a) average $\bar{r}$ and b) its variation coefficient δr. The Ms temperature exhibits a quasi-linear variation with respect to the average atomic radius $\bar{r}$. For the case of ternary alloys, a parabolic variation trend of Ms temperature is observed with the variation coefficient of the atomic radius.

increase the Ms temperature. In other words, the larger the mismatch between the electronegativities of the alloying elements, the higher the Ms can be reached. This could be correlated to the elements substituted to Ni and Ti whose electronegativities are far from their average value. For HEAs, this trend is less clear despite a much higher variation coefficient than ternary and quaternary alloys.

Fig. 7 shows the effect of the atomic radius on the Ms temperature in terms of the average value and the coefficient of variation. The effect of alloying elements induces an increase of the average atomic radius since the substitution alloying elements of Ni and Ti have a higher atomic radius (Fig. 7a). In fact, Pd, Pt and Au have a larger atomic radius than Ni, and Zr and Hf have a larger atomic radius than Ti (see appendix). Consequently, the increase of number of elements will increase the average atomic radius and the Ms temperature until four elements. For the HEAs alloys, there is no clear tendencies anymore. Fig. 7b represents the effect of the variation coefficient of the atomic radius. There is a parabolic variation trend similar to the average electronegativity, (Fig. 6a). The minimum of δr, up to 0.08 is reached for the reference equiatomic binary alloy $Ni_{50}Ti_{50}$. Similarly, to the electronegativity, the farther the coefficient of variation is from the reference value, the higher the Ms.

Finally, the effect of the mixing enthalpy is investigated and highlighted in Fig. 8. The tendency is more pronounced than the other materials descriptors, because regardless the number of elements, the greater the mixing enthalpy, the higher is the Ms. The increase of mixing enthalpy is explained by the combination of pair of elements with a high

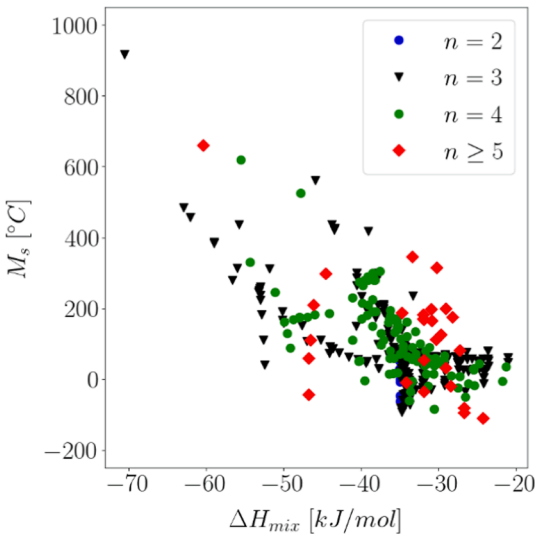


Fig. 8. Effect of the mixing enthalpy on the Ms temperature. The decrease of ΔH_{mix} induces an increase of the Ms temperature regardless the number of alloying elements.

mixing enthalpy such as Pd combined with Hf or Zr, for instance (Table B2). In fact, the alloys with a mixing enthalpy ranging between −40 and −20 kJ/mol, the Ms does not exceed 300 °C. Nevertheless, for a mixing enthalpy below −60 kJ/mol, the Ms is necessarily above 300 °C. For the case between −60 and −40 kJ/mol there are some discrepancies in the prediction of Ms, inducing a dependence on other parameters such as the critical number of elements, thus inhibiting the increase of Ms temperature.

To summarize the previous discussions, the following points can be drawn:

- The position of the substitution elements in the periodic table of the elements influences the Ms temperature. In fact, an alloy with a *vec* equal to 7, which means that the element is in the same column and has the lowest *vec*, as well as the highest average atomic number, is necessary to increase the Ms temperature. For instance, the substitution of Ni with Pd, Au, or Pt will drastically increase Ms temperature above 300 °C. Similarly, the substitution of Ti with Zr or Hf also increase Ms. Conversely, the elements Cu, Fe, Co have no effect on Ms temperature because they are very close to Ni with the same atomic number and similar physical properties.
- The calculation of the average and the variation coefficient of physical properties of the alloys (atomic radius, electronegativity, and mixing enthalpy) highlighted a strong correlation with the Ms temperature. Several physical material descriptors considering the mismatch between elements is essential for Ms prediction.

3. Results and analysis

3.1. Description of the machine learning algorithm

The Ms prediction algorithms are generally based on two machine learning algorithms either by neural network or by regression methods. The latter are more suitable for Ms prediction as the amount of data is not high enough to train a neural network. Various regression methods have been investigated such as linear regression, scalar vector regression (SVR), K-nearest neighbors (KNN), AdaBoosting, XGBoost, random forest (RF) and extremely randomized trees (ExRT) [11,42,45,52]. RF and ExRT, based on decision trees, are known to successfully interpolate high-dimensional inputs but are rather poor at extrapolating data beyond the learning domain. He et al. [42] and Honrao et al. [45] have shown that this method was particularly suitable for Ms temperature prediction.

The two algorithms both consist of selecting N decision trees and computing the prediction as the average of the predictions given by each decision tree, see Fig. 9. Note that another parameter is the depth of the decision tree. The differences between the two approaches are the selection and the building of decision tree, for more details, see [53] and [54]. ExRT is known to have lower variance compared to RF. Therefore, ExRT was selected and implemented for this study using scikit-learn python package [55]. The number of trees and the maximum depth were both set to 10 for this the study as a compromise between computation time and accuracy of the predictions (see Appendix C).

The ML algorithm proposed for this study was divided into several successive steps. The ML algorith flowchart is illustrated in Fig. 10. Firstly, the database was randomly divided into a portion for "training" inputs and the remainder for "validation" inputs. Since the amount of binary, ternary, quaternary, and high entropy alloys is not equal, a proportion of 80% is extracted for each class of alloy to be representative of the database. In addition, as the number of alloys in the database is small, 100 random splits have been performed using the same procedure to make the algorithm more robust. Secondly, the model has been trained on the "training" inputs and applied to the remaining "validation" inputs. Finally, 6 unpublished home-made HESMAs prepared specifically for this study are used as "test" cases (Table 2) to demonstrate the predictive capabilities of the developed physics-informed machine learning (PIML) based approach.

The mean absolute error (MAE) is taken as the scoring metric because it penalizes all errors equally contrary to root-mean-squared error. The MAE is given against each number of elements. The predictive ability of the model is discussed by averaging of the MAE over the 100 models obtained for the 100 split database samples. Moreover, the effect of the input data is also investigated by considering the composition of the 13 alloying elements, then the 11 physical mechanical descriptors and finally an optimized number of descriptors. The feature importance of each input data will be also given based on the Gini importance implemented in scikit-learn.

3.2. Sensitivity to the input data

The effect of the chemical alloying elements as input data is firstly investigated. The 100 models have led to an average MAE of 17 ± 1 °C for the training alloys, 35 ± 5 °C for the validation alloys and 48 ± 17 °C for the 6 HEAs of test. Fig. 11 provides more details on the distribution of errors sorted by the number of alloying elements (for the case of test and validation). The model predicts error increasing with the number of elements. In fact, for ternary and quaternary alloys the average error is 31 ± 6 °C and 32 ± 8 °C respectively that is under the limit of 50 °C. In addition, the greatest error is made with the validation HEAs whose error is 84 ± 35 °C. Fig. 11b shows the influence of the alloying elements in the Ms temperature prediction. Ni is the highest correlated parameter with Ms. Then, one can notice that Ti, Pd, Pt and Hf are then the four other elements the more correlated with Ms. This trend is coherent since Ni and Ti are the base elements of "NiTi-like" alloys and Pd, Pt and Hf increase Ms temperature, see Fig. 4. On the other side, Fe, Co, Ta, Nb have low effect on the Ms prediction that is also confirmed with the results presented in Fig. 4. Finally, Au, Cu and Zr have a slight effect on the transformation temperature.

Secondly, the effect of the 11 physical material descriptors is investigated. The average MAE is equal to 8 °C ± 1 °C for the training alloys, 29 ± 5 °C for the validation alloys and 37 ± 16 °C for the 6 HEAs of test. Similarly, to the case with the alloying elements as input data, the average error is increasing with the number of elements. Indeed, for ternary and quaternary alloys the error is 24 ± 5 °C and 30 ± 8 °C respectively. However, the error is more important for the HEAs of validation (77 ± 32 °C) and especially for the HEAs of test with 37 ± 16 °C respectively. Fig. 12 shows the influence of the material descriptors depicting that the coefficient of variation of electronegativity

