Composition driven machine learning for unearthing high-strength lightweight multi-principal element alloys

Mengxing Li[*1,2], Xiu Kun Quek[*2], Hongli Suo[1], Delvin Wuu[2], Jing Jun Lee[2], Wei Hock Teh[2], Fengxia Wei[2], Riko I Made[2], Dennis Cheng Cheh Tan[2], Si Rong Ng[2], Siyuan Wei[2], Andre Kai Yuan Low[2,4], Kedar Hippalgaonkar[2,4], Yee-Fun Lim[#3], Pei Wang[#2], Chee Koon Ng[#2]

[1]*College of Materials Science and Engineering, Beijing University of Technology, 100 Pingleyuan, Chaoyang District, Beijing 100124, China*

[2]*Institute of Materials Research and Engineering, Agency for Science, Technology and Research (A*STAR), Fusionopolis Way, #08-03 Innovis, Singapore 138634, Singapore*

[3] *Institute of Sustainability for Chemicals, Energy and Environment (ISCE[2]), Agency of Science, Technology and Research (A*STAR), 1 Pesek Road, Singapore 627833, Singapore*

[4] *School of Materials Science and Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore, 639798, Singapore*

[*]These authors contributed equally

[#]Corresponding authors. Emails: ngck@imre.a-star.edu.sg, wangp@imre.a-star.edu.sg, limyf@isce2.a-star.edu.sg

Abstract

High-strength, lightweight alloys are highly desired in the aerospace, automotive and marine industries. The optimization of such alloys is a multifaceted process, characterized by the consideration of diverse objectives and constraints. Various optimization methodologies exist, spanning from intricate first-principle approaches to the application of sophisticated

machine learning algorithms. These algorithms might incorporate input features encompassing elemental composition, microstructural attributes, and thermodynamic properties to enhance prediction accuracies. In this work, we aim to streamline this complexity by employing solely the alloy's elemental composition as the input feature for the machine learning algorithm, improving the hardness while reducing the density of the alloy. We have curated a comprehensive database comprising 544 multi-principal element alloys and developed a robust surrogate model based on these compositions. This composition-driven model is subsequently coupled with principal component analysis (PCA) to facilitate the selection process. Remarkably, through a mere three iterations involving 14 samples, we successfully identified an alloy with an effective specific hardness surpassing the training database maximum by 8.6%. The proposed composition-driven machine learning delineates a simplified approach for conducting optimization across multiple target material properties.

Keywords: Machine learning; Bayesian optimization; Multi-principal element alloys; Effective specific hardness

1 Introduction

In the automotive and aerospace industries, weight is an extremely important factor impacting vehicle performance, and weight reduction is typically seen as the primary method of improving performance and fuel efficiency. The ever-increasing need and demand for stronger materials, coupled with increased pressure for reduction of greenhouse gas emissions from increasingly stringent regulations worldwide has led researchers to maintain considerable interest in the pursuit of these high-strength, light-weight materials. Multi-principal element alloys (MPEAs) are a type of alloy material composed of multiple principal elements that were proposed in 2004 by Yeh *et al*. and Cantor *et al.* [1, 2] which have shown to possess high corrosion resistance [3-5] as well as unprecedented hardness and strength [6, 7]. However, MPEAs' complex combination of elements result in an immense search space. The increase in elements increases the dimensionality of the search space as compared to traditional alloys, becoming cost prohibitive for researchers to experimentally validate. There are approximately 143,000 possible 5-component alloy systems with 30 starting elements to choose from [8], and the search space for potential alloys exponentially increases as the number of elements increase. However, inexpensive computing power and widespread availability of machine learning frameworks and libraries has allowed researchers to efficiently narrow their search space and increase their experimental throughput. Rickman *et al*. [9] successfully designed new MPEAs with extremely high hardness by combining ML surrogate model with experimental validation. ML models of MPEAs predicting phase selection [10, 11], hardness [12-15], ductility [16, 17] and catalytic performances [18] are often seen with high accuracy (R^2 > 0.85). The high-dimensionality of the elemental composition search space led researchers to introduce additional descriptors (usually derivatives of alloy composition) such as atomic radius difference (δ), entropy of mixing (ΔS_{mix}), enthalpy of mixing (ΔH_{mix}), valence electron concentration (VEC), cohesive energy (E_c), electronegativity mismatch (χ_P), free electron

