Can Machine Learning Predict the Liquidus Temperature of Binary Alloys?

Yifei He[1,7], Isay Katsman[2], Sungwoo Sohn[1,7], Mark D. Shattuck[3], Corey S. O'Hern[1,4,5,7], Anna Gilbert[2,6], Jan Schroers[1,7]

[1]Department of Mechanical Engineering, Yale University, New Haven, CT 06520, USA

[2]Department of Statistics & Data Science, Yale University, New Haven, CT 06520, USA

[3]Benjamin Levich Institute and Physics Department, City College of CUNY, New York, NY 10031, USA

[4]Department of Applied Physics, Yale University, New Haven, CT 06520, USA

[5]Department of Physics, Yale University, New Haven, CT 06520, USA

[6]Department of Electrical Engineering, Yale University, New Haven, CT 06520, USA

[7]Department of Materials Science, Yale University, New Haven, CT 06520, USA

Abstract: Accurate prediction of the liquidus temperature (T_L) of alloys remains a challenge despite numerous theoretical models. Here, we explore and analyze the degree to which machine learning, ML, strategies can be used to predict T_L. We use established literature data on liquidus temperatures of 85523 binary alloys to train ML models using various feature vectors to represent the alloys. While our results are comparable to previous studies, the persistent ~8% error underscores the limitations of current ML models for practical usage. The suboptimal accuracy leads us to question how well-defined the problem is and to what degree fundamental limitations prevent us from attaining more accurate predictions. We identify two major challenges in predicting the liquidus temperature of binary alloys through supervised ML algorithms. One challenge is representing the relevant characteristics of an alloy that determines liquidus temperature through appropriate features. The other fundamental challenge is the discreteness of atoms properties. The difference between two elements and thereby alloy systems is significant, which makes it difficult to learn from one alloy system to predict properties of another. We argue that these problems can be reduced to some extent, however these challenges are common in complex materials science problems and constitute a fundamental challenge in applying supervised ML strategies in this context.

Keywords: Machine Learning, Liquidus Temperatures, Phase Diagrams, Alloys, Binary Alloy Systems, Random Forest, Neural Network.

1. Introduction

For pure elements, at least for the more practical ~100 elements, essentially all technologically relevant properties have been determined. However, in real-world applications, most materials we use are alloys, or combinations of these elements [1, 2, 3, 4]. Due to the vast number of alloys that can be formed [5], the fraction of alloy properties that have been measured for alloys is dramatically smaller than for pure elements. To address this knowledge gap, machine learning methods have emerged a potential tool to explore the vast compositional space of alloys [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18].

Problems in materials science that have been addressed using ML can be categorized into two classes. The first class consists of problems that can be reduced to properties that originate from a small number of atoms. Such problems can often be addressed through ab initio approaches: they include predicting elastic constants [19, 20, 21, 22], band structure [23, 24], diffusion [25, 26], and thermal expansion coefficient [27, 28]. Even though ab initio approaches still suffer limitations in representation, synthesizability, and accuracy [29, 30], combining these with ML models has been a fruitful area of research. Several examples of accurate predictions at low cost that accelerate speed and further led to the discovery of materials at unconventional chemical compositions have been given [31, 32, 33, 34, 35].

The second class consists of complex materials science problems in which properties and mechanisms originate from a large number of atoms. Examples for alloys include their microstructure, glass forming ability, plastic deformation, liquidus temperature, and phase diagrams. To address complex problems, models and guiding principles have been developed to estimate some of the alloys' properties [1, 2, 36, 37, 38, 39, 40, 41]. Generally, such model predictions are limited in their application range. If predictions are widely applicable, they are at best only semi quantitative and typically these models do not allow comparison across alloy systems. Hence, there is a strong need to develop prediction methods for alloy properties beyond today's state-of-the-art. As machine learning algorithms have been widely used to predict material properties, we consider ML here to predict the liquidus temperature, T_L, of alloys. We reduce the problem to binary alloys, as for most of the "practical alloy systems," T_L has been determined and hence can be compared with the prediction of the ML models. Hence, this problem is an ideal test bed to research if and how ML can be used to predict complex alloy properties.

Machine learning strategies have been used in the past to address complex materials science problems including the prediction of glass formation [11, 17, 42, 43, 44, 45], critical temperature of superconductivity [46], high entropy alloy formation [13, 18, 47, 48] and alloy liquidus temperature [14, 49, 50]. Generally, these first attempts have concluded that ML is effective in predicting complex materials and properties [43, 44, 46]. However, follow up work has often revealed limitations in the ability to truly predict, i.e. extrapolate [17, 49, 51, 52] as opposed to interpolate [8, 38, 53, 54, 55]. This is also evident by the fact that, despite wide usage of ML in materials science, few if any materials developed through ML strategies have been reported as front runners. When using ML to predict T_L, most predictions result in ~8-10% MAPE (mean absolute percentage error) [14, 49, 50], which is too large for practical applications. Some studies report lower errors, however more careful analysis reveals that significant information of the predicted alloy has been used in in the training set, hence the predictions have been interpolated [55] rather than extrapolated.

In this article, we explore the origins that limit the use of ML for complex materials science problems through the example problem of predicting the liquidus temperature of binary alloys. When using standard ML approaches like in previous work [14], i.e., random forest or neural network models, where feature representations of the alloys include their mixing behavior, we obtain a similar prediction error of MAPE $\sim$ 8%, which is similar to the previously reported MAPE of $\sim$ 10% [14, 49, 50]. We compare these MAPE values with the results from our simplest model that linearly interpolates the melting temperatures of alloys' constituents. The simplest model yields a MAPE of $\sim$ 13%, which reveals that the effectiveness of the ML models is low. We identify two core reasons for the high MAPE. One is that a priori known features to represent the behavior controlling T_L, only poorly represent T_L. The other is much more fundamental and originates from the discreteness of atoms. We confirm and quantify the latter by using "artificial atoms" which can continuously, through their features, vary between two real atoms or two real alloys.

2. Procedures

2.1. Data

We consider 55 elements based on their "practicality" (Fig. 1a), which result in $\binom{55}{2} = 1485$ possible binary alloy combinations. We then identify out of these 1485 combinations all binary systems for which a complete set of T_L data is available in the ASM Alloy Phase Diagram Database [56], which totals 429 binary alloy systems. For each of these complete 429 binary phase diagrams we extract T_L for $\sim$ 201 alloys in each alloy system, and these alloys are $\sim$ equally spaced in composition using WebPlotDigitizer [57] (Fig. 1b). Altogether, we curate such data, $(A_p\, B_{100-p},\ T_L^{A_p\, B_{100-p}})$, for 85523 binary alloys from 429 binary alloy systems.

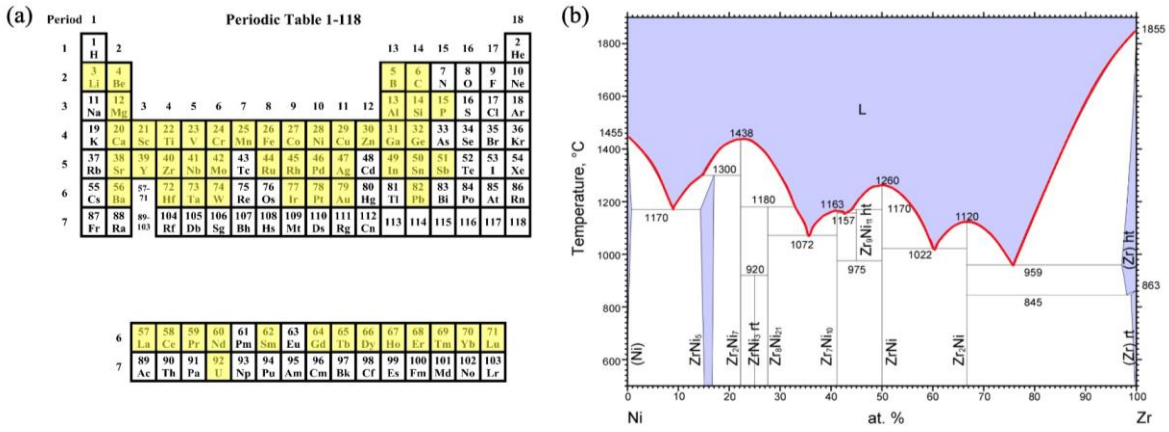

Figure 1: Data and data preparation process for considered binary alloys: **(a)** We selected 55 "practical" elements that comprise 429 binary alloy systems with a complete, for all compositions, T_L line. **(b)** T_L is obtained using WebPlotDigitizer [57] from the ASM binary phase diagrams [56]. For each binary alloy system, e.g. NiZr, we consider ~ 201 equally spaced compositions.