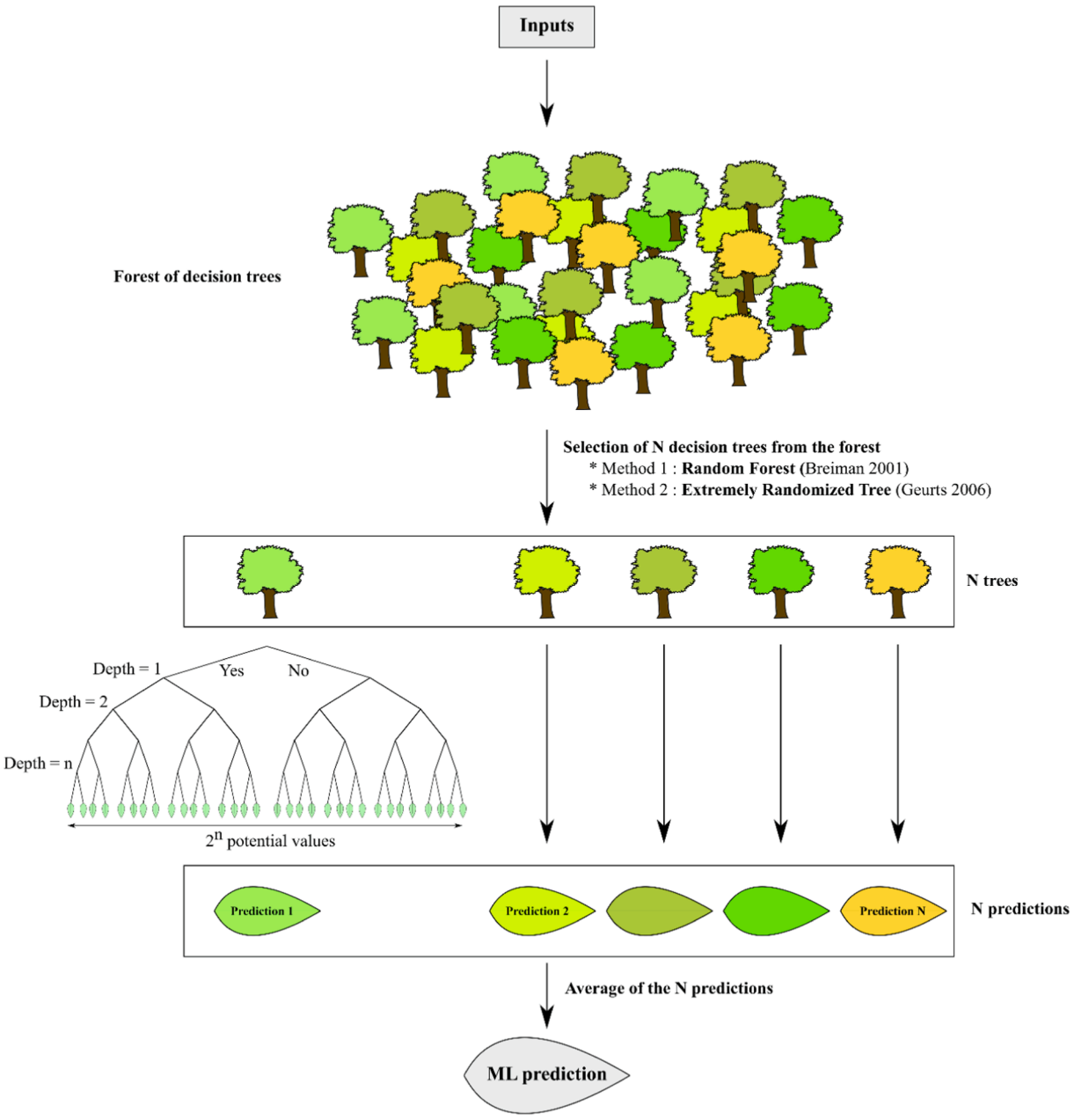


Fig. 9. Flow chart of the Decision Tree based algorithms (RF and ExRT). The hyperparameters of this model are the number of decision trees and the maximum depth of each tree. The number of trees and the maximum depth were both set to 10 for this the study as a compromise between computation time and accuracy of the predictions (see Appendix C).

$\delta\chi$ and the mixing enthalpy $\Delta Hmix$ contribute to more than 50 % to the Ms prediction as already assumed (Fig. 6 and Fig. 8). Then, *vec*, δr, $\overline{\chi}$ and $\overline{r}$ are the four other most influent material descriptors. On the contrary, the number of elements and *ven* are poorly correlated to Ms prediction despite an indirect effect. Similarly, the proportion of Ni-like, Ti-like and Nb-like are no influent since their proportions are quite constants for each alloy with values around 50 %, 50 % and 0 % respectively.

Fig. 13 shows the sensitivity of the number of material descriptors on the average MAE for the "validation" and "test" alloys. Based on the results using the whole material descriptors, the latter were ranked hierarchically according to their feature importance on the Ms predictions (see Fig. 12). A number k of descriptors refers to the k most important descriptors in the Ms prediction. For example, three descriptors correspond to δX, ΔHmix and *vec* that have the three highest feature importance (see Fig. 12).

Fig. 13 shows that the Ms predictions based on the three most important descriptors (δX, ΔHmix, vec) provide MAE below the acceptable value of 50 °C for the "validation" of ternary, quaternary alloys, as well as for the "test" of HEAs. However, the MEA appears above the acceptable limit of 50 °C for the "validation" HEAs. This error exhibits a low decrease when adding the other descriptors as inputs in the prediction, namely: $\overline{r}$, δr, $\overline{X}$, n, *ven*, $\overline{f}_{Ni}$, $\overline{f}_{Ti}$ and $\overline{f}_{Nb}$ due to the intercorrelation between some descriptors (see Fig. 2). However, for "validation" HEAs, the error does not decrease regardless of the number of inputs. As a compromise between good prediction and minimizing the intercorrelation between descriptors, a choice of 6 descriptors is considered.

3.3. Toward optimization of Ms for ternary, quaternary, and quinary alloys

A model with the 6 most important physical descriptors, *i.e.,* the

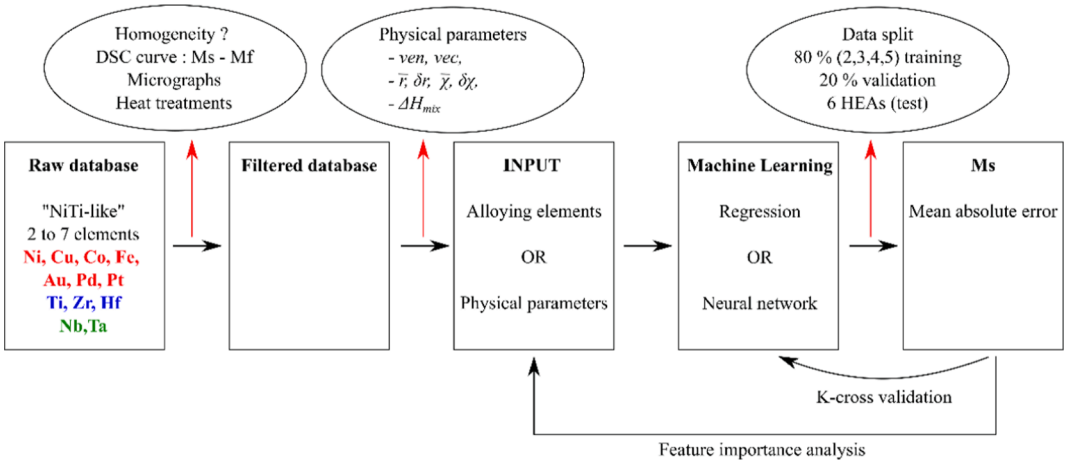


Fig. 10. Flow chart of the general machine learning algorithm used for this study.

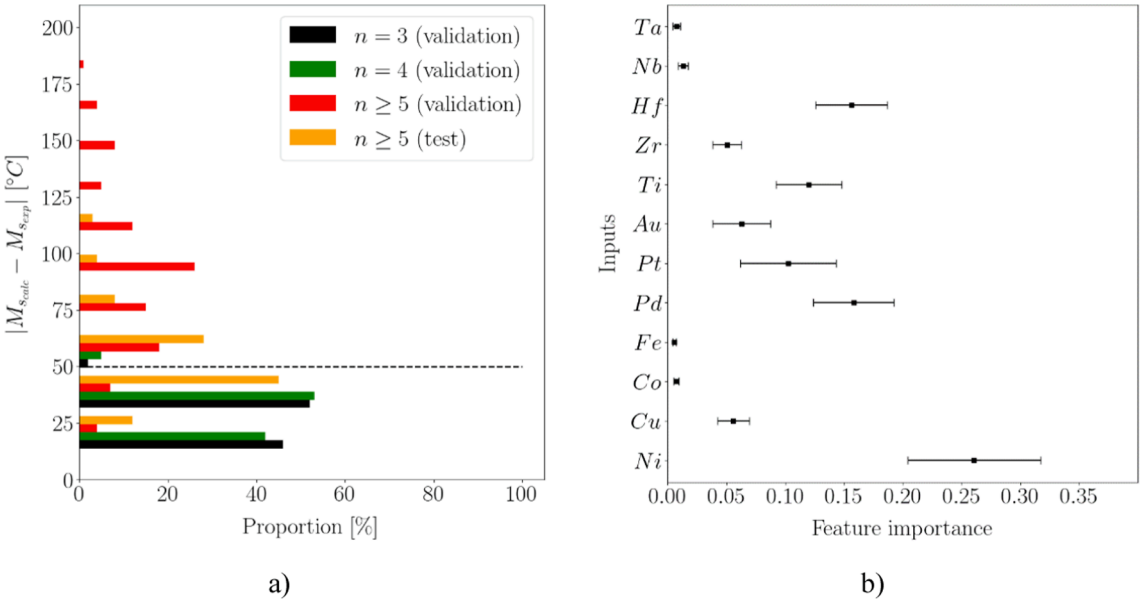


Fig. 11. Effect of the alloying elements on the Ms temperature prediction: a) MAE for validation and test data, b) feature importance.

radius ($\overline{r}$ *and* δr), the electronegativity ($\overline{\chi}$ *and* $\delta\chi$), the mixing enthalpy $\Delta Hmix$, and the valence electron concentration *vec* have been considered sufficient to predict Ms temperature (see Appendix C). Fig. 14 shows the Ms predicted against the experimental Ms with a representative convenient split database. This model provides a very reliable and robust prediction with an average error of less than 30 °C for the validation and the test alloys.