concentration (e/a) and work function (w) in an effort to improve predictive accuracy for ML algorithms. These descriptors along with computational tools such as CALculation of PHAse Diagrams (CALPHAD) have been used in the modelling of MPEA properties [9, 11, 19-22]. However, it is worth noting that the Hume-Rothery and Pauling's rules used in the formation of some of these derivatives were substantiated using binary alloys and has shown to be inconsistent when applied on multicomponent MPEAs [10]. Modelling of MPEA properties (such as hardness) is possible using composition-only [12, 23] or with additional descriptors [9, 13-15, 21, 24]. Wen *et al.* [21] utilised a material design strategy that combines a ML surrogate mode with design algorithms to optimize for MPEA hardness and managed to exceed the maximum hardness of their database by 10% in 7 experiments.

There are various techniques we can employ in exploring the vast search space of MPEAs. Exhaustive search techniques iterate over all permutations within a provided search space and restrictions. E.g., by using a restriction of nine elements (Al, Co, Cr, Cu, Fe, Mn, Ni, Ti, W) at a step size of 5 at.% and maximum concentration of 35 at.%, there are 1,992,195 possible synthetic compositions. However, these methods are often slow and computationally expensive to evaluate. Researchers have tried methods including the use of Bayesian Optimization (BO), Genetic Algorithms, Particle Swam Optimization (PSO) as well as Design of Experiments search heuristics to optimize experimental work [15, 21, 25]. BO refers to a class of optimization strategies that utilizes a surrogate to model an expensive black-box function cheaply [26]. One of the most popular approaches is the "Kriging Method", which utilizes the robustness, flexibility and uncertainty prediction of the Gaussian Process (GP) [27]. Using probabilistic methods to compute the surrogate's predictions with an uncertainty distribution, various acquisition functions can be built to maximize the desired target. Compared with evolution-based algorithms such as GA and PSO, BO's strength is in the relatively low number of function evaluations during the optimization process [28]. Furthermore, BO can also be

extended to different implementations, e.g. parallelization, multi-fidelity hyperparameter tuning, optimization of multiple noisy objectives [29, 30], and has been used for high throughput experimental platforms in materials science [31].

In this work, we explore the idea of ML-assisted experimentation to design strong yet lightweight alloys. We trained composition-only ML surrogate models paired with BO to identify optimal alloy compositions with high effective specific hardness (hardness divided by density of alloy). Subsequently, we employed Principal Component Analysis (PCA) data reduction techniques to screen alloys for experimental validation. The screening techniques via PCA are intuitive for both alloy optimization and surrogate model improvement. Our work has produced an alloy with an effective specific hardness of 187.6 $HV/g.cm^{-3}$ (844 HV / 4.5 $g.cm^{-3}$), which is 8.6% higher than the database maximum of 172.8 $HV/g.cm^{-3}$, with just three rounds of iterations. Notably, our approach emphasizes the utilization of elemental composition exclusively, avoiding additional complicated features. This deliberate simplification not only enhances computational efficiency but also broadens accessibility for researchers lacking access to computationally demanding procedures.

2 Design Strategy

We adopted an incremental modelling framework with an emphasis on both computational and experimental efficiency. Our framework incrementally increases the complexity of the design strategy to simplify the already complex field of multi-principal element alloys. The strategy was implemented in four distinct phases (Fig. 1): the compilation and updating of multi-principal element alloy hardness data available in literature, training of surrogate models, search techniques for the virtual alloy search space, and screening techniques to determine which alloys to experimentally validate. Iteration 1 coupled plug & play models with exhaustive search, selecting only the highest predictions to fabricate experimentally. Iteration 2 utilized

tuned-models with exhaustive search followed by Principal Component Analysis (PCA) guided sampling in the latent space for experimental fabrication. Iteration 3 utilized the best performing model in iteration 2 coupled with Bayesian Optimization (BO) to shortlist the top search results followed by PCA guided sampling for experimental fabrication.