2.2. Features

We consider two fundamentally different feature vector sets. One is the Physically Informed feature vector that considers the mixing of the elements. The other is the Elemental Properties feature vector that only considers the properties of the constituent elements of the alloy. For benchmarking and comparison, we create Composition Fractions (Comp) features which only consider the composition of an alloy but don't include information about the alloys' properties. As an additional benchmark features, we also consider Random (Rand) features, i.e. random numbers that do not reflect any information about the alloy. Additional details of these different feature vectors are provided in the Supplementary Materials B.2. We describe below the seven Physically Informed features which constitute the Physically Informed feature vector describing the mixing of the constituent elements in binary alloys and its effect on the thermal stability.

(1) **Atomic size difference:**

$$\delta_{atomic\ size} = \sqrt{p_A \cdot \left(1 - \frac{r_A}{r_{avg}}\right)^2 + p_A \cdot \left(1 - \frac{r_B}{r_{avg}}\right)^2} \tag{1}$$

where p_A and p_B are the composition fractions of elements A and B; r_A and r_B are their atomic sizes and the weighted average of the atomic size is:

$$r_{avg} = p_A \cdot r_A + p_B \cdot r_B \tag{2}$$

The atomic size difference has been widely identified as critical in controlling various alloy characteristics including solubility limits in solid solutions [1, 58], metallic glass formation [2, 38], and high-entropy alloy formation [2, 3, 59, 60].

(2) **Atomic size range:**

$$\Delta r = \frac{|r_A - r_B|}{p_A \cdot r_A + p_B \cdot r_B} \tag{3}$$

(3) **Average melting temperature:**

$$T_{avg} = p_A \cdot T_A + p_B \cdot T_B \tag{3}$$

Where T_A and T_B as the melting temperatures of the elements A and B.

(4) **Difference in melting temperature:**

$$\Delta T = \frac{|T_A - T_B|}{p_A \cdot T_A + p_B \cdot T_B} \quad (5)$$

(5) **Heat of mixing [61]:**

$$H = \epsilon_{AB} \cdot p_A \cdot p_B \quad (6)$$

where ϵ_{AB} is calculated using Miedema's model for the pairs of elements in the alloys [62].

(6) **Normalized heat of mixing:**

$$H = \epsilon_{AA} \cdot p_A + \epsilon_{BB} \cdot p_B + H \quad (7)$$

where ϵ_{AA} and ϵ_{BB} are the cohesive energies of element A, B in eV/atom. Normalized heat of mixing is a key indicator of alloys mixing behavior which then effects phase diagram and hence T_L. As a high-level guiding rule, T_L deviates more from T_{avg} with increasing H_{AB}, as it deviates more from the ideal mixing behavior, where $T_L \sim T_{avg}$.

(7) **Crystal structure mismatch:**

$$CSM = CS \cdot p_A \cdot p_B \quad (8)$$

$CS = 0$ if the crystal structures of A and B are the same and $CS = 1$ if the crystal structures of A and B are different. CS has been recently identified as a key parameter that controls glass formation under conditions far from thermal equilibrium [63].

2.3. Models

We employ three fundamental machine learning (ML) models: linear regressions [64], random forest [65], and deep neural networks models [66]. Here, we give a minimal exposition of the methods to familiarize the reader and provide more detail in Supplementary B.3.

To develop the models and optimize hyperparameters, we employ 10-fold cross-validation. We use two fundamentally different approaches to separate the data for our problem. The first approach separates the training and testing data by the alloy systems (for more details, see Supplementary B.4). In this case, alloys of one alloy system can either be in the training set or in the testing set, but not in both. Alternatively, data are separated by alloys. In this case, and for the here used 90/10 split for the cross-validation, statistically ~180 alloys out of the 201 alloys per alloy system are already in the training set. An example would be that $A_{50}B_{50}$, $T_L(A_{50}B_{50})$ and $A_{48}B_{52}$, $T_L(A_{48}B_{52})$ are in the training data and the ML algorithm would predict $A_{49}B_{51}$, $T_L(A_{49}B_{51})$. Hence, predicting the missing ~ 20 alloys of one alloy system is a simple interpolation task. For this reason, we separate the data by alloy systems for our main analysis, thereby allowing us to evaluate the extrapolative ability of the ML models. In addition to a 90/10 split for the cross-validation, we also use a leave-one-system-out validation method (for more details, see Supplementary B.6.1), where all systems are used to train, except only one into which we then predict.

As a general benchmark model, we compare all predictions to a model that uses linear interpolation of the average melting temperature, $T_{avg} = p_A \cdot T_A + p_B \cdot T_B$ (Supplementary B.5). A summary of our methodology including data extraction, feature selection, and model development is provide in Fig. 2.

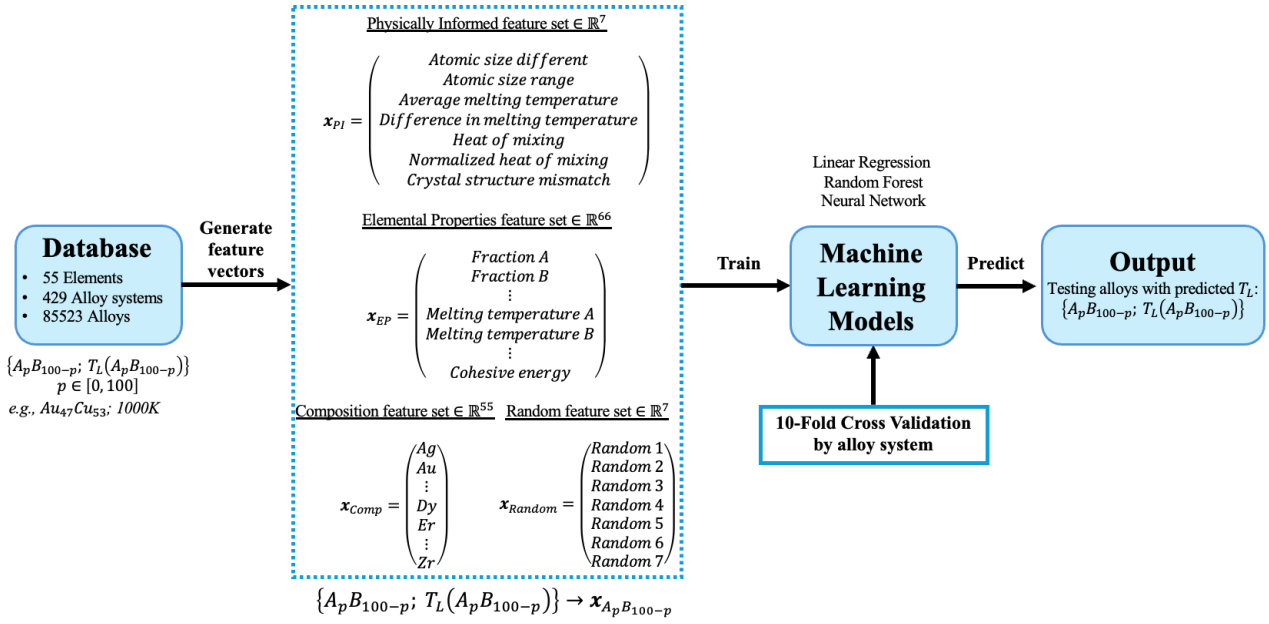

Figure 2: Outline of the here applied ML strategies to predict liquidus temperatures of binary alloys. Liquidus temperatures used in the training set are from the Alloy Phase Diagram Database (ASM) [56]. Overall, we consider 429 alloy systems AB and each system comprises of ~ 201 data, where each data comprises of the chemical composition A_pB_{100-p} of a binary alloy and its corresponding T_L. We represent the alloys through Physically Informed feature vector ($\in \mathbb{R}^7$), where we use mixing properties of alloys such as the average atomic size and heat of mixing. We also use Elemental Properties feature vector ($\in \mathbb{R}^{66}$), which only incorporates composition and elemental properties of the alloys' elements such as their melting temperature or atomic radius. For comparison, to benchmark the feature representation, we construct random feature vectors ($\in \mathbb{R}^7$), which are generated by assigning a random number ($\in \{0, 1\}$) to a feature. An additional benchmark feature representation is the Composition Fraction feature ($\in \mathbb{R}^{55}$) where an alloy is only represented by its chemical composition. Using such various feature representations, we train machine learning models through a 10-fold cross validation. These models are used to predict "left out data" and their predictability is quantified via the mean absolute percentage error (MAPE).