The predictive capabilities of the model are now demonstrated through the design of HESMAs with a target Ms temperature. As an illustrative case, the virtual fabrication of HESMAs containing Cu, Zr and Pd is proposed to reach a Ms temperature above 400 °C. At first, the model is applied for ternary alloys and its predictions have been compared with experimental results from literature (Fig. 15). Then, the model is extrapolated to quaternary (Fig. 16) and quinary alloys (Fig. 17) to determine the chemical composition required to reach the target temperature of 400 °C.

Fig. 15 illustrates the comparison of the Ms predicted by the model and the experimental Ms of ternary alloys with Cu, Zr and Pd. For the three ternary alloy families, the predictions are in very good agreement with the experimental results. For the case of NiTiCu (Fig. 15a and Fig. 15b), similar to the experimental results, it is emphasized that the Cu addition has no significant effect on the Ms temperature. This would explain that Cu has been widely substituted for Ni in the design of HEAs. Fig. 15c and Fig. 15d show the Ms prediction for NiTiZr alloys. The model highlights that this family of alloys can reach an Ms temperature between 200 °C and 300 °C for a minimum Zr content of 15 at.% and a Ti content such that Ti-like is greater than 50 at.%. Fig. 15e and Fig. 15f show that the addition of Pd can increase the Ms above 300 °C when the fraction is greater than 30 at.%, as observed experimentally for a Ti-like fraction of 50 at.%.

Fig. 16 presents the case of the optimization of quaternary alloys such that $(NiCu)_{50}(TiZr)_{50}$, $(NiPd)_{50}(TiZr)_{50}$ and $(NiCuPd)_{50}Ti_{50}$ alloys. Concerning the $(NiCu)_{50}(TiZr)_{50}$, the model predicts that Ms can reach more than 200 °C for Cu content below 20 at.% with simultaneously a Zr content greater than 30 at.%, (Fig. 16a). If Pd is replaced by Cu, *i.e.,* the case of $(NiPd)_{50}(TiZr)_{50}$ (Fig. 16b), Ms could reach very high temperature above 400 °C if Pd and Zr contents are sufficiently important. For the case combining Cu and Pd, *i.e.* $(NiCuPd)_{50}Ti_{50}$ (Fig. 16c), it would require a very high amount of Pd, greater than 40%, to reach temperature above 400 °C. Reversely, for lower Pd content (at.%), Ms temperature remains less than 100 °C.

Finally, the $(NiCuPd)_{50}(TiZr)_{50}$ HESMAs family is investigated to

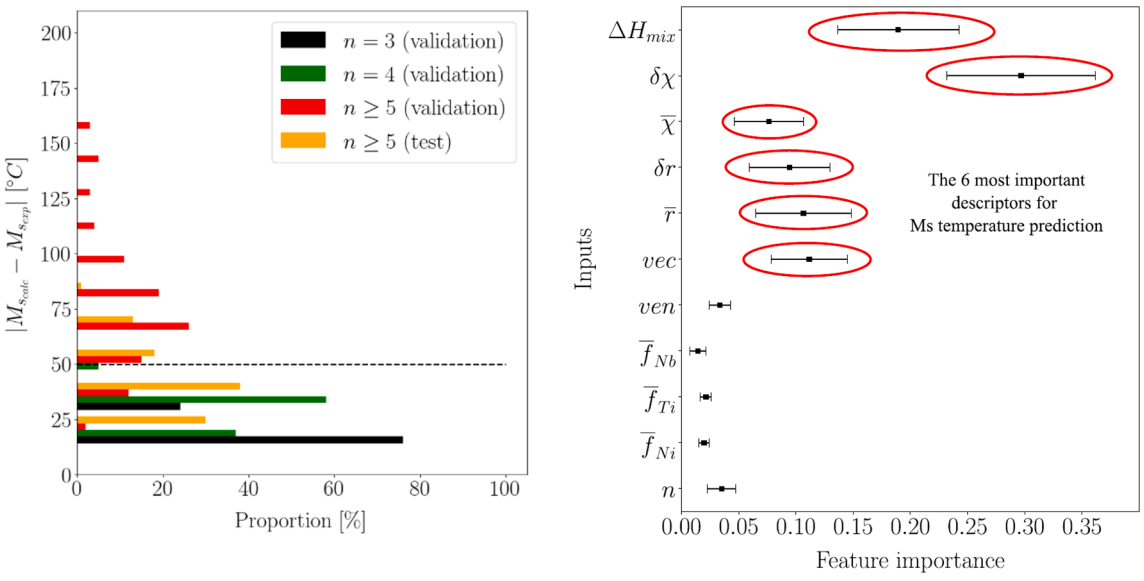


Fig. 12. Effect of the physical material descriptor for the Ms temperature prediction: a) MAE for validation and test data, b) feature importance. The 6 most important material descriptors for the Ms temperature prediction, namely δX, ΔHmix, vec, $\overline{r}$, δr and $\overline{X}$ are circled in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

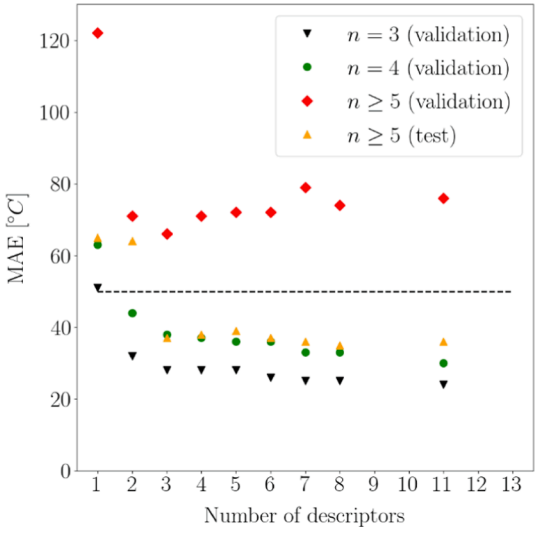


Fig. 13. Effect of the number of inputs for the error in the Ms temperature prediction. A number k of descriptors refers to the k most important descriptors in the Ms prediction. For example, three descriptors correspond to δX, ΔHmix and vec that have the 6 highest feature importance (see Fig. 12).

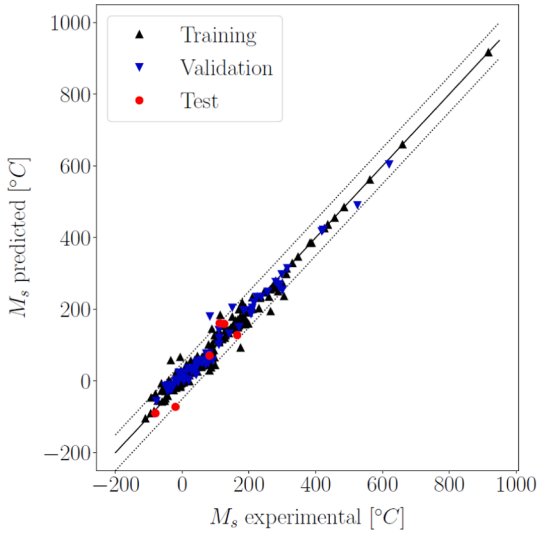


Fig. 14. Ms calculated vs Ms experiment with 6 material descriptors.

optimize the chemical composition range toward reaching a target Ms with specific conditions. As an illustrative case, let's consider a target Ms above 400 °C combined with some conditions on Cu, Pd and Zr contents. The elaboration of HEAs requires a proportion between 5 and 35 at.%. In this example, a more restrictive condition of 25 at.% is also considered to minimize the cost of the pure elements. Fig. 17 illustrates the methodology of optimization of the $(NiCuPd)_{50}(TiZr)_{50}$ HESMAs based on the proposed specifications requirement. Ms evolution was firstly plotted against the different combinations Cu, Pd and Zr contents satisfying only the condition of HEAs elaboration. It is noticed that the greater is the Cu content, the lower Ms is since the potential remaining amount of Pd would be less important to have a Ni-equivalent proportion $\overline{f}_{Ni} = f_{Ni} + f_{Cu} + f_{Pd}$ to 50 at.%. In addition, Ms is maximized if the amount of Zr and Pd are both high enough as already explained for the quaternary system $(NiPd)_{50}(TiZr)_{50}$ (Fig. 16b).

Then, among all the proposed HESMAs compositions, the alloys that did not fulfill the specifications requirements have been removed that led to a restricted range of alloys (Fig. 17). It appeared that for having Ms above 400 °C, it requires more than 10 at.% Zr, at least 15 at.% Pd and a variable proportion of Cu depending the composition of Zr and Pd.