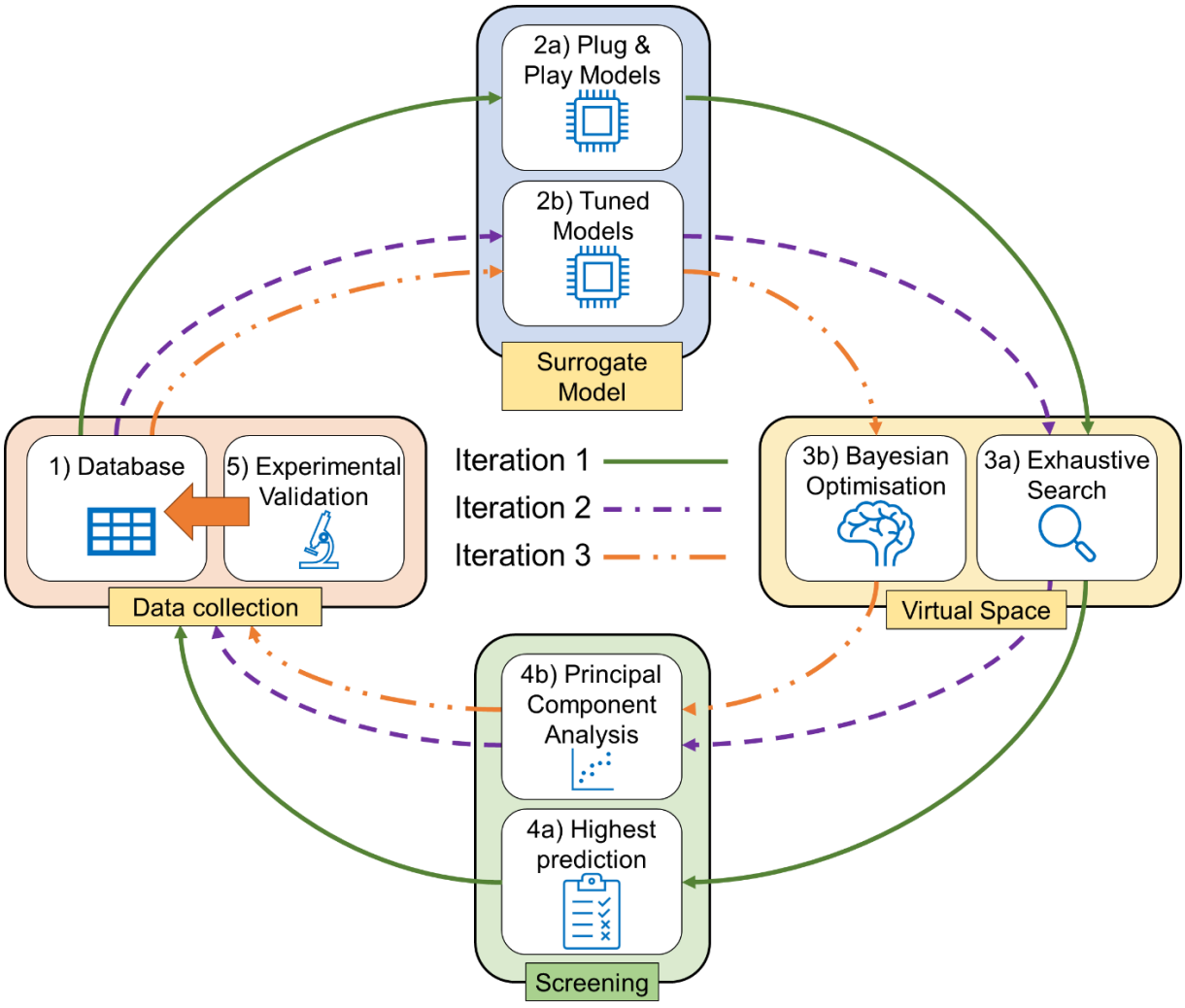

Fig. 1: Schematics of our composition-driven machine learning framework.

2.1 Data collection

We collected a total of 2,253 MPEA alloy compositions from published academic literature [8, 15, 32]. All compositions are expressed via atomic percentage (at. %) and the constituent elements in any composition sum up to 100%. We included only alloys synthesized via arc melting/casting from iteration 2 onwards, as there were insufficient details on specific heat

treatment and fabrication methods in the alloy database. Additionally, all reported hardness measurements were obtained at room temperature. For compositions with duplicate entries, the mean hardness values of the entries were taken and data with relative error exceeding 10% were discarded. This resulted in 544 unique compositions. The screened data consists of 4 ternary, 70 quaternary, 234 quinary, 188 six-component alloys and 44 seven-component alloy systems. The remaining 4 alloys consisted of alloys with eight or more components. A total of 25 elements from the periodic table were present. In the interest of simplicity, apart from compositional data, no additional descriptors or features were generated. The training subset consisted of a randomly selected 90% of the overall data and the remaining 10% was used as the holdout validation subset.