3. Results

The considered data originate from 429 binary alloy systems and each alloy system is represented by ~ 201 compositions resulting in a total of 85523 binary alloys altogether. ML models are developed for each of the feature vectors and their predictions are quantified through MAPE (Fig. 3a). The best predictions, averaged over all alloys, are for random forest ML model using PI features with a MAPE $= 7.96\% \pm 0.01\%$. Using a neural network model with the PI features results in a similar $MAPE = 8.73\% \pm 0.12\%$. The benchmark model $T_{avg} = p_A \cdot T_A + p_B \cdot T_B$ gives a MAPE of 12.84%. Using a linear regression model results in MAPE values similar or worse than achieved by the benchmark model. Going beyond the average values for the predictions we choose Co-Ti as an example binary alloy system (Fig. 3b). We evaluate the effect of four different feature vectors (Random, Comp, EP, PI) using Random Forest to predict T_L. These predictions are compared to measured T_L (considered as the true values) a linear regression model and the benchmark model. The best prediction, quantified in lowest MAPE of ~ 4% (average MAPE for all composition within this alloy system) is obtained by the random forest model trained on PI features. The EP features, without considering mixing of the elements, achieve a slightly higher MAPE of approximately ~ 5%. In comparison, features derived only from composition yield a significantly higher MAPE ~ 15% and Random features produce a MAPE of ~ 12%. For EP and PI feature vectors the predictions mimic to various extends the composition dependent of T_L within this system. This is not the case when using random forest model with Random and Composition features, linear regression model (MAPE ≈ 12%), and the benchmark model (MAPE ≈ 20%), which are unable to capture the specifics of this system.

(a)

	PI	EP	Comp	Rand
Benchmark	12.84 ± 0.00%	12.84 ± 0.00%	12.84 ± 0.00%	12.84 ± 0.00%
LR	11.37 ± 0.10%	14.93 ± 0.20%	11.44 ± 0.05%	32.73 ± 0.41%
NN [15]	8.73 ± 0.12%	-	-	-
RF	7.96 ± 0.01%	8.85 ± 0.02%	13.46 ± 0.01%	34.16 ± 0.56%

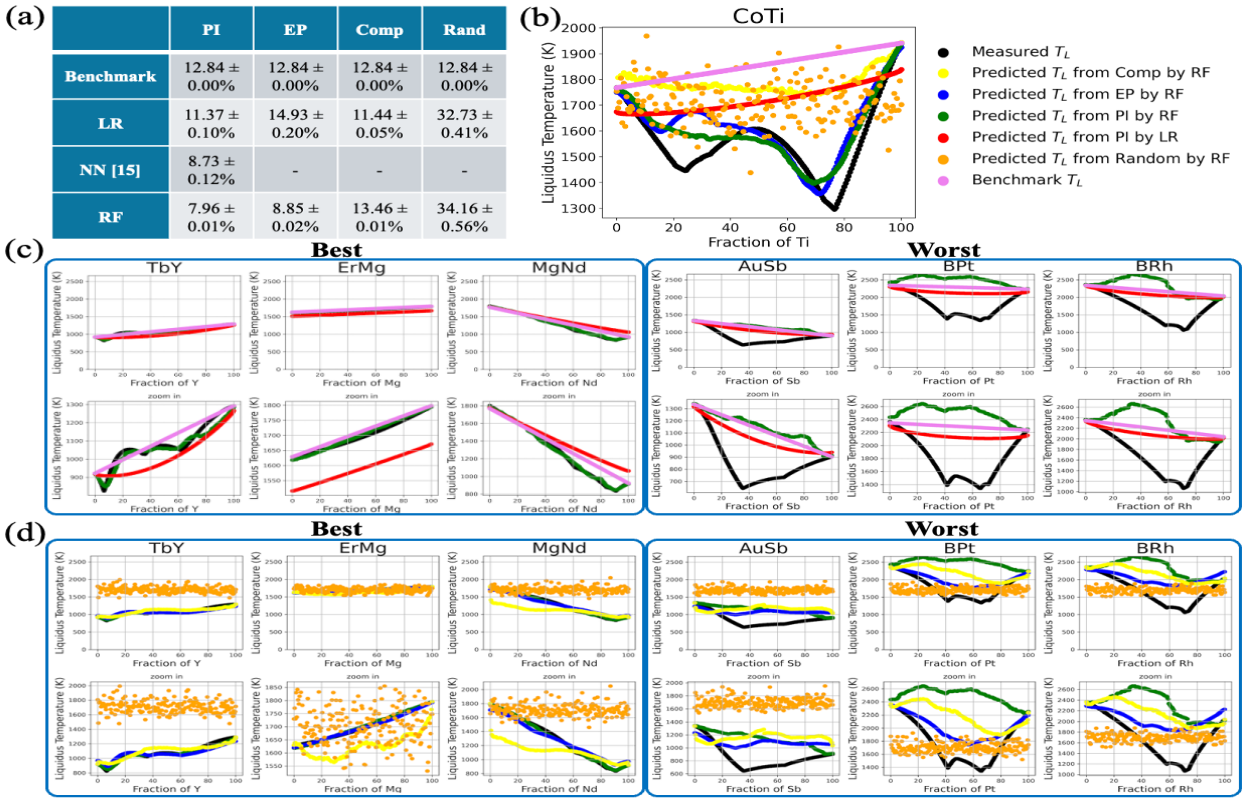

Figure 3: Summary of the ML predictions of T_L for binary alloys: **(a)** Summary of average predictions from linear regression, neural network and random forest with 10-fold CV where training/testing data splitting is based on alloy systems. The benchmark model (linear interpolation), $T_{avg} = p_A \cdot T_A + p_B \cdot T_B$, is also shown. Standard deviations are given over five trials for all models. **(b)** Random forest predictions of T_L using four different features vectors (Comp, EP, PI, Random) for Co-Ti as an example binary alloy. These predictions are compared to T_L obtained from the linear regression and benchmark models. The best prediction, quantified by the lowest MAPE of ~ 4% is obtained using the random forest model trained on PI features. EP features result in a MAPE of ~ 5%. Composition features result in a MAPE of ~ 15% and random features result in a MAPE of ~ 12%. **(c)** Three examples are shown of the most accurate and least accurate predictions for an alloy system with Physically Informed features, the leave-one-system-out validation method and different models. The best three alloy systems prediction are for TbY (MAPE = 0.41%) and ErMg (MAPE = 1.37%) and MgNd (MAPE = 2.25%). The three worst predictions for alloy systems are for BRh (MAPE = 44.61%), BPt (MAPE = 44.27%) and AuSb (MAPE = 38.00%). These worst predictions are even worse than the predictions using the benchmark model. **(d)** Three examples of the most accurate and least accurate Random Forest predictions for an alloy system with different features and the leave-one-system-out validation method. The PI features give the best predictions.

We also consider the extremes of the predicted MAPE distribution by considering three examples of the most accurate and least accurate predictions for an alloy system, based on the various models using Physically Informed features (shown in Fig. 3c with different models). Using the leave-one-system-out validation method and Random Forest model with the Physically Informed features, the three alloy systems with the best prediction are for TbY (MAPE = 0.41%) and ErMg (MAPE = 1.37%) and MgNd (MAPE = 2.25%). Also, three of the alloy systems with the worst predictions are for AuSb (MAPE = 38.00%), BPt (MAPE = 44.27%) and BRh (MAPE = 44.61%). The alloy systems with the worst predictions possess even larger MAPE values than the MAPE determined using the benchmark model.

When considering different feature vectors for the best and worst predictions per alloy system with Random Forest model, PI features result in smallest MAPE and appear to have the most ‘‘insight”, which is reflected in the shape

of the liquidus temperature curve (Fig. 3d). When using Elemental Properties as features, MAPE is higher, which is not surprising as mixing is not considered in these features. When using Random features, predication drops significantly, to a MAPE > 30% for both the linear regression and random forest ML model. Composition features achieve better performance than Random features, as they preserve the continuous elemental fraction information of the binary alloys. Random features do not have any physical meaning, and as a consequence, lead to noisy and unstructured predictions.

4. Discussion

Our results reveal that even the best predictions we can achieve are consistently above ~ 8% MAPE. These results are obtained using the best possible methodologies (i.e. random forest model with the most informative Physically Informed feature vectors using leave-one-system-out validation). However, this predictability is only a factor of two better than that obtained from the $T_{avg} = p_A \cdot T_A + p_B \cdot T_B$. Such, at best, moderate predictability is highly surprising. Further, our data reveal that there is not a considerable difference between using a simple nonlinear model like a Random Forest and a more complicated model like a well-tuned neural network model. Although this may seem surprising, we believe it is a consequence of the fact that the problem of predicting T_L of binary alloys itself has several fundamental limitations that prevent one from obtaining more accurate predictions through machine learning.