4. Discussion

4.1. Relevance of the materials descriptors

The results obtained in the section have demonstrated a strong effect of the inputs for the Ms prediction. Firstly, all the alloying elements, *i.e.*, 13 inputs have been considered. It led to small MAE below 30 °C for all

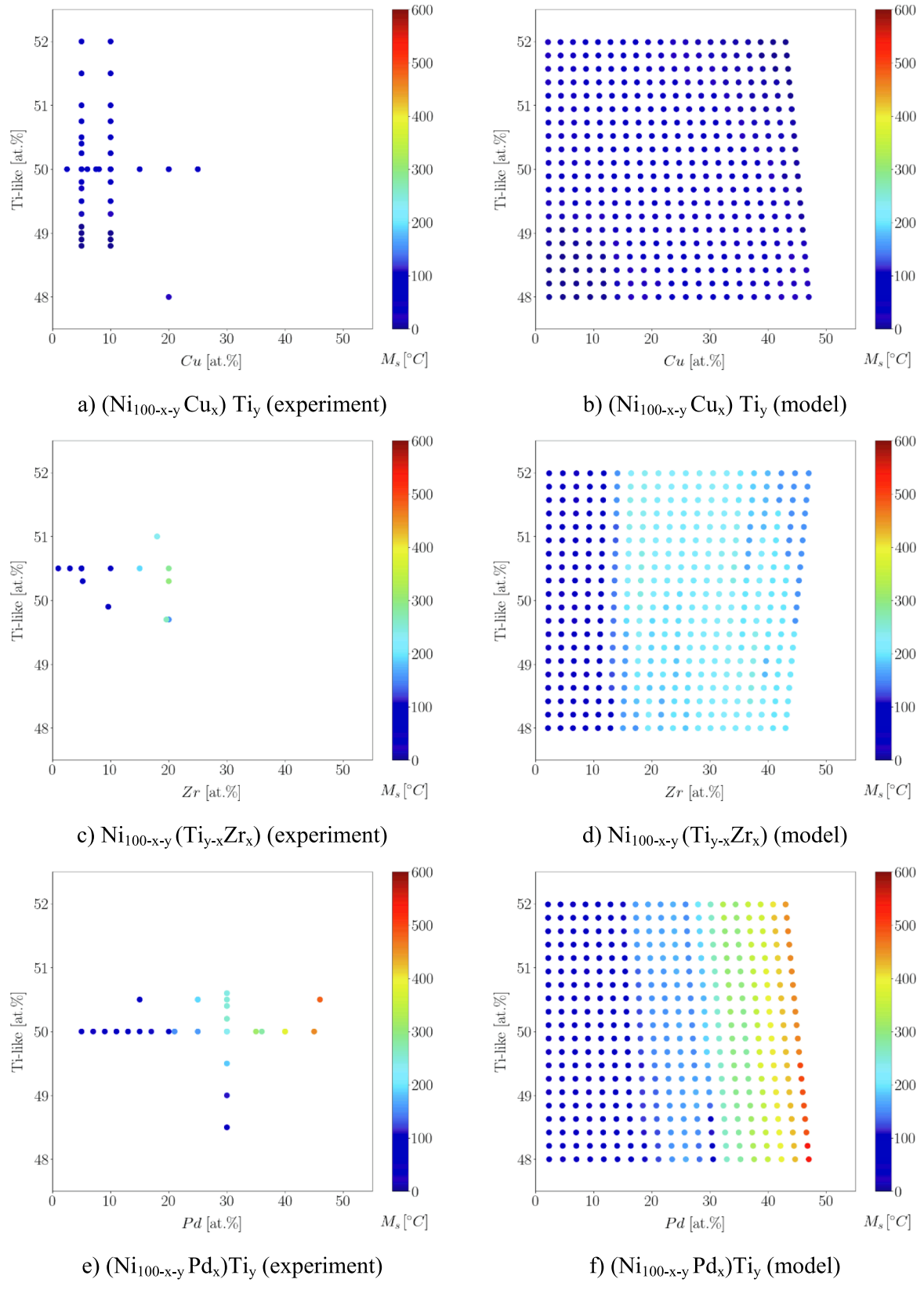


Fig. 15. Ms experimental vs Ms predicted for ternary alloys (NiCu)Ti, Ni(TiZr) and (NiPd)Ti. Note that, in all the graphs, the colorbars stand for the Ms temperature ranging between 0 and 600 °C.

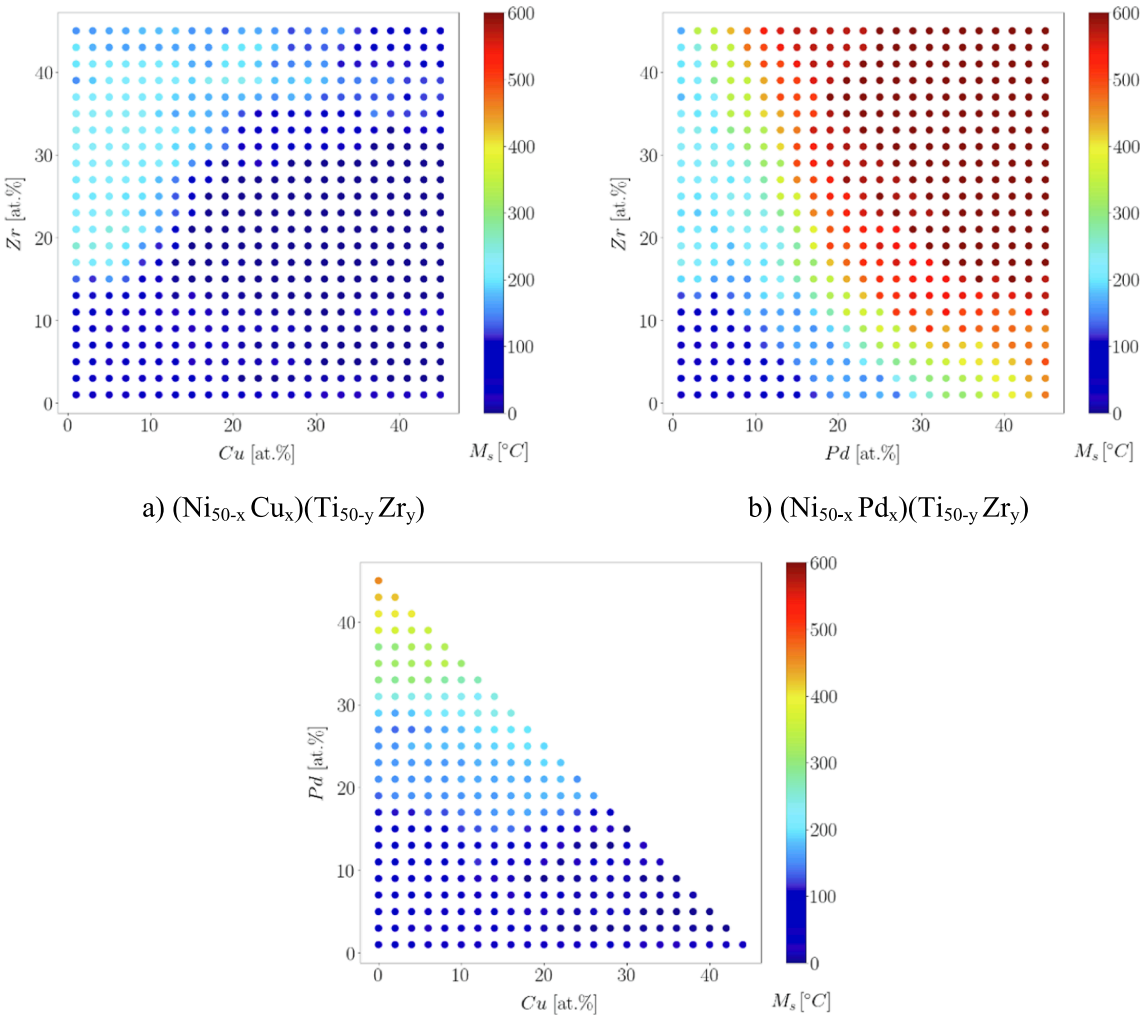


Fig. 16. Ms prediction for quaternary alloys $(NiCu)_{50}(TiZr)_{50}$, $(NiPd)_{50}(TiZr)_{50}$ and $(NiCuPd)_{50}Ti_{50}$.

the alloys containing less than 4 elements. Moreover, the study of feature importance depicted that Fe, Co, Nb and Ta does not have a strong influence on Ms prediction. Reversely, after Ni and Ti, Pd, Pt and Hf are the most influent parameters on Ms prediction. Indeed, those elements are already known to increase Ms.

Secondly, 11 physical material descriptors were then considered as inputs for the Ms prediction. Comparing the "alloying elements" with the "physical material descriptors" used as inputs, the error was slightly smaller for alloys with no more than 5 elements. Note that for the HEAs used as test alloys, the prediction was also better using the "physical material descriptors" as inputs. In addition, the feature importance analysis showed that at least 6 parameters are sufficient for having a good Ms prediction. The inputs are the atomic radius (mean and variation coefficient), the electronegativity (mean and variation coefficient), the valence electron concentration and the mixing enthalpy.

The introduction of the alloying elements from one side and the physical material descriptors gave two complementary approaches for Ms analysis and prediction. On the one hand, the use of alloying elements allowed to distinguish the importance of each element in the Ms prediction without any physical data. On the other hand, the material descriptors provided some physical properties to better understand the effect of specific elements. For instance, it is known that Pd is generally used to replace Ni for a Ms above 400 °C. This effect is probably related to the proximity of the Pd and Ni elements combined with a higher atomic number, higher electronegativity and higher mixing enthalpy with elements compared to Ni. Conversely, to reach a temperature above 400 °C, a combined set of material descriptors is required that does not agree with all alloying elements. For this temperature range, the addition of Pd is often required.

4.2. *Effect of the number of elements*

Unlike the feature importance analysis, the number of elements is important for the understanding of the Ms prediction. For instance, increasing the number of elements increases the error, especially when the number of elements is greater than 5. The first assumption is the criteria of homogeneity of the alloys. The alloys containing up to 4 elements are generally quite homogeneous even without heat treatment. In these cases, the predicted error is small. Contrarywise, Peltier et al. [11] have argued that HESMAs must always be heat treated for a sufficient time and at a high enough temperature. Otherwise, the measurements of HESMAs properties are meaningless and would depend on the preparation process itself. This explains the increasing error with the number of elements. The homogeneity criteria would require the analysis of micrographs which is time-consuming but more efficient. Indeed, the 6 homemade HEAs, have shown the lowest errors and validation of the proposed methodology. Moreover, the criteria of homogeneity allowed a first filtering of the results.