2.2 Surrogate Model

Here we build ML surrogate model denoted as $\hat{y} = \hat{f}(\boldsymbol{x}, w)$. This model aimed to establish a connection between the elemental composition and alloy properties, approximating the relationship between MPEA compositions and their corresponding properties, $y = f(\boldsymbol{x})$, where y represents target property (Vickers Hardness, Yield Strength, Elongation) and $\boldsymbol{x}$ is the input vector consisting of n-dimensions in the form $(x_1, x_2, x_3, \ldots, x_n)$, $\hat{y}$ is the predicted target properties and w is the model set of hyperparameters. The exact form of f is unknown and challenging to determine. The use of surrogate model automates this process with reasonable prediction accuracy through training. Our study involves the implementation and comparison of commonly used ML models: Neural Networks (NN), Random Forest (RF), Gradient Boosting (GB) and Gaussian Process (GP).

2.3 Model selection and evaluation

Typically, several ML models are created and compared by evaluating their performance on the holdout subset. This is achieved through two metrics: Mean squared error (MSE) and

coefficient of determination (R^2). The mean squared error (Eqn. 1) [33] is the average of the squares of errors and is one of the most used loss function for regression problems in ML. The ML algorithms seek to minimise the loss function during the training process to achieve an accurate mapping of the input and target properties. R^2 (Eqn. 2) [33] is simply a statistical measure used to benchmark the goodness-of-fit performance of a model. Given the expected i-th value y_i, the predicted value $\hat{y}_i$ and the mean value $\bar{y}$:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \qquad \text{Eqn. 1}$$

$$R^2(y, \hat{y}) = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \qquad \text{Eqn. 2}$$

Training subset is a randomly selected 90% of the overall dataset. To prevent overfitting or selection bias, 10-fold cross validation is employed by further partitioning the training subset into 10 folds. For each of the 10 folds, one-fold is selected as the testing subset and the model is trained on the other nine folds. The training and testing results across the 10 folds are thereafter averaged to obtain the optimal model.

NN models mimic the anatomy of the human brain through layers of artificial neurons fully connected to each other. Supplementary Materials Fig. S1 shows that for a neuron i in a hidden layer, inputs x_j are multiplied by a specified weight w_j and a bias constant θ is added to obtain the output $y = \sum w_j x_j + \theta$. It is important to note that the weights and bias values of the neurons are randomly generated using a pseudo-random number generator (PRNG) upon initialization of the neural network models. The weights are modified during the training process until the best-fitting output is attained for a known input. In addition, the stochastic gradient descent (SGD) algorithm used to train NNs drastically simplifies the convergence method by randomly selecting weights to modify. It has also been shown that varying the random weights can affect model performance [34] and it is these sources of randomness

inherent in the NN architecture that make it is extremely unlikely that two models will converge in exactly the same manner, holding all other variables constant, leading to variance and noise in single-model predictions. Ensemble models seek to reduce this variance by constructing multiple models in parallel and averaging the output of multiple models may provide better approximation of the ground truth than a single model [35].

Iteration 1 utilized surrogate-based optimization with screening across all feasible alloy compositions. CatBoost (CB) and Gaussian Process (GP) were selected to model our desired target (Effective Specific Hardness), achieving R^2 above 0.85. The input features and output target used for modelling each iteration can be found in Table S1. In iteration 2, the surrogate-based optimization was guided by sampling in the PCA latent space (elaborated in section 2.4 below). We compared the performance of commonly used models, including Random Forest (RF), Gradient Boosting (GB), XGBoost (XGB), Gaussian Process (GP), Neural Network (NN) and Neural Network Ensemble (NNE) models and optimized their hyperparameters (Table S2). Our hyperparameters tuning results (Table S2) showed that NN-based models exhibited some of the best performances (R^2 = 0.93). Fig. 2 depicts a single instance of iteration 2's NN model, five of these models were constructed in parallel and their average output is used for the NNE. We trained a custom surrogate model using Keras' Neural Network (NN) regressor written within sklearn "BaseEstimator" template. For a single NN, output of a hidden layer is given by [36]:

$$Z_m = \sigma\left(a_{m0} + \sum_{p=1}^{p} \alpha_{mp} \mathrm{X}_p\right), m = 1, \ldots, M$$

where Z_m is the output of the *m*th node of a hidden layer with M nodes, X_p are the inputs of this hidden layer, α_{mp} are the weights connecting the previous layer with this layer, and σ is the Rectified Linear Unit (ReLU) activation function which can handle nonlinear relationships between input and output parameters:

$$\sigma(x) = \max(0, x)$$

A single trained NN is susceptible to variance due to the random reinitialization of weights [37], causing slightly different predictions every time the model is reinitialized. Neural Network Ensembles (NNE) seek to reduce this variance by averaging the output of multiple NN models. In our NNE implementation, we averaged the output of five NNs to obtain our final predictions.

In iteration 3, BO was incorporated to reduce computational overhead by only proposing top 40 candidates instead of screening the entire composition space. BO as an optimization strategy is extremely popular in ML-assisted experimentation due to its sample efficient approach. BO optimization techniques also allow for searching over continuous search spaces as opposed to the discrete search space used in synthetic composition generation. BO comprises of an acquisition function that controls the optimization run through the exploration vs exploitation trade-off [38]. Good exploration allows the model to better learn the search space, while good exploitation prioritizes better objective value. There are 3 traditional acquisition functions [39]; Probability of Improvement (PI), which seeks to improve the score of the next best point; Expected Improvement (EI) which selects the next point most likely to improve gradient; and Lower Confidence Bound (LCB), which seeks to minimize regret/loss across optimization. In iteration 3, we opted for the EI acquisition function in our implementation of BO. Our Bayesian Optimization (BO) implementation was achieved by using skopt's library's built-in "gp_minimize" routine. We used the "Expected Improvement" (EI) acquisition function to select the next point for sampling, defined by [40]:

$$EI_n(x) = E[f(x) - f(x_t^+)]$$

Here, E[f(x)] represents the expected value of the posterior function f(x), which is derived by fitting the NNE to the observed function values at the previous sample points $x_1, x_2, \ldots x_n$, and x_t^+ is the highest value observed so far from *t* evaluations performed. The subsequent point x_{n+1} is chosen where the value of x provides the maximum amount of expected improvement:

$$x_{n+1} = argmax|EI_n(x)|$$

The optimization routine is run for 100 iterations and repeated 10 times. Whilst BO was used to generate 40 compositions with high effective specific hardness (more than 175 HV/g.cm^{-3}), we deferred to PCA in the sampling of alloys for experimental validation.