We argue that the reason for the poor predictability achieved by machine learning models originate mainly from two contributions. One is the suboptimal representation of the alloys through the features to represent T_L. Theoretically, as stated by the universal approximation theorem [67], even a poor presentation of the output label (here T_L) through its features can result in high predictability as long as the prediction is over a compact domain and sampling can be carried out densely. We argue here, and this is a fundamental limitation of machine learning for complex materials science problems that involve discrete combinations of atoms, that the quantity of data cannot be arbitrarily increased but instead clustered and reduced to only some locations in the feature space. This is because complex problems involve combinations of atoms, which can only lead to some combinations of features, but others can never be realized. This is the second contribution, and it originates from the fact that atoms are discrete, and hence their feature representations.

In the following we discuss the feature representation and discreteness of atoms and the resulting accessible feature space.

4.1. Feature Representation

4.1.1. Feature and Liquidus Line Distance Calculation

We argue that the poor feature representation can be quantified in two ways. One is reflected in the overall poor predictability of T_L through our models reflected in the high MAPE values (Fig. 3a). The other is the argument that a measure of goodness for feature representations in our context should reflect in the following: *If the feature representations for two alloy systems are similar, then one can conclude that their T_L's must be similar.*

Looking at alloys divorced from their system results in information loss and can lead to redundant analysis in our case, since intermediate alloy features are synthesized from the alloy system's constituent atoms. As a consequence, and to reduce problem, we will generalize our analysis to the level of alloy systems instead of individual alloys. Hence, the measure of goodness changes to: *If the feature representations for two alloy systems are similar, then one can conclude that their liquidus lines must be similar.*

Specifically, we will represent an alloy system *AB* by a concatenation of the features of its constituent atoms. In addition, the PI feature of an alloy system, $\boldsymbol{X}_{PI}$, is in $\mathbb{R}^{201\times7}$, which consists of 201 alloy compositions, each represented by seven features (Details in Section 2.2). However, we define the feature vector for a binary alloy system AB as $\boldsymbol{X}_{AB} = \begin{bmatrix}\boldsymbol{X}_A \\ \boldsymbol{X}_B\end{bmatrix}$, where $\boldsymbol{X}_A \in \mathbb{R}^7$ are the Elemental Properties (EP) vectors of pure elements A and B, and $\boldsymbol{X}_{AB} \in \mathbb{R}^{14}$. Since our goal is to build a model that generalizes to unseen alloy systems, we intentionally avoid using

Physically Informed (PI) features. The PI feature contains the mixing information, which may cause biases. Instead, we extract the EP feature vector $\boldsymbol{X}_{AB}$ only from the endpoints of the composition range (the pure elements $A_{100}B_0$ and A_0B_{100}. This approach ensures that the feature vectors are derived solely from elemental properties, independent of any knowledge of the system's liquidus behavior. Also, EP feature vectors provide a cleaner and more generalizable basis for comparing alloy systems. It allows us to compute distances between systems in feature space based only on elemental similarities, without being affected by data artifacts or fitting noise in the PI domain. Similarly, we define the liquidus temperature vector for a binary system AB as $\boldsymbol{T}_L^{AB} \in \mathbb{R}^{201}$, representing its temperature values across the composition range.

We propose that a good alloy system representation should preserve the continuity of the mapping from features to liquidus line. Specifically, if two alloy systems have similar EP feature vectors (i.e., $\boldsymbol{X}_{AB} \approx \boldsymbol{X}_{CD}$), then their corresponding liquidus temperature vectors should also be similar (i.e., $\boldsymbol{T}_L^{AB} \approx \boldsymbol{T}_L^{CD}$). In other words, the map from feature vectors of an alloy system to its liquidus line has some continuity property.

Alloy System Feature Correlation with Liquidus Line: To quantify the degree to which the alloy system feature space represents the liquidus line space, we first need to define metrics on the spaces of alloy system features and liquidus line. For the sake of simplicity to define a distance matrix between feature vectors of each two alloy systems, we will use the simple Euclidean ℓ_2 distance on the space of features:

$$\ell_2(\boldsymbol{X}_{AB}, \boldsymbol{X}_{CD}) = \sqrt{(\boldsymbol{X}_{AB} - \boldsymbol{X}_{CD})^2}$$

$\boldsymbol{X}_{AB} \in \mathbb{R}^{14}$ and $\boldsymbol{X}_{CD} \in \mathbb{R}^{14}$ represent the feature vectors of the alloy systems AB and CD.

To represent the liquidus line of an alloy system $\boldsymbol{T}_L^{AB} \in \mathbb{R}^{201}$, we use a 201-dimensional real-valued vector where each dimension represents a liquidus temperature for a particular alloy (or composition p) $T_L^{A_pB_{100-p}}$. To properly describe the difference between the liquidus lines of two different alloy systems, we argue that the absolute difference and the difference in local slopes must be considered. The absolute difference is quantified through:

$$\ell_1(\boldsymbol{T}_L^{AB}, \boldsymbol{T}_L^{CD}) = \sum_{p=0}^{100} \left| T_L^{A_pB_{100-p}} - T_L^{C_pD_{100-p}} \right| \quad \text{(where p = 0, 0.5, 1, …, 100)}$$

$\boldsymbol{T}_L^{AB} \in \mathbb{R}^{201}$ and $\boldsymbol{T}_L^{CD} \in \mathbb{R}^{201}$ are the liquidus line vectors of alloy systems, AB and CD. $T_L^{A_pB_{100-p}}$ and $T_L^{C_pD_{100-p}}$ represent the liquidus temperatures of alloys A_pB_{100-p} and C_pD_{100-p}.

The difference in local slopes of the liquidus lines of two different alloy systems is quantified through:

$$\ell_1(\boldsymbol{S}_L^{AB}, \boldsymbol{S}_L^{CD}) = \sum_{p=0}^{100} \left| T_L^{A_pB_{100-p}} - T_L^{A_{p-1}B_{101-p}} - T_L^{C_pD_{100-p}} + T_L^{C_{p-1}D_{101-p}} \right| \quad \text{(where p = 0, 0.5, 1, …, 100)}$$

$\boldsymbol{S}^{AB}, \boldsymbol{S}^{CD} \in \mathbb{R}^{200}$ are the local slope vectors of alloy systems AB and CD.

We propose a weighted linear combination of the ℓ_1 difference between two liquidus lines and the local slope information between them. We argue that the local slopes are critical as they carry important information about the specifics of the alloy system. We define the liquidus line distance as:

$$T_{L_{\text{dist}}}^{(\alpha,\beta)}(\boldsymbol{T}_L^{AB}, \boldsymbol{T}_L^{CD}) = \alpha \cdot \ell_1(\boldsymbol{T}_L^{AB}, \boldsymbol{T}_L^{CD}) + \beta \cdot \ell_1(\boldsymbol{S}^{AB}, \boldsymbol{S}^{CD})$$

where $\alpha, \beta \in \mathbb{R}$ are learned fitting parameters and represent the relative importance of the two contributions, 0^{th} and 1^{st} order information in constructing the liquidus line metric.

To obtain the learned α and β parameters we minimize the following mean squared error loss over all pairs of alloy systems:

$$\ell_2(\alpha,\beta) = \frac{1}{\binom{362}{2}} \sum_{(AB,CD)\in\binom{362}{2}\ systems} \left(\ell_2(\boldsymbol{X}_{AB},\boldsymbol{X}_{CD}) - T_{L_{dist}}^{(\alpha,\beta)}(\boldsymbol{T}_L^{AB},\boldsymbol{T}_L^{CD})\right)^2$$

$$(\alpha^*,\beta^*) = \arg\min_{\alpha,\beta} \ell\,(\alpha,\beta)$$

We use *EP* features and normalize all data by mean and standard deviation (as discussed in Supplementary B.1). We optimized this loss with PyTorch [68] to obtain a converged average mean squared error of 1.34 and the best parameters of α and β: $\alpha^* = 0.0039, \beta^* = 0.7720$. We argue these learned values for α and β are reasonable and interpretable: the first order information (slope: β) between liquidus lines is highly informative as it reflects the specifics of the liquidus line. It is reasonable that the slope carries more information than the absolute value. Moreover, the mean squared error of the α and β fitting of 1.34 is reasonably good, given that we normalized all features.