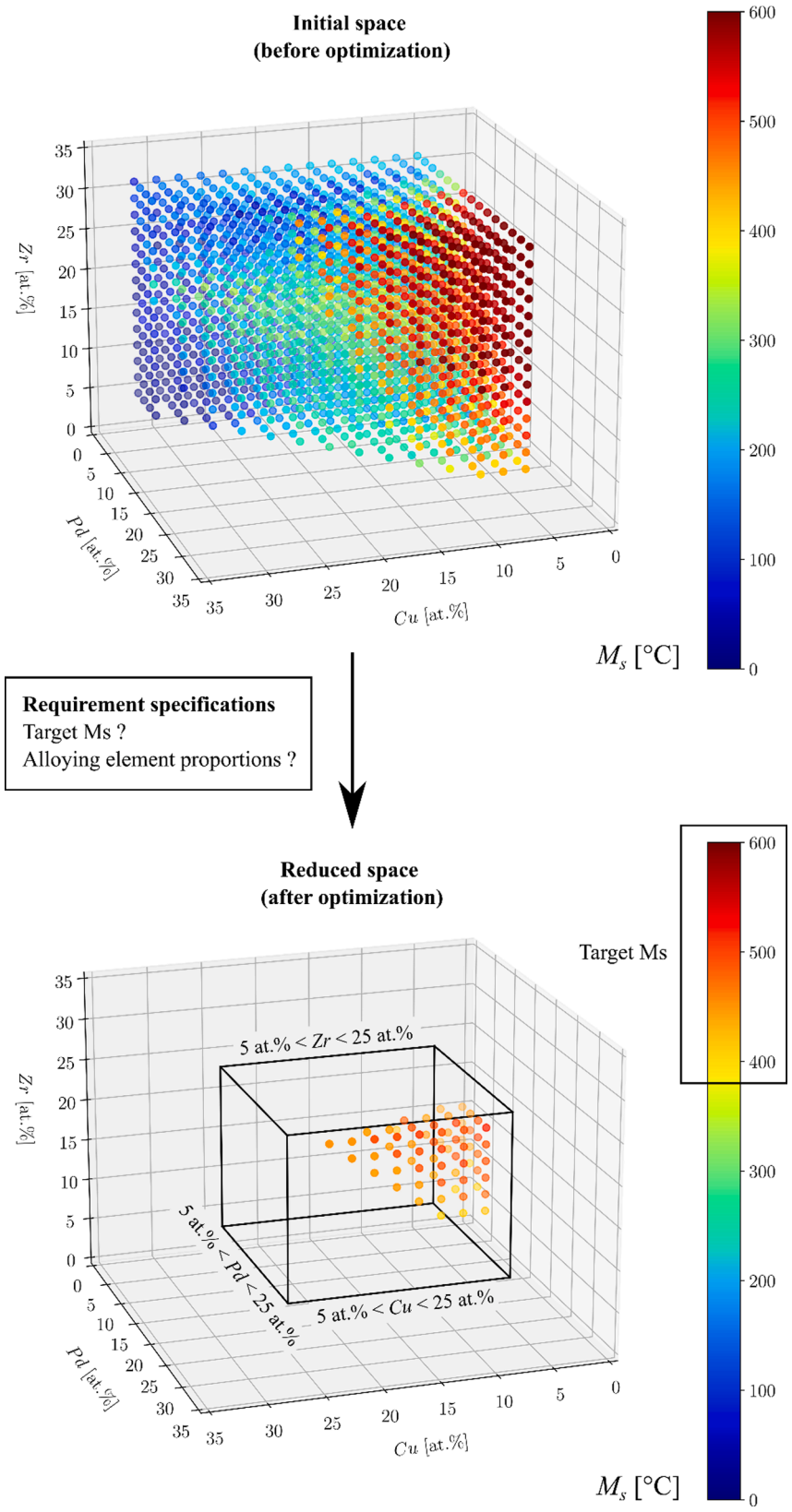
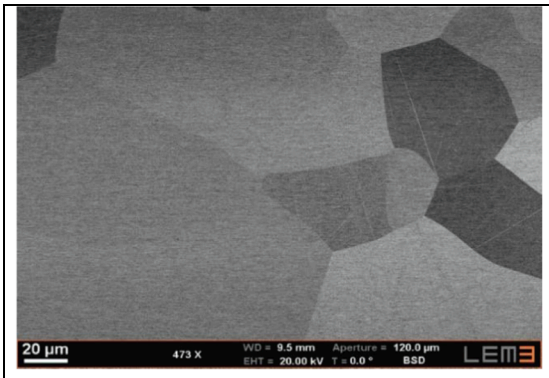
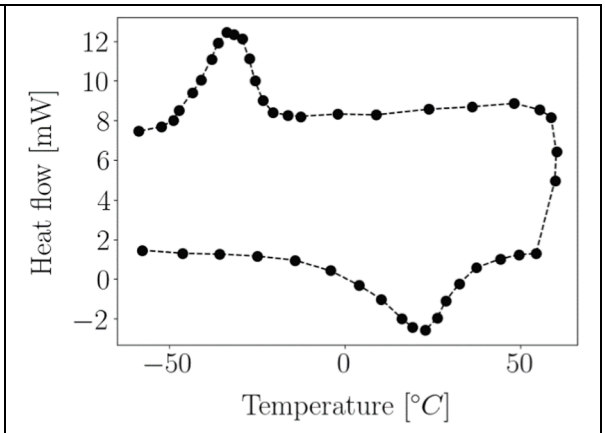


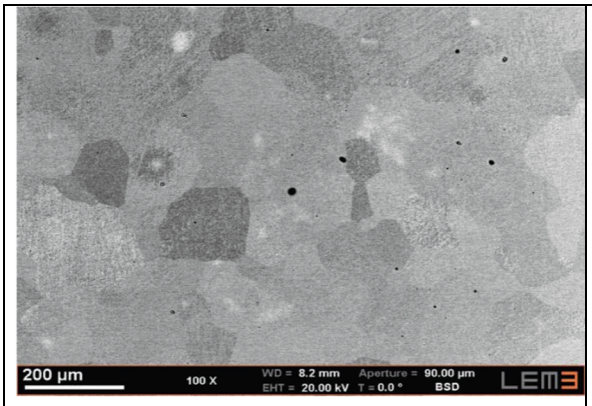
Fig. 17. (Ni_{50-x-y} Cu_xPd_y)(Ti_{50-z} Zr_z) HESMAs design with the following requirement specifications: Ms greater than 400 °C and the alloying elements fraction of Cu, Pd and Cu below 25 at.%.

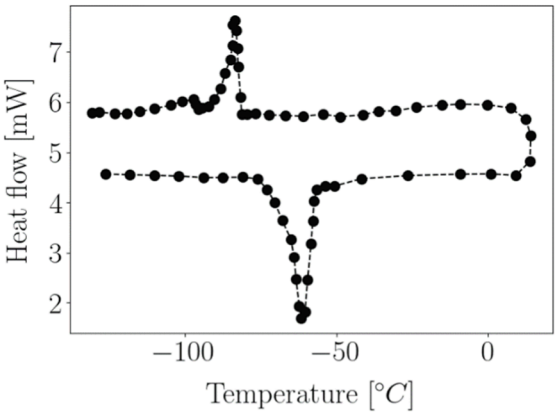




a) Alloy 1 $(Ni_{25}Cu_{20}Co_5)(Ti_{30}Zr_{20})$

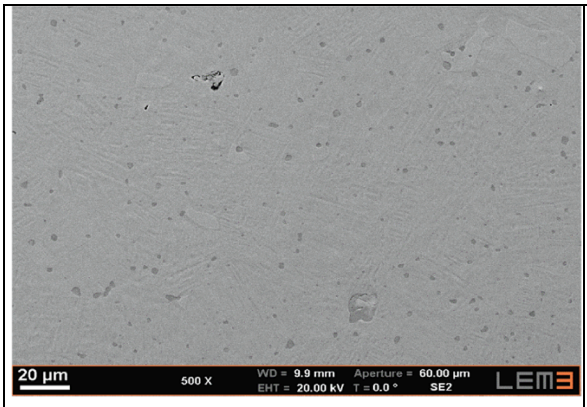
b) Ms = -20 °C and Ms - Mf = 25 °C

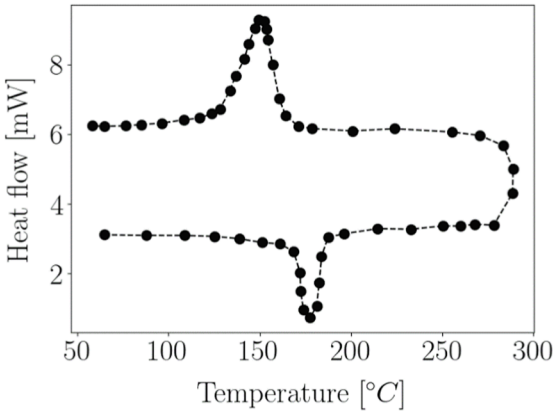




c) Alloy 2 $(Ni_{25}Cu_{25})(Ti_{25}Zr_{15}Hf1_{10})$

d) Ms = -80 °C and Ms - Mf = 15 °C





e) Alloy 3 $(Ni_{25}Cu_{17.5}Co_{7.5})(Ti_{16.7}Zr_{16.7}Hf_{16.7})$

f) Ms = 165 °C and Ms - Mf = 39 °C

Fig. 18. Micrographs at room temperature and the corresponding DSC curves of the 6 unpublished homemade prepared and homogenized HESMAs specifically for this study.