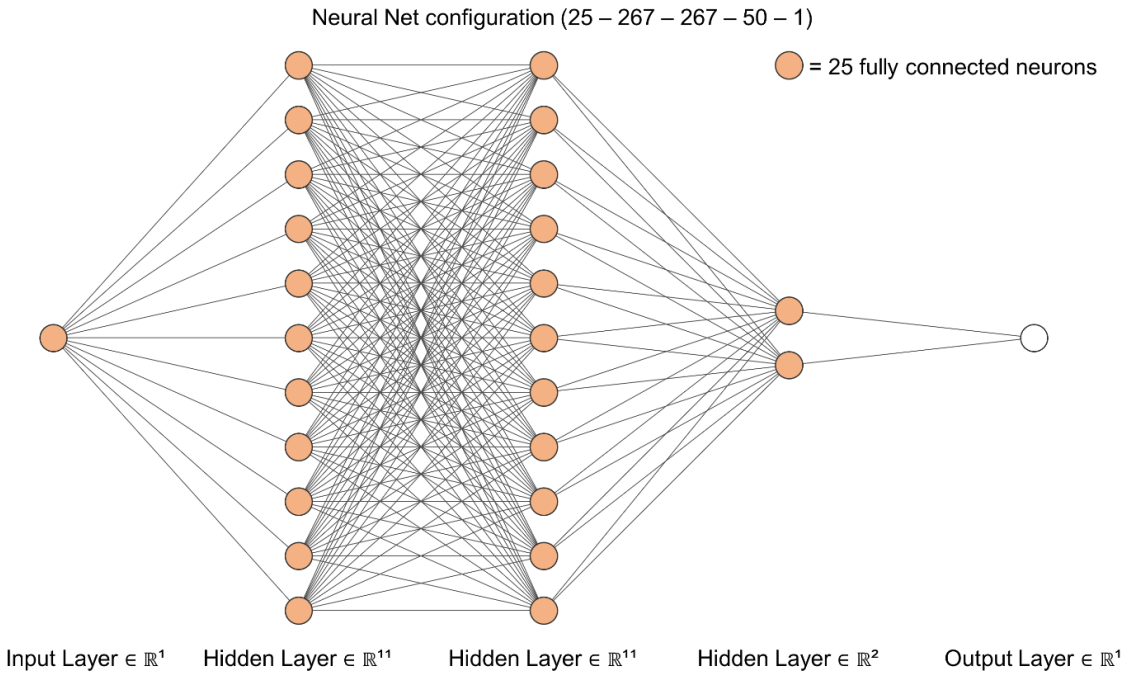

Fig. 2: Iteration 2 single neural network configuration prior to ensemble construction. Due to the width of the model, each orange node represents approximately 25 fully connected neurons.

2.4 Principal Component Analysis

Principal Component Analysis (PCA) is a data transformation technique used for dimensionality reduction. A principal component axis is obtained by mean-centering and rotation of the dataset to obtain an axis with the highest variance. This first axis is fixed whilst the process is then repeated until all axes (and variance in the dataset) is obtained. Typically, only the first few axes containing the greatest cumulative variance is retained for plotting. When used in this manner, PCA is used as a data reduction method that projecting a large, complex dataset onto a smaller set of smaller principal components [41]. PCA allows us to reduce our 25-dimension input parameters (consisting of multiple elemental compositions) into a readily

accessible three-dimension dataset for visualisation. The cumulative variance retained by PCA performed on iterations 2 and 3 were 67.5% and 66.1%, respectively.

The resulting plots on the PCA enables us to visually differentiate similar (lower Euclidean distance) or dissimilar (further Euclidean distance) alloys according to their compositions. The key design requirements for high-strength lightweight MPEAs in this work is to maximize the hardness while minimizing the density of the alloy, i.e. to maximize the effective specific hardness of these alloys. To do so, we could select either "exploitation" or "exploration" points from the BO generated alloy compositions. Alloy compositions with lower Euclidean distance to datapoints with high effective specific hardness was considered "exploitation". Conversely, high Euclidean distance to datapoints with high effective specific hardness was considered "exploration". In iterations 2 and 3, at least half of the selections were selected to serve as exploration points. These results, even if unfavourable, serve to improve the accuracy of subsequent iterations through exploring undiscovered regions in the search space [42]. For iteration 1, the compositions selected for fabrication were chosen to ensure the largest variance in composition in the top 25 results with the highest effective specific hardness. For iterations 2 and 3, three-dimensional PCA was used in the selection process.

2.5 Experimental validation

All the raw materials with purities of higher than 99.9% (wt.%) were weighed according to the designed MPEA compositions. The ingots (25 mm in diameter and 10 mm in height, each about 20 g) were prepared by arc melting and then remelted at least five times in a hemispherical water-cooled Cu mould to ensure the microstructure homogeneity. All samples were sequentially ground and polished using sandpapers of #400, #600, #800, #1000, #2000, and #4000, followed by vibratory polishing with a 0.05 μm colloidal silica suspension to achieve a final mirror finish for SEM and EBSD samples. The phase structure of the samples was

determined using a Bruker D8 Advance X-ray diffractometer (XRD) with Cu Kα radiation, scanning from 20° to 100° with a step size of 0.02° and a dwell time of 0.5 s per step. The microstructure was observed using a JEOL JSM-7600F scanning electron microscope (SEM) equipped with energy-dispersive X-ray spectroscopy (EDX) for elemental and phase analysis and mapping, and an IT500HR system with an electron backscatter diffraction (EBSD) detector. The initial EBSD data were processed using HKL Channel 5 software (Oxford). Vickers hardness was measured using a Falcon 5000HVRB hardness tester under a 1 kgf load, with measurements taken at 2 mm intervals across the sample center and a dwell time of 15 s. The representative hardness value for each alloy is reported as the average of at least 10 measurements.