4.1.2. Visualizing Explainability of Features:

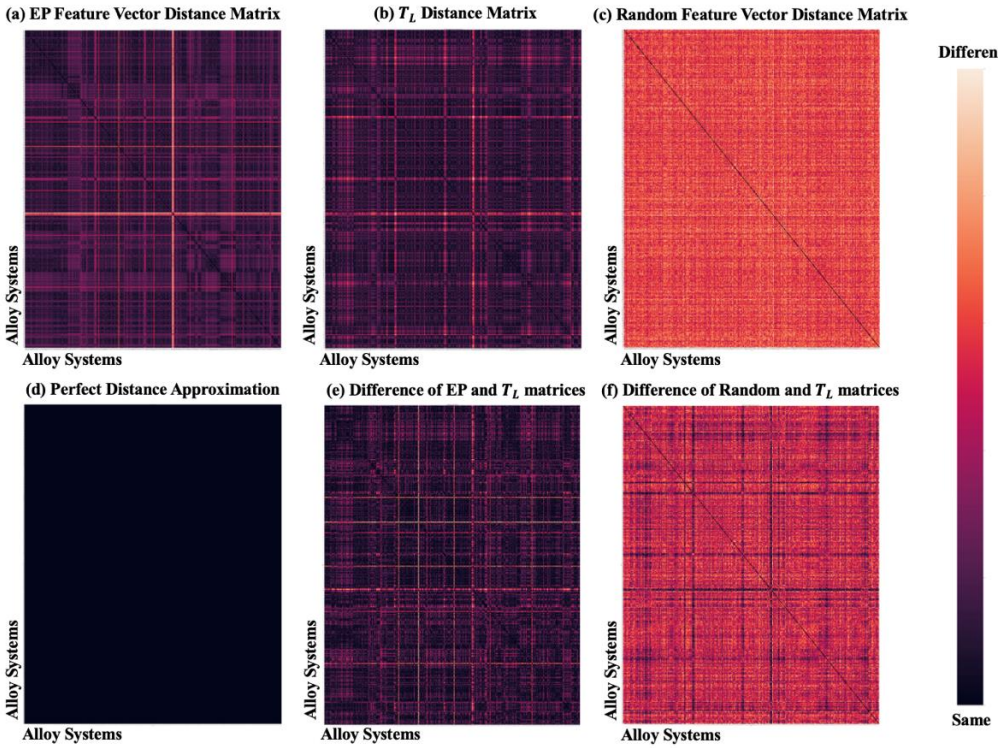

Figure 4: (a): 362×362 feature vector distance matrix when using Elemental Properties features. Each entry corresponds to the distance between the alloy systems represented by the i^{th} row and j^{th} column. Details about the quantification of the distance matrix are given in Section 4.1.1. **(b)**: Liquidus line distance matrix between all 362 binary alloy systems. **(c)**: The feature distance matrix when random features are used. **(d)**: Schematic of difference of EP features and Liquidus lines distances matrix that shows perfect correlation; $\boldsymbol{X}_{AB} \approx \boldsymbol{X}_{CD} \Rightarrow \boldsymbol{T}_L^{AB} \approx \boldsymbol{T}_L^{CD}$, i.e. a matrix of all zeros. **(e)**: Distance matrix for the correlation of the difference between the EP feature distance matrix and the liquidus line distance matrix. **(f)**: Distance matrix for the correlation of the random feature distance matrix and the liquidus line distance matrix.

Now that we have set up relevant alloy system feature and liquidus line vectors and identified appropriate quantities to represent distances, we are able to quantify how "representative" our features are, i.e., how well they codify the intuition "$\boldsymbol{X}_{AB} \approx \boldsymbol{X}_{CD} \Rightarrow \boldsymbol{T}_L^{AB} \approx \boldsymbol{T}_L^{CD}$" (how well our feature distances approximate the liquidus line distances). To visualize this, we create distance matrices for the distance of all features and liquidus lines for all the 362 alloy systems for which a complete set of liquidus temperatures (all 201) are available (Fig. 4). Fig. 4a shows the EP feature vector distance matrix, where the (i,j)th element of the matrix represents the ℓ_2 distance, which is represented as a heat map, between the EP features of the i^{th} alloy system and the j^{th} alloy system. Similar for the liquidus line distance matrix, where we use the learned $T_{L_{dist}}^{(\alpha^*,\beta^*)}$ to compute the distance between liquidus lines of alloy systems (Fig. 4b). For comparison, a random feature distance matrix is shown in Fig. 4c. To quantify "$\boldsymbol{X}_{AB} \approx \boldsymbol{X}_{CD} \Rightarrow \boldsymbol{T}_L^{AB} \approx \boldsymbol{T}_L^{CD}$" (Fig. 4e), we subtract the EP feature distance matrix (Fig. 4a) from the liquidus line distance matrix (Fig. 4b). A perfect correlation of "$\boldsymbol{X}_{AB} \approx \boldsymbol{X}_{CD} \Rightarrow \boldsymbol{T}_L^{AB} \approx \boldsymbol{T}_L^{CD}$" would result in vanishing values of the subtracted matrix (to appear in uniform black in Fig. 4d). The other extreme of no correlation (Fig. 4f) is present between the liquidus line distance matrix (Fig. 4b) and random feature vector distance matrix (Fig. 4c). The matrix representing the subtraction of feature distance matrix based on Elemental Properties (Fig. 4a) from the liquidus line distance matrix (Fig. 4b), shown in Fig. 4e falls in between those two extremes (no correlation vs. perfect correlation). Therefore, we can conclude that the EP features represent T_L to some extent, however not perfectly.

4.2. Atom Discreteness

We argue that the other, and very fundamental reason that ML models are ineffective to solve complex materials science problems (reflected in high MAPE), originates from the discreteness of atoms. Such discreteness reflects and can be quantified in the discreteness of their features and prevents that the feature space can be arbitrarily sampled, as a fraction of the feature space will not be accessible even in the limit of infinite data. This is very different from the classical machine learning setting, where one can in theory sample arbitrarily many points which are arbitrarily distributed within the feature space from the underlying distribution.

If there existed, theoretically, continuous atoms that interpolate between actual atoms, and hence a continuous compact space that contains all of our data points, one could always learn a function over this space simply by sampling it densely [67]. This is true and a high accuracy can be achieved, as long as a sufficiently large number of samples will be used even with a feature representation that represents the data poorly. However, in our physical world that is not the case. We will now provide arguments that such discreteness indeed prevents sampling of the entire feature space in the limit of infinite training data.

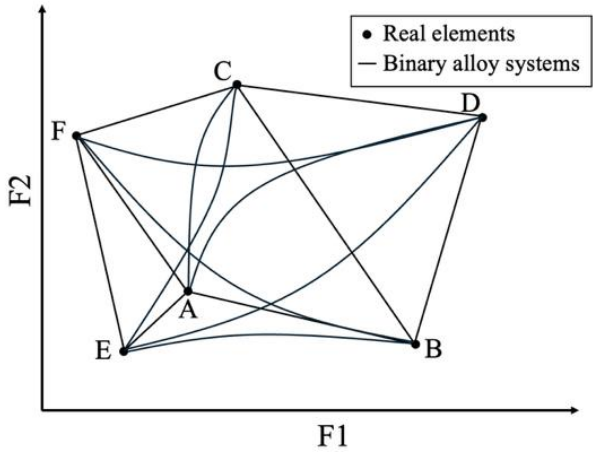

Figure 5: Schematics of a representation of binary alloy systems represented by two features, F1 and F2. Only combinations of F1 and F2 that are on the lines connecting in a binary fashion the elements A, B, C, D, E, F exist. Additional data are always on these lines, even in the limit of infinite data and the majority of the feature space remains unrepresented.

Fig. 5 shows that the space of features of binary alloy systems is discrete. That is why, even as the number of data points approaches infinity, the MAPE does not approach zero. In the case of T_L prediction, the features lie along a low-dimensional manifold effectively a "line" and increasing the data size only results in denser sampling along that line, without providing information beyond it.