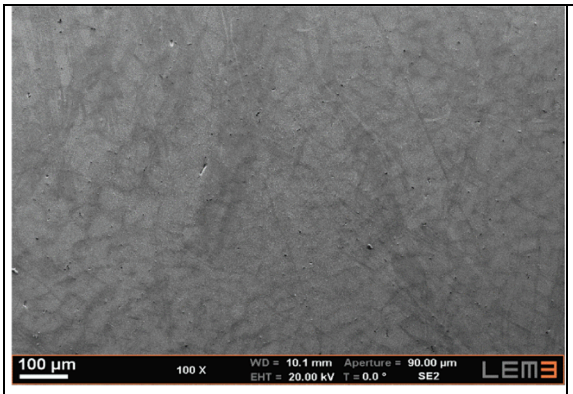
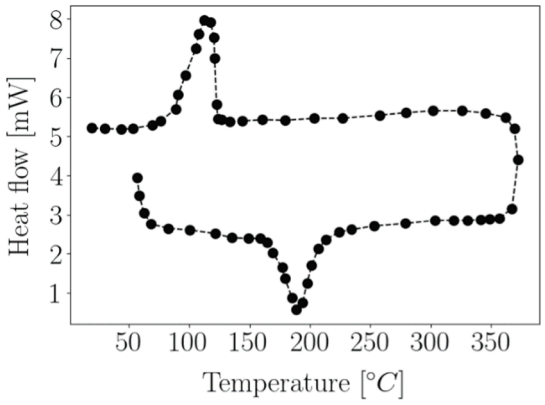
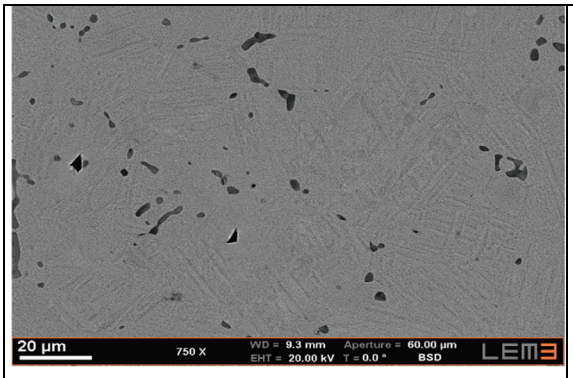
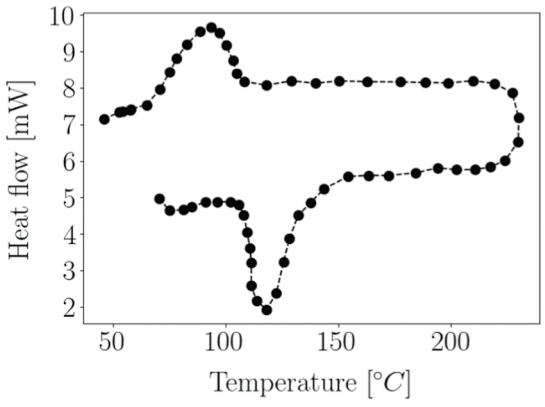
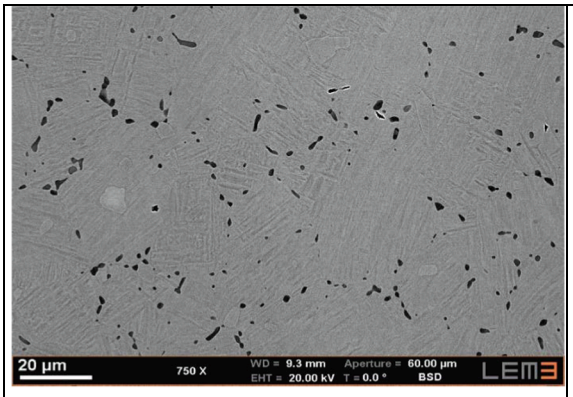
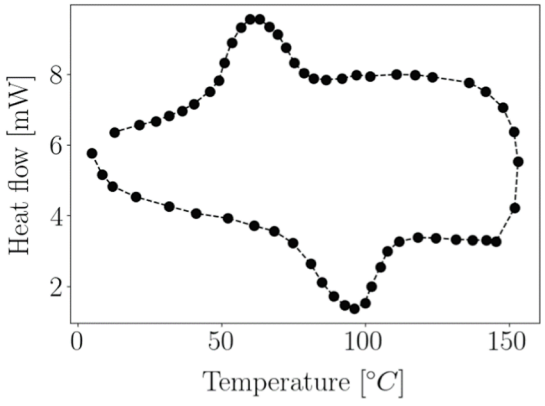
	
g) Alloy 4 $(Ni_{24.8}Cu_{15.5}Co_{8.5})(Ti_{16.5}Zr_{15.5}Hf_{13.5})Ta_{5.7}$	h) Ms = 126 °C and Ms - Mf = 48 °C
	
i) Alloy 5 $(Ni_{24.7}Cu_{18.3}Co_{7.4})(Ti_{17.9}Zr_{16.4}Hf_{15.3})$	j) Ms = 112 °C and Ms - Mf = 45 °C
	
k) Alloy 6 $(Ni_{23.8}Cu_{26.4})(Ti_{18.1}Zr_{16.5}Hf_{15.2})$	l) Ms = 82 °C and Ms - Mf = 35 °C

Fig. 18. (*continued*).

The lack of reference in the literature for HEAs is the second explanation for the bad predictions, since the database for ternary and quaternary alloys has not sufficiently enriched the database. In fact, the effect of the splitting data of HEAs showed that depending on the choice of the alloys, there are huge discrepancies in the results. Finally, the transition from 4 to 5 elements greatly increases the entropy of the systems and is favorable to more unpredictable results such as the cocktail effect.

Table B1
Properties of the pure elements used in this study.

	Ni	Cu	Co	Fe	Pd	Pt	Au	Ti	Zr	Hf	Nb	Ta
Atomic number	28	29	27	26	46	78	79	22	40	72	41	73
Valence electron	10	11	9	8	10	10	11	4	4	4	5	5
Atomic radius [Å]	1.246	1.278	1.251	1.241	1.375	1.387	1.442	1.462	1.603	1.578	1.429	1.43
Electronegativity	1.91	1.9	1.88	1.83	2.2	2.38	2.54	1.54	1.33	1.3	1.6	1.5

4.3. A strong tool for NiTi-like alloys design

The present study proposed a strong algorithm to predict Ms temperature for NiTi-like alloys. The relevance and the scientific interest of such a tool is demonstrated for the high temperature HESMAs. This model is then implemented to the design of new HESMAs alloys to meet a challenging issue to reach Ms above 400°C. As observed in the literature, Pd is a good candidate for the elaboration of such HESMAs. The predictive capabilities of the model to elaborate HESMAs with Ms above 400 °C are demonstrated using an illustrative case of $(NiCuPd)_{50}(TiZr)_{50}$ HESMAs. Other specification requirements on the chemical element ranges also constrained the model and provide relevant indications on the composition of HESMAs.

The developed methodology can be easily generalized to other elements known to also increase Ms such as Au or Pt for Ni substitution or Hf for Ti substitution. In addition, other specifications regarding phase stability, phase structure, the processing costs should also be incorporated into the algorithm for an industrial application.

5. Conclusions

A database of NiTi-like HESMAs has been created and has been enriched with alloys of less complex chemical composition (binary, ternary and quaternary). The main novelty of this database is the addition of a metallurgical criterion designed to maintain only homogenized alloys. It brings a metallurgical point of view to the data, which guarantees a greater robustness of the experimental data from different research works.

The Ms evolution has been analyzed through two complementary points of view, the alloying elements, and the material descriptors that are of first order of importance for Ms temperature prediction. These 6 parameters are founded on the metallurgy and related to: the mismatch between each element in terms of mixing enthalpy, atomic radius, electronegativity, and the atomic number.

Furthermore, the Ms temperature prediction is related to the number of alloying elements, in particular, for binary and ternary alloys. However, it has been observed that the transition from medium to high entropy (from 4 to 5 elements) brings about some unclear trends in the evolution of Ms temperature. This complex evolution of Ms with respect to the number alloying elements can be explained by the cocktail effect in HEAs. The use of material descriptors instead of the chemical elements brought better physical explanations for Ms prediction.

The ExRT algorithm is a promising ML algorithm for Ms prediction since the errors are very low (below 30 °C) for alloys containing up to 4 elements. For more elements, the errors increased significantly and could probably be explained by a lack of homogenized HESMAs leading to the uncertainty of Ms prediction. Conversely, the Ms temperature is predicted quite well for the home-made HESMAs, defined in this study as "test" alloys, whose homogeneity has been systematically checked by SEM and DSC measurements.

Based on the predictions of the model, a powerful tool was developed to design HESMAs with a desired Ms temperature and specification requirements. As an illustrative case, the design of $(NiCuPd)_{50}(TiZr)_{50}$ HESMAs was carried out and resulted in compositions with Ms above 400 °C including other requirement specifications.

The model revealed a significant sensitivity to the number of alloying elements whose Ms prediction decreases sharply for HEAs. To address this issue, thermodynamic calculations [28] are planned to virtually define and extend a larger database. In the future, to fill the gap regarding the lack of certain combined compositions, the present work will be extended and enriched by introducing experimental data from 4D printing combined with thermodynamic calculations. This prospective will open the way for an innovative design of HESMAs towards high temperature applications avoiding the use of expensive elements such as Pd, Pt or Au elements.

Finally, this methodology approach could be extended to other alloying elements such as Mn, Al-like (Al, Zn, Ga, In) and Mo-like (Mo, V, Cr, W) elements without adding a significant number of parameters. In addition, the entire methodology can be applied to other families of SMAs such as TiNb, CuAl or FeMn-like alloys.

CRediT authorship contribution statement

Léo Thiercelin: Writing – original draft, Methodology, Formal analysis, Investigation, Validation. **Laurent Peltier:** Data curation, Writing – review & editing, Formal analysis, Investigation. **Fodil Meraghni:** Conceptualization, Methodology, Writing - review & editing, Formal analysis, Supervision. Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial

Table B2
Mixing enthalpy table of the studied couple of alloying elements.