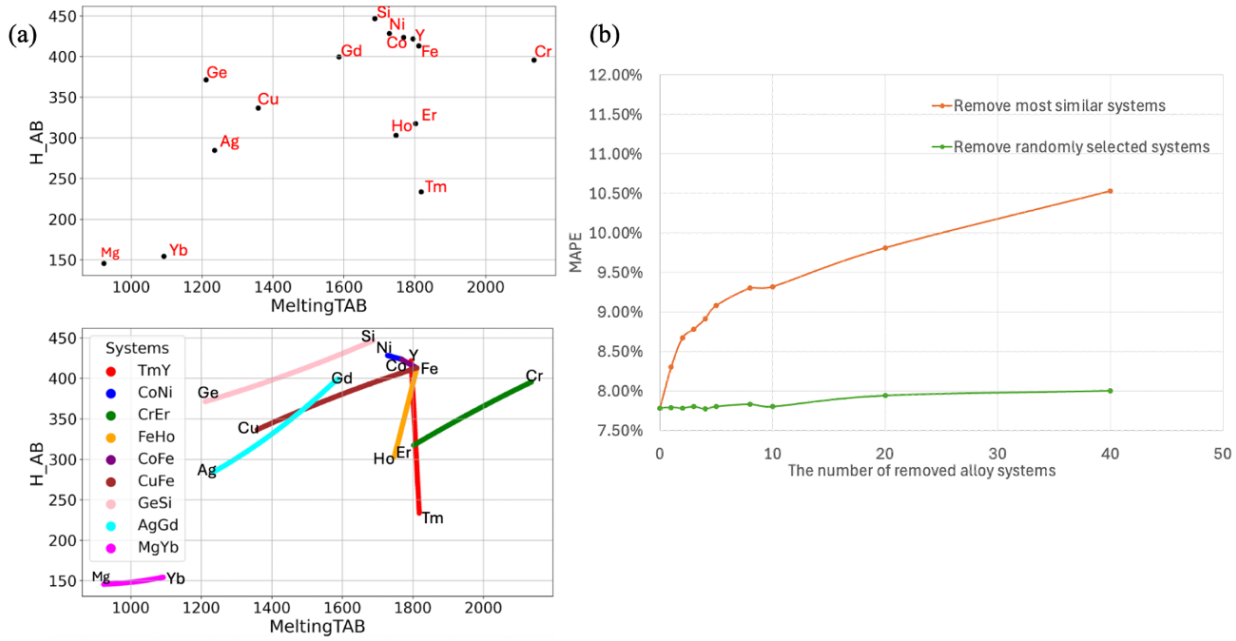

Figure 6: (a) Example of discreteness of features space considering average melting temperature and heat of mixing for some selected elements. Only a small fraction of the features space is available. Some elements and their binary combinations have similar features. (e.g. CoFe and CoNi) whereas others are very different from all other considered (e.g. MgYb). **(b)** The impact on the prediction of a ML model of such discreteness of the available features and their distribution in feature space can be observed when comparing the ML models trained on alloy systems from which some are removed. Either the most similar systems are removed (most-similar) (Section 4.1.1 shows how to find the most similar systems), or random systems are removed (Random). The values plotted are average out over all test systems.

Experimental evidence suggesting such discreteness can be obtained when selectively eliminating systems from the training set (Fig. 6). We argue, and Fig. 6a provides examples, that such discrete feature values reflect in the quality, here similarity to the alloy system left out for predicting. When we randomly select and eliminate an alloy system from the training set and train a RF-ML model, the prediction is essentially (statistically) unaffected (Fig. 6b). However, if we selectively choose the most similar alloy system to the one left out for prediction, the prediction rapidly worsens. This result not only shows the importance of a meaningful feature representation, systems with similar features are indeed more informative for prediction, but also implicitly reflects the role of atomic discreteness. As only a discrete set of alloy systems is available, removing similar systems leads to abrupt performance drops, underscoring the limitations imposed by the inherent discreteness of the dataset.

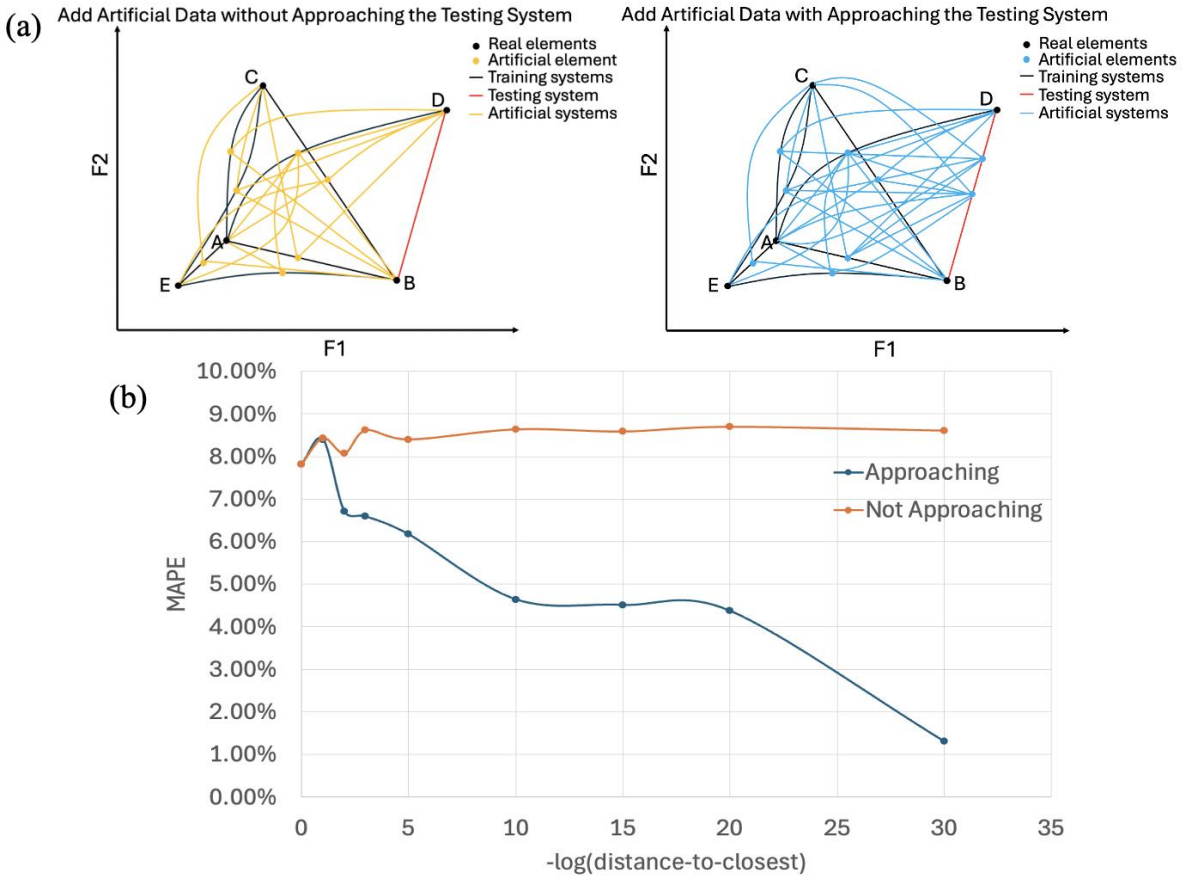

Figure 7: Effect of adding artificial atoms to the training data. **(a)** Left: Add artificial data but do not approach the testing system, BD. Right: Add artificial data and approach the testing system, BD. **(b)** Effect of the number of artificial atoms added to the training set on MAPE. Just adding artificial atoms but not approaching the left-out system does not affect predictability. Only if we approach the left-out system, the accuracy of the prediction increases.

To provide further evidence that the discreteness of atoms leads to discreteness in phase diagrams, and their corresponding feature representation fundamentally limits predictability of ML models to predict the liquidus temperature of alloys, we use artificial atoms and alloy systems to simulate a continuous alloy, alloy system and corresponding feature representation scenario. Specifically, we define artificial atoms and describe them as linear interpolations of the feature vectors of neighboring real atoms (see supplementary D.2).

We make the following important differentiation. At first (Fig. 7a, left), artificial atoms in between atoms are created in between atoms (and their features) in the training set but do not approach the left-out system (BD in Fig. 7a). Essentially, we add feature density in the training set but don't specifically approach the left-out system. Such adding of data has essentially no effect on the MAPE (Fig. 7b). Here we argue that generally, when we add features to the training set away from the left-out system, such feature carries insignificant information as they only interpolate between already know information and are too different from the left-out system (see also Fig. 6b). An improvement in predictability is only possible when adding features directly approaching the left-out system. In that case, information of the left-out system is in the training set. Such approaching of the left-out system would naturally happen in a world of continuous atoms when increasing the number of test data. In such a world, in the limit of infinite data the predictability would be perfect. However, in the existing world of discrete atoms to truly predict, hence not approaching, and incorporating information from the alloy system into which one predicts, just adding artificial atoms interpolating between atoms in the training set has essentially no effect on the predictability

of a machine learning model, hence highlights the severeness of atom discreteness limiting machine learning approaches.

5. CONCLUSIONS

When using machine learning strategies to predict the liquidus line of a binary alloy system, considering 85523 alloys for training, which are represented through various feature vectors, we find that our predictions are limited to an error of ~ 8%, comparable to previously reported results. Such suboptimal predictability, unsatisfactory for these models to be of practical use, suggests two limitations of supervised machine learning strategies to predict the liquidus line of an alloy system. One originates from the challenge of representing the relevant characteristics of an alloy system that determines liquidus line through features. The other deeply fundamental challenge is the discreteness of atoms. The difference between two elements and thereby alloy systems is significant and hence makes it generally challenging to learn from one alloy system to predict another. We argue that these challenges are common in complex materials science problems and constitute a fundamental challenge in applying supervised ML strategies in this context. To reduce such limitations, one must develop more accurate descriptions of the complex problem, often related to its mixing behavior to achieve better feature representations. This is what the metallurgical community has been doing over the last century. To mitigate the discreteness problems, reducing the problem through subgrouping into more similar systems is one strategy. The trade off is that predictions become less general as the application space of the models are reduced. Another strategy may be to use model systems and alloys, for example in MD simulations, where the discreteness of atoms is lifted and arbitrarily continuous interactions which approximate atoms and alloys can be synthesized. Such modeling relies on faithful atom representations through potentials, where the simulation community has been making significant progress in the last decades.

References

[1] M. F. Ashby. Materials in mechanical design. *MRS Bulletin*, 18:43–54, 1993.

[2] Akihisa Inoue. Stabilization of metallic supercooled liquid and bulk amorphous alloys. *Acta materialia*, 48(1):279–306, 2000.

[3] M. F. Ashby and Alan Lindsay Greer. Metallic glasses as structural materials. *Scripta Materialia*, 54:321–326, 2006.

[4] Daniel B. Miracle and Oleg N. Senkov. A critical review of high entropy alloys and related concepts. *Acta Materialia*, 122:448–511, 2016.

[5] Yanglin Li, Shaofan Zhao, Yanhui Liu, Pan Gong, and Jan Schroers. How many bulk metallic glasses are there? *ACS combinatorial science*, 19 11:687–693, 2017.

[6] Ruoqian Liu, Abhishek Kumar, Zhengzhang Chen, Ankit Agrawal, Veera Sundararaghavan, and Alok Choudhary. A predictive machine learning approach for microstructure optimization and materials design. *Scientific reports*, 5(1):11551, 2015.

[7] Felix A Faber, Alexander Hans Gustav Lindmaa, O. Anatole von Lilienfeld, and Rickard Armiento. Machine learning energies of 2 million elpasolite (abc_{2}d_{6}) crystals. *Physical review letters*, 117 13:135502, 2015.

[8] Tian Xie and Jeffrey C. Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters*, 120 14:145301, 2017.

[9] Weike Ye, Chi Chen, Zhenbin Wang, Iek-Heng Chu, and Shyue Ping Ong. Deep neural networks for accurate predictions of crystal stability. *Nature Communications*, 9, 2017.

[10] Chi Chen, Weike Ye, Yunxing Zuo, Chen Zheng, and Shyue Ping Ong. Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials*, 2018.

[11] Fang Ren, Logan Ward, Travis Williams, Kevin J Laws, Christopher Wolverton, Jason Hattrick-Simpers, and Apurva Mehta. Accelerated discovery of metallic glasses through iteration of machine learning and highthroughput experiments. *Science advances*, 4(4):eaaq1566, 2018.

[12] K. Kaufmann, D. Maryanovsky, W.M. Mellor, C. Zhu, A.S. Rosengarten, T.J. Harrington, C. Oses, C. Toher, S. Curtarolo, K.S. Vecchio, Discovery of high-entropy ceramics via machine learning, Npj Comput Mater 6(1) (2020) 42.

[13] K. Kaufmann, K.S. Vecchio, Searching for high entropy alloys: A machine learning approach, Acta Mater 198 (2020) 178-222.

[14] Pin-Wen Guan and Venkatasubramanian Viswanathan. Meltnet: Predicting alloy melting temperature by machine learning. *arXiv preprint arXiv:2010.14048*, 2020.

[15] Zongrui Pei, Kyle A Rozman, Ömer N Doğan, Youhai Wen, Nan Gao, Elizabeth A Holm, Jeffrey A Hawk, David E Alman, and Michael C Gao. Machine-learning microstructure for inverse material design. *Advanced Science*, 8(23):2101207, 2021.

[16] Ziyuan Rao, Po-Yen Tung, Ruiwen Xie, Ye Wei, Hongbin Zhang, Alberto Ferrari, TPC Klaver, Fritz Körmann, Prithiv Thoudden Sukumar, Alisson Kwiatkowski da Silva, et al. Machine learning–enabled high-entropy alloy discovery. *Science*, 378(6615):78–85, 2022.

[17] Guannan Liu, Sungwoo Sohn, Sebastian A Kube, Arindam Raj, Andrew Mertz, Aya Nawano, Anna Gilbert, Mark D Shattuck, Corey S O'Hern, and Jan Schroers. Machine learning versus human learning in predicting glass-forming ability of metallic glasses. *Acta Materialia*, 243:118497, 2023.

[18] O.A. Graeve, M.S. García-Vázquez, A.A. Ramírez-Acosta, Z. Cadieux, Latest Advances in Manufacturing and Machine Learning of Bulk Metallic Glasses, Adv Eng Mater 25(9) (2023) 2201493.

[19] Artem R Oganov, John P Brodholt, and G David Price. The elastic constants of mgsio3 perovskite at pressures and temperatures of the earth's mantle. *Nature*, 411(6840):934–937, 2001.

[20] GV Sin'Ko and NA Smirnov. Ab initio calculations of elastic constants and thermodynamic properties of bcc, fcc, and hcp al crystals under pressure. *Journal of Physics: Condensed Matter*, 14(29):6989, 2002.

[21] Rostam Golesorkhtabar, Pasquale Pavone, Jürgen Spitaler, Peter Puschnig, and Claudia Draxl. Elastic: A tool for calculating second-order elastic constants from first principles. *Computer Physics Communications*, 184(8):1861–1873, 2013.

[22] Henrik Levämäki, Ferenc Tasnadi, DG Sangiovanni, LJS Johnson, Rickard Armiento, and IA Abrikosov. Predicting elastic properties of hard-coating alloys using ab-initio and machine learning methods. *npj Computational Materials*, 8(1):17, 2022.

[23] Joachim Paier, Martijn Marsman, Kerstin Hummer, Georg Kresse, Iann C Gerber, and János G Ángyán. Screened hybrid density functionals applied to solids. *The Journal of chemical physics*, 124(15), 2006.

[24] Wahyu Setyawan and Stefano Curtarolo. High-throughput electronic band structure calculations: Challenges and tools. *Computational materials science*, 49(2):299–312, 2010.

[25] Henry Wu, Tam Mayeshiba, and Dane Morgan. High-throughput ab-initio dilute solute diffusion database. *Scientific data*, 3(1):1–11, 2016.

[26] Xi Zhang, Sergiy V Divinski, and Blazej Grabowski. Ab initio machine-learning unveils strong anharmonicity in non-arrhenius self-diffusion of tungsten. *Nature Communications*, 16(1):394, 2025.

[27] Zi-Kui Liu, Yi Wang, and ShunLi Shang. Thermal expansion anomaly regulated by entropy. *Scientific reports*, 4(1):7043, 2014.

[28] Jong Hyun Jung, Prashanth Srinivasan, Axel Forslund, and Blazej Grabowski. High-accuracy thermodynamic properties to the melting point from ab initio calculations aided by machine-learning potentials. *npj Computational Materials*, 9(1):3, 2023.

[29] Wenhao Sun, Stephen T. Dacek, Shyue Ping Ong, Geoffroy Hautier, Anubhav Jain, William Davidson Richards, Anthony Collins Gamst, Kristin A. Persson, and Gerbrand Ceder. The thermodynamic scale of inorganic crystalline metastability. *Science Advances*, 2, 2016.

[30] Jonathan Schmidt, Jingming Shi, Pedro Borlido, Limin Chen, Silvana Botti, and Miguel A. L. Marques. Predicting the thermodynamic stability of solids combining density functional theory and machine learning. *Chemistry of Materials*, 29:5090–5103, 2017.