	Ni	Cu	Co	Fe	Pd	Pt	Au	Ti	Zr	Hf	Nb	Ta
Ni	0	4	0	-2	0	-5	7	-35	-49	-42	-30	-29
Cu	4	0	6	13	-14	-12	-9	-9	-23	-17	3	2
Co	0	6	0	-1	-1	-7	7	-28	-41	-35	-25	-24
Fe	-2	13	-1	0	-4	-13	8	-17	-25	-21	-16	-15
Pd	0	-14	-1	-4	0	2	0	-65	-91	-80	-53	-52
Pt	-5	-12	-7	-13	2	0	4	-74	-100	-90	-67	-66
Au	7	-9	7	8	0	4	0	-47	-74	-63	-32	-32
Ti	-35	-9	-28	-17	-65	-74	-47	0	0	0	2	1
Zr	-49	-23	-41	-25	-91	-100	-74	0	0	0	4	3
Hf	-42	-17	-35	-21	-80	-90	-63	0	0	0	4	3
Nb	-30	3	-25	-16	-53	-67	-32	2	4	4	0	0
Ta	-29	2	-24	-17	-52	-66	-32	1	3	3	0	0

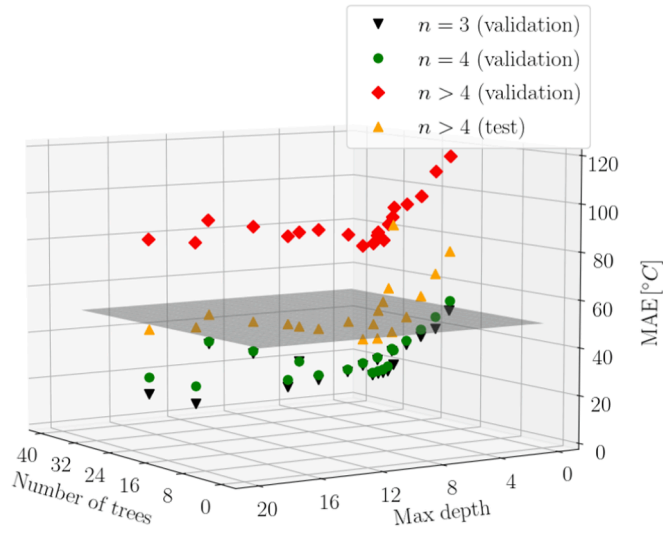


Fig. 19. Sensitivity of the hyperparameters, the max depth and the number of trees on the Ms temperature prediction. The good compromise chosen is 10 trees and a maximum depth of 10 nodes for each decision tree.

interests or personal relationships that could have appeared to influence the work reported in this paper. The authors acknowledge the use of the Cassiopee Arts et Métiers Institute of Technology HPC Center made available for conducting the research reported in this paper.

Data availability

The database and the related references can be found at: https://data.mendeley.com/datasets/z6j6g3b9yf/draft?a=00b9d4d9-3e19-4552-b670-cf763375cc25

Appendices A.

Appendix A. Homemade HEAs: micrographs and DSC curves.

Fig. 18 illustrates the micrographs (a,c,e,g,i,k) and the related DSC thermograms (b,d,f,h,j,l) of the 6 developed HEAs exploited as "test" alloys for the proposed ML algorithm. The microstructural observations combined with transformation temperature measurements emphasized the homogenized state of the studied alloys. This confirms the metallurgical criterion for the database filtering (see Table B1-B2).

Appendix B. Properties of the alloying elements used in this study

Appendix C. Optimization of the ML parameters.

C.1. Sensitivity study of the ExRT hyperparameters.

Fig. 19 reflects the effects of the number of decision trees and the maximum depth on the Ms prediction for the alloys used in validation and testing process of the ML algorithm. For these simulations, the alloying elements are chosen as reference. The accuracy of the results increases with the number of decision trees until a saturation point is reached after 10 trees. Similarly, the maximum depth reduces the errors in Ms prediction and converges to an asymptotic value for 10 branches. Moreover, the accuracy of the model is clearly correlated with the number of elements, since for ternary and quaternary alloys the errors are less than 30 °C, whereas for HEAs used for validation the error saturates at 80 °C. For the HEAs developed specifically for the present work, which are utilized for testing, the accuracy is better and decreases below 50 °C when the number of trees and the maximum depth are below 50 °C. The number of trees and maximum depth are then both set to 10 for the remainder of the study as a compromise between computation time and accuracy.

References

[1] S. Saravanan, Iron : from metal to metallurgy, Heritage: J. Multidiscip. Stud. Archaeol. (2017) 514–518.
[2] A. Ölander, The Crystal Structure of AuCd, Zeitschrift für Kristallographie - Crystalline Materials. 83 (1932) 145–148, https://doi.org/10.1524/zkri.1932.83.1.145.
[3] A.B. Greninger, V.G. Mooradian, Strain transformation in metastable beta copper-zinc and beta copper-tin alloys, American Institute Min. Metall. Eng. (1938).
[4] L.C. Chang, T.A. Read, Plastic deformation and diffusionless phase changes in metals — the gold-cadmium beta phase, JOM 3 (1951) 47–52, https://doi.org/10.1007/BF03398954.
[5] W.J. Buehler, F.E. Wang, A summary of recent research on the nitinol alloys and their potential application in ocean engineering, Ocean Eng. 1 (1968) 105–120, https://doi.org/10.1016/0029-8018(68)90019-X.
[6] J. Mohd Jani, M. Leary, A. Subic, M.A. Gibson, A review of shape memory alloy research, applications and opportunities, Materials & Design (1980-2015). 56 (2014) 1078–1113. https://doi.org/10.1016/j.matdes.2013.11.084.
[7] J. Frenzel, E.P. George, A. Dlouhy, C.h. Somsen, M.-F.-X. Wagner, G. Eggeler, Influence of Ni on martensitic phase transformations in NiTi shape memory alloys, Acta Mater. 58 (2010) 3444–3458, https://doi.org/10.1016/j.actamat.2010.02.019.
[8] J. Frenzel, A. Wieczorek, I. Opahle, B. Maaß, R. Drautz, G. Eggeler, On the effect of alloy composition on martensite start temperatures and latent heats in Ni–Ti-based shape memory alloys, Acta Mater. 90 (2015) 213–231, https://doi.org/10.1016/j.actamat.2015.02.029.
[9] J. Ma, I. Karaman, R.D. Noebe, High temperature shape memory alloys, Int. Mater. Rev. 55 (2010) 257–315, https://doi.org/10.1179/095066010X12646898728363.
[10] A.N. Bucsek, G.A. Hudish, G.S. Bigelow, R.D. Noebe, A.P. Stebner, Composition, compatibility, and the functional performances of ternary NiTiX high-temperature shape memory alloys, Shap. Mem. Superelasticity. 2 (2016) 62–79, https://doi.org/10.1007/s40830-016-0052-5.
[11] L. Peltier, F. Meraghni, S. Berveiller, P. Lohmuller, P. Laheurte, Relationship between Chemical Composition and Ms Temperature in High-Entropy Shape Memory Alloys, Shap. Mem. Superelasticity. 7 (2021) 438–446, https://doi.org/10.1007/s40830-021-00342-1.
[12] F.K. Achard, Recherches sur les propriétés des alliages métalliques, 1788.https://books.google.fr/books?id=3649nQEACAAJ.
[13] B. Cantor, I.T.H. Chang, P. Knight, A.J.B. Vincent, Microstructural development in equiatomic multicomponent alloys, Mater. Sci. Eng. A 375–377 (2004) 213–218, https://doi.org/10.1016/j.msea.2003.10.257.
[14] E.P. George, D. Raabe, R.O. Ritchie, High-entropy alloys, Nat Rev Mater. 4 (2019) 515–534, https://doi.org/10.1038/s41578-019-0121-4.
[15] L. Peltier, P. Lohmuller, F. Meraghni, S. Berveiller, E. Patoor, P. Laheurte, Investigation and Composition Characterization of a "NiTi-like" Alloy Combining High Temperature Shape Memory and High Entropy, Shap. Mem. Superelasticity. 6 (2020) 273–283, https://doi.org/10.1007/s40830-020-00290-2.
[16] L. Peltier, S. Berveiller, F. Meraghni, P. Lohmuller, P. Laheurte, Martensite Transformation and Superelasticity at High Temperature of (TiHfZr)74(NbTa)26 High-Entropy Shape Memory Alloy, Shap. Mem. Superelasticity. 7 (2021) 194–205, https://doi.org/10.1007/s40830-021-00323-4.
[17] L. Peltier, P. Lohmuller, F. Meraghni, E. Patoor, P. Laheurte, S. Berveiller, Damping Behavior in a Wide Temperature Range of FeMn-Like High Entropy Shape Memory Alloys, Shap. Mem. Superelasticity. 8 (2022) 335–348, https://doi.org/10.1007/s40830-022-00381-2.
[18] Q. Han, Z. Lu, S. Zhao, Y. Su, H. Cui, Data-driven based phase constitution prediction in high entropy alloys, Comput. Mater. Sci 215 (2022), 111774, https://doi.org/10.1016/j.commatsci.2022.111774.
[19] Y. Zhang, C. Wen, C. Wang, S. Antonov, D. Xue, Y. Bai, Y. Su, Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models, Acta Mater. 185 (2020) 528–539, https://doi.org/10.1016/j.actamat.2019.11.067.
[20] W. Huang, P. Martin, H.L. Zhuang, Machine-learning phase prediction of high-entropy alloys, Acta Mater. 169 (2019) 225–236, https://doi.org/10.1016/j.actamat.2019.03.012.
[21] Y. Li, W. Yang, R. Dong, J. Hu, Mlatticeabc: generic lattice constant prediction of crystal materials using machine learning, ACS Omega 6 (2021) 11585–11594, https://doi.org/10.1021/acsomega.1c00781.
[22] K. Takahashi, L. Takahashi, J.D. Baran, Y. Tanaka, Descriptors for predicting the lattice constant of body centered cubic crystal, J Chem Phys. 146 (2017), 204104, https://doi.org/10.1063/1.4984047.
[23] Y. Ye, Y. Li, R. Ouyang, Z. Zhang, Y. Tang, S. Bai, Improving machine learning based phase and hardness prediction of high-entropy alloys by using Gaussian noise augmented data, Comput. Mater. Sci 223 (2023), 112140, https://doi.org/10.1016/j.commatsci.2023.112140.
[24] G. Vazquez, P. Singh, D. Sauceda, R. Couperthwaite, N. Britt, K. Youssef, D. D. Johnson, R. Arróyave, Efficient machine-learning model for fast assessment of elastic properties of high-entropy alloys, Acta Mater. 232 (2022), 117924, https://doi.org/10.1016/j.actamat.2022.117924.
[25] C. Wen, Y. Zhang, C. Wang, D. Xue, Y. Bai, S. Antonov, L. Dai, T. Lookman, Y. Su, Machine learning assisted design of high entropy alloys with desired property, Acta Mater. 170 (2019) 109–117, https://doi.org/10.1016/j.actamat.2019.03.010.
[26] J. Zhang, C. Cai, G. Kim, Y. Wang, W. Chen, Composition design of high-entropy alloys with deep sets learning, Npj Comput Mater. 8 (2022) 89, https://doi.org/10.1038/s41524-022-00779-7.