[31] Geoffroy Hautier, Christopher C. Fischer, Anubhav Jain, Tim Mueller, and Gerbrand Ceder. Finding nature's missing ternary oxide compounds using machine learning and density functional theory. *Chemistry of Materials*, 22:3762–3767, 2010.

[32] Yuping He, Ekin Dogus Cubuk, Mark D. Allendorf, and Evan J. Reed. Metallic metal-organic frameworks predicted by the combination of machine learning methods and ab initio calculations. *The journal of physical chemistry letters*, 9 16:4562–4569, 2018.

[33] Nilesh Varadan Orupattur, Samir H. Mushrif, and Vinay Prasad. Catalytic materials and chemistry development using a synergistic combination of machine learning and ab initio methods. *Computational Materials Science*, 174:109474, 2020.

[34] Andrew S. Rosen, Shaelyn Iyer, Debmalya Ray, Zhenpeng Yao, Alán Aspuru-Guzik, Laura Gagliardi, Justin M. Notestein, and Randall Q. Snurr. Machine learning the quantum-chemical properties of metal– organic frameworks for accelerated materials discovery. *Matter*, 4:1578–1597, 2020.

[35] Gus L. W. Hart, Tim Mueller, Cormac Toher, and Stefano Curtarolo. Machine learning for alloys. *Nature Reviews Materials*, 6:730–755, 2021.

[36] William Hume-Rothery. Xxx. a note on the connexion between chemical valency, electron grouping, and crystal structure. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 3(14):301– 305, 1927.

[37] NJ Petch. Citation classic-the cleavage strength of polycrystals. *Current Contents/Engineering Technology & Applied Sciences*, (19):24–24, 1982.

[38] T Egami and Y_ Waseda. Atomic size effect on the formability of metallic glasses. *Journal of non-crystalline solids*, 64(1-2):113–134, 1984.

[39] JJ Lewandowski*, WH Wang, and AL Greer. Intrinsic plasticity or brittleness of metallic glasses. *Philosophical Magazine Letters*, 85(2):77–87, 2005.

[40] W.P. Weeks, K.M. Flores, Using characteristic structural motifs in metallic liquids to predict glass forming ability, Intermetallics 145 (2022) 107560.

[41] Sungwoo Sohn, Naijia Liu, Geun Hee Yoo, Aya Ochiai, Jade Chen, Callie Levitt, Guannan Liu, Samuel Charles Schroers, Ethen Thomas Lund, Eun Soo Park, et al. A framework for plasticity in metallic glasses. *Materialia*, 31:101876, 2023.

[42] Catherine Dreyfus and Gérard Dreyfus. A machine learning approach to the estimation of the liquidus temperature of glass-forming oxide blends. *Journal of non-crystalline solids*, 318(1-2):63–78, 2003.

[43] Yi-Tao Sun, Hai-Yang Bai, Mao-Zhi Li, and Wei-Hua Wang. Machine learning approach for prediction and understanding of glass-forming ability. *The journal of physical chemistry letters*, 8(14):3434–3439, 2017.

[44] Logan Ward, Stephanie C O'Keeffe, Joseph Stevick, Glenton R Jelbert, Muratahan Aykol, and Chris Wolverton. A machine learning approach for engineering bulk metallic glass alloys. *Acta Materialia*, 159:102–111, 2018.

[45] Guannan Liu, Sungwoo Sohn, Corey S O'Hern, Anna C Gilbert, and Jan Schroers. Effective subgrouping enhances machine learning prediction in complex materials science phenomena: Inoue's subgrouping in discovering bulk metallic glasses. *Acta Materialia*, 265:119590, 2024.

[46] Valentin Stanev, Corey Oses, A Gilad Kusne, Efrain Rodriguez, Johnpierre Paglione, Stefano Curtarolo, and Ichiro Takeuchi. Machine learning modeling of superconducting critical temperature. *npj Computational Materials*, 4(1):29, 2018.

[47] Jia Li, Baobin Xie, Qihong Fang, Bin Liu, Yong Liu, and Peter K Liaw. High-throughput simulation combined machine learning search for optimum elemental composition in medium entropy alloy. *Journal of Materials Science & Technology*, 68:70–75, 2021.

[48] Y. Sohail, C. Zhang, D. Xue, J. Zhang, D. Zhang, S. Gao, Y. Yang, X. Fan, H. Zhang, G. Liu, J. Sun, E. Ma, Machine-learning design of ductile FeNiCoAlTa alloys with high strength, Nature 643(8070) (2025) 119-124.

[49] Guillaume Deffrennes, Kei Terayama, Taichi Abe, Etsuko Ogamino, and Ryo Tamura. A framework to predict binary liquidus by combining machine learning and calphad assessments. *Materials & Design*, page 112111, 2023.

[50] Vahe Gharakhanyan, Luke J Wirth, Jose A Garrido Torres, Ethan Eisenberg, Ting Wang, Dallas R Trinkle, Snigdhansu Chatterjee, and Alexander Urban. Discovering melting temperature prediction models of inorganic solids by combining supervised and unsupervised learning. *The Journal of Chemical Physics*, 160(20), 2024.

[51] Shufeng Kong, Dan Guevarra, Carla P Gomes, and John M Gregoire. Materials representation and transfer learning for multi-property prediction. *Applied Physics Reviews*, 8(2), 2021.

[52] Eric S Muckley, James E Saal, Bryce Meredig, Christopher S Roper, and John H Martin. Interpretable models for extrapolation in scientific machine learning. Digital Discovery, 2(5):1425–1435, 2023.

[53] Chandramouli Nyshadham, Matthias Rupp, Brayden Bekker, Alexander V Shapeev, Tim Mueller, Conrad W Rosenbrock, Gábor Csányi, David W Wingate, and Gus LW Hart. Machine-learned multi-system surrogate models for materials prediction. *npj Computational Materials*, 5(1):51, 2019.

[54] Albert P Bartók, Sandip De, Carl Poelking, Noam Bernstein, James R Kermode, Gábor Csányi, and Michele Ceriotti. Machine learning unifies the modeling of materials and molecules. *Science advances*, 3(12):e1701816, 2017.

[55] Akira Miura, Tsukasa Hokimoto, Masanori Nagao, Takashi Yanase, Toshihiro Shimada, and Kiyoharu Tadanaga. Prediction of ternary liquidus temperatures by statistical modeling of binary and ternary ag–al–sn–zn systems. *ACS omega*, 2(8):5271–5282, 2017.

[56] Hiroaki Okamoto, TB Massalski, et al. Binary alloy phase diagrams. *ASM International, Materials Park, OH, USA*, 12, 1990.

[57] Ankit Rohatgi. Webplotdigitizer user manual version 3.4. *URL http://arohatgi. info/WebPlotDigitizer/app*, pages 1–18, 2014.

[58] Ethen Thomas Lund, Salena Huang, Sebastian A Kube, Guannan Liu, Nathan Johnson, Wade Colley, Apurva Mehta, Barbara K Reck, Sungwoo Sohn, and Jan Schroers. A general indicator for the tolerance to impurities of metals and alloys. Materialia, 33:102037, 2024.

[59] ON Senkov and DB Miracle. A new thermodynamic parameter to predict formation of solid solution or intermetallic phases in high entropy alloys. *Journal of Alloys and Compounds*, 658:603–607, 2016.

[60] Yong Zhang, Yun Jun Zhou, Jun Pin Lin, Guo Liang Chen, and Peter K Liaw. Solid-solution phase formation rules for multi-component alloys. *Advanced engineering materials*, 10(6):534–538, 2008.

[61] Robert Floyd Sekerka. *Thermal physics: thermodynamics and statistical mechanics for scientists and engineers*. Elsevier, 2015.

[62] Akira Takeuchi and Akihisa Inoue. Classification of bulk metallic glasses by atomic size difference, heat of mixing and period of constituent elements and its application to characterization of the main alloying element. *Materials transactions*, 46(12):2817–2829, 2005.

[63] Salena Huang, Sebastian A Kube, Nathan S Johnson, Sungwoo Sohn, Apurva Mehta, and Jan Schroers. Glass formation during combinatorial sputtering in binary alloys. Acta Materialia, page 121240, 2025.

[64] Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. *Introduction to linear regression analysis*. John Wiley & Sons, 2021.

[65] Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

[66] Yann LeCun, Yoshua Bengio, and Geoffrey E. Hinton. Deep learning. 2015.

[67] Jooyoung Park and Irwin W. Sandberg. Universal approximation using radial-basis-function networks. *Neural Computation*, 3:246–257, 1991.

[68] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. *ArXiv*, abs/1912.01703, 2019.