[27] A. Roy, G. Balasubramanian, Predictive descriptors in machine learning and data-enabled explorations of high-entropy alloys, Comput. Mater. Sci 193 (2021), 110381, https://doi.org/10.1016/j.commatsci.2021.110381.
[28] Y. Zeng, M. Man, K. Bai, Y.-W. Zhang, Revealing high-fidelity phase selection rules for high entropy alloys: A combined CALPHAD and machine learning study, Mater. Des. 202 (2021), 109532, https://doi.org/10.1016/j.matdes.2021.109532.
[29] F. Biermair, V.I. Razumovskiy, G. Ressel, Influence of alloying on thermodynamic properties of AlCoCrFeNiTi high entropy alloys from DFT calculations, Comput. Mater. Sci 202 (2022), 110952, https://doi.org/10.1016/j.commatsci.2021.110952.
[30] Z. Wu, J.W. Lawson, Theoretical investigation of phase transitions in the shape memory alloy NiTi, Phys. Rev. B. 106 (2022) L140102, https://doi.org/10.1103/PhysRevB.106.L140102.
[31] L. Zhang, K. Qian, J. Huang, M. Liu, Y. Shibuta, Molecular dynamics simulation and machine learning of mechanical response in non-equiatomic FeCrNiCoMn high-entropy alloy, J. Mater. Res. Technol. 13 (2021) 2043–2054, https://doi.org/10.1016/j.jmrt.2021.06.021.
[32] L. Qiao, Y. Liu, J. Zhu, A focused review on machine learning aided high-throughput methods in high entropy alloy, J. Alloy. Compd. 877 (2021), 160295, https://doi.org/10.1016/j.jallcom.2021.160295.
[33] G. Plummer, M.I. Mendelev, J.W. Lawson, Molecular dynamics simulations of austenite-martensite interface migration in NiTi alloy, Phys. Rev. Materials. 6 (2022), 123601, https://doi.org/10.1103/PhysRevMaterials.6.123601.
[34] X. Tian, D. Shi, K. Zhang, H. Li, L. Zhou, T. Ma, C. Wang, Q. Wen, C. Tan, Machine-learning model for prediction of martensitic transformation temperature in NiMnSn-based ferromagnetic shape memory alloys, Comput. Mater. Sci 215 (2022), 111811, https://doi.org/10.1016/j.commatsci.2022.111811.
[35] O. Eyercioglu, E. Kanca, M. Pala, E. Ozbay, Prediction of martensite and austenite start temperatures of the Fe-based shape memory alloys by artificial neural networks, J. Mater. Process. Technol. 200 (2008) 146–152, https://doi.org/10.1016/j.jmatprotec.2007.09.085.
[36] B. Xiong, X. Zhao, Y. Hu, H. Huang, Y. Liu, Y. Su, Machine learning assisted empirical formula augmentation, Mater. Des. 210 (2021), 110037, https://doi.org/10.1016/j.matdes.2021.110037.
[37] P.L. Narayana, S.-W. Kim, J.-K. Hong, N.S. Reddy, J.-T. Yeom, Estimation of Transformation Temperatures in Ti–Ni–Pd Shape Memory Alloys, Met. Mater. Int. 24 (2018) 919–925, https://doi.org/10.1007/s12540-018-0109-4.
[38] U.M.H.U. Kankanamge, J. Reiner, X. Ma, S.C. Gallo, W. Xu, Machine learning guided alloy design of high-temperature NiTiHf shape memory alloys, J Mater Sci. 57 (2022) 19447–19465, https://doi.org/10.1007/s10853-022-07793-6.
[39] S. Liu, B.B. Kappes, B. Amin-ahmadi, O. Benafan, X. Zhang, A.P. Stebner, Physics-informed machine learning for composition – process – property design: Shape memory alloy demonstration, Appl. Mater. Today 22 (2021), 100898, https://doi.org/10.1016/j.apmt.2020.100898.
[40] A.A. Catal, E. Bedir, R. Yilmaz, D. Canadinc, Design of a NiTiHf shape memory alloy with an austenite finish temperature beyond 400 °C utilizing artificial intelligence, J. Alloy. Compd. 904 (2022), 164135, https://doi.org/10.1016/j.jallcom.2022.164135.
[41] D. Xue, D. Xue, R. Yuan, Y. Zhou, P.V. Balachandran, X. Ding, J. Sun, T. Lookman, An informatics approach to transformation temperatures of NiTi-based shape memory alloys, Acta Mater. 125 (2017) 532–541, https://doi.org/10.1016/j.actamat.2016.12.009.
[42] S. He, Y. Wang, Z. Zhang, F. Xiao, S. Zuo, Y. Zhou, X. Cai, X. Jin, Interpretable machine learning workflow for evaluation of the transformation temperatures of TiZrHfNiCoCu high entropy shape memory alloys, Mater. Des. 225 (2023), 111513, https://doi.org/10.1016/j.matdes.2022.111513.
[43] R. Machaka, P.M. Radingoana, Prediction of narrow HT-SMA thermal hysteresis behaviour using explainable machine learning, Mater. Today Commun. 35 (2023), 105806, https://doi.org/10.1016/j.mtcomm.2023.105806.
[44] W. Trehern, R. Ortiz-Ayala, K. Atli, R. Arroyave, I. Karaman, Data-driven shape memory alloy discovery using artificial intelligence materials selection (AIMS) framework, Acta Mater. 228 (2022), https://doi.org/10.1016/j.actamat.2022.117751.
[45] S.J. Honrao, O. Benafan, J.W. Lawson, Data-Driven Study of Shape Memory Behavior of Multi-Component Ni–Ti Alloys in Large Compositional and Processing Space, Shap. Mem. Superelasticity. (2022), https://doi.org/10.1007/s40830-022-00405-x.
[46] Zadeh, et al., An Interpretable Boosting-based Predictive Model. https://doi.org/10.1016/j.jallcom.2020.157467.
[47] D. Piorunek, O. Oluwabi, J. Frenzel, A. Kostka, H.J. Maier, C. Somsen, G. Eggeler, Effect of off-stoichiometric compositions on microstructures and phase transformation behavior in Ni-Cu-Pd-Ti-Zr-Hf high entropy shape memory alloys, J. Alloy. Compd. 857 (2021), 157467, https://doi.org/10.1016/j.jallcom.2020.157467.
[48] D. Piorunek, J. Frenzel, N. Jöns, C. Somsen, G. Eggeler, Chemical complexity, microstructure and martensitic transformation in high entropy shape memory alloys, Intermetallics 122 (2020), 106792, https://doi.org/10.1016/j.intermet.2020.106792.
[49] D. Canadinc, W. Trehern, J. Ma, I. Karaman, F. Sun, Z. Chaudhry, Ultra-high temperature multi-component shape memory alloys, Scr. Mater. 158 (2019) 83–87, https://doi.org/10.1016/j.scriptamat.2018.08.019.
[50] M. Zarinejad, K. Wada, F. Pahlevani, R. Katal, S. Rimaz, Valence Electron Ratio for Design of Shape Memory Alloys with Desired Phase Transformation Temperatures, Shap. Mem. Superelasticity. 7 (2021) 179–189, https://doi.org/10.1007/s40830-021-00319-0.
[51] M. Zarinejad, Y. Liu, Dependence of Transformation Temperatures of NiTi-based Shape-Memory Alloys on the Number and Concentration of Valence Electrons, Adv. Funct. Mater. 18 (2008) 2789–2794, https://doi.org/10.1002/adfm.200701423.
[52] Kankanamge, et al., - 2022 - Machine learning guided alloy design of high-tempe. https://link.springer.com/10.1007/s10853-022-07793-6.
[53] L. Breiman, Random Forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[54] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Mach. Learn. 63 (2006) 3–42, https://doi.org/10.1007/s10994-006-6226-1.
[55] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.