3 Results and Discussion

We completed three iterations of our framework. After each iteration, the experimental results are integrated into the training data for subsequent iterations. The variations in proposed vs experimental alloy compositions (determined via SEM-EDX) are likely due to experimental errors (Table S3). The experimental compositions and hardness of previous iterations are then fed back into the database to train the next iteration. The summary of the experimental validation results of all three iterations can be found in Table 1 below.

3.1 Iterations

Table 1: Iteration 1, 2 and 3 experimental validation results

Alloy	Predicted Effective Specific Hardness ($HV/g.cm^{-3}$)	Experimental Effective Specific Hardness ($HV/g.cm^{-3}$)
Sample 1-1	128.6	138.9 ± 2.2

Sample 1-2	123.6	132.9 ± 2.4
Sample 1-3	133.8	144.9 ± 2.3
Sample 1-4	147.2	158.9 ± 5.2
Sample 2-1	154.8	156.7 ± 2.6
Sample 2-2	160.1	154.1 ± 5.7
Sample 2-3	152.9	142.5 ± 2.2
Sample 2-4	140.4	159.5 ± 2.6
Sample 2-5	155.8	176.4 ± 3.0
Sample 3-1	225.1	89.9 ± 2.6
Sample 3-2	214.3	132.2 ± 2.2
Sample 3-3	197.2	187.6 ± 4.7
Sample 3-4	180.4	156.5 ± 7.8
Sample 3-5	173.8	149.4 ± 10.7

In iteration 1, CatBoost (R^2 = 0.87) and Gaussian Process (R^2 = 0.85) were used as surrogate models and trained with 466 datapoints from the Borg dataset [8]. Due to the limited amount of hardness data available in the dataset, the alloys were selected regardless of their processing method (casted, wrought, powder, etc…). CatBoost was selected as it was one of the best performing regression models ranked by PyCaret [43] and Gaussian Process was selected based on our previous work on its superior extrapolation capabilities [44]. Exhaustive search of the synthetic alloy compositions was conducted, and four compositions (CatBoost for samples 1-1 and 1-2, Gaussian Process for samples 1-3 and 1-4) were chosen for experimental validation (Table 1) based on their highest maximum effective specific hardness and largest observed compositional variation among the top 25 results for each model.

In iteration 2, refinements were made to the database as some material descriptors (e.g. valence electron configuration, electronegativity, work function, etc...) and elemental

compositions (Ag, Ca, Ga, Li, Mg) did not significantly affect model performance. These descriptors were removed, additional datapoints from relevant literature [15, 32] were added into our updated database and data inconsistencies from the different databases were rectified (refer to section 2.1). In iterations 2 and 3, only the as-casted alloys were used for training due to an increase in number of hardness data available, leading to an increase to 544 datapoints used. The baseline and hyperparameter tuned models were re-evaluated in iteration 2 using these updated datapoints, with the NNE (n = 5) being one of the best performing models (R^2 = 0.93) and was subsequently selected as the surrogate model.

Fig. 3 shows the scatter plot of the ML model's performance using the holdout validation subset. The exhaustive search of the alloy compositions was repeated, and the top 200 predictions based on effective specific hardness were plotted on a 3-D PCA plot. PCA guided selection was performed to determine which alloys were experimentally validated. Readers are encouraged to use the interactive .html file provided in the Supplementary Materials for clarity in the selection process due to the limitations and intricacies of displaying three-dimension plots on a two-dimension medium. Samples 2-1 and 2-5 were selected to exploit their relative proximity to high effective specific hardness datapoints (mostly more than 140 HV/g.cm^{-3}). Samples 2-3 and 2-4 lay near regions that had datapoints less than 100 HV/g.cm^{-3} and Sample 2-2 lay in a region with no datapoints at all. These points may represent either: a) local maxima in the explored compositional space that have not previously been discovered or b) unexplored regions in the compositional space. Validation of these points serve to improve the ML model. Sample 2-5 exhibited a theoretical effective specific hardness of 176.4 HV/g.cm^{-3}, which has exceeded the database maximum by 2%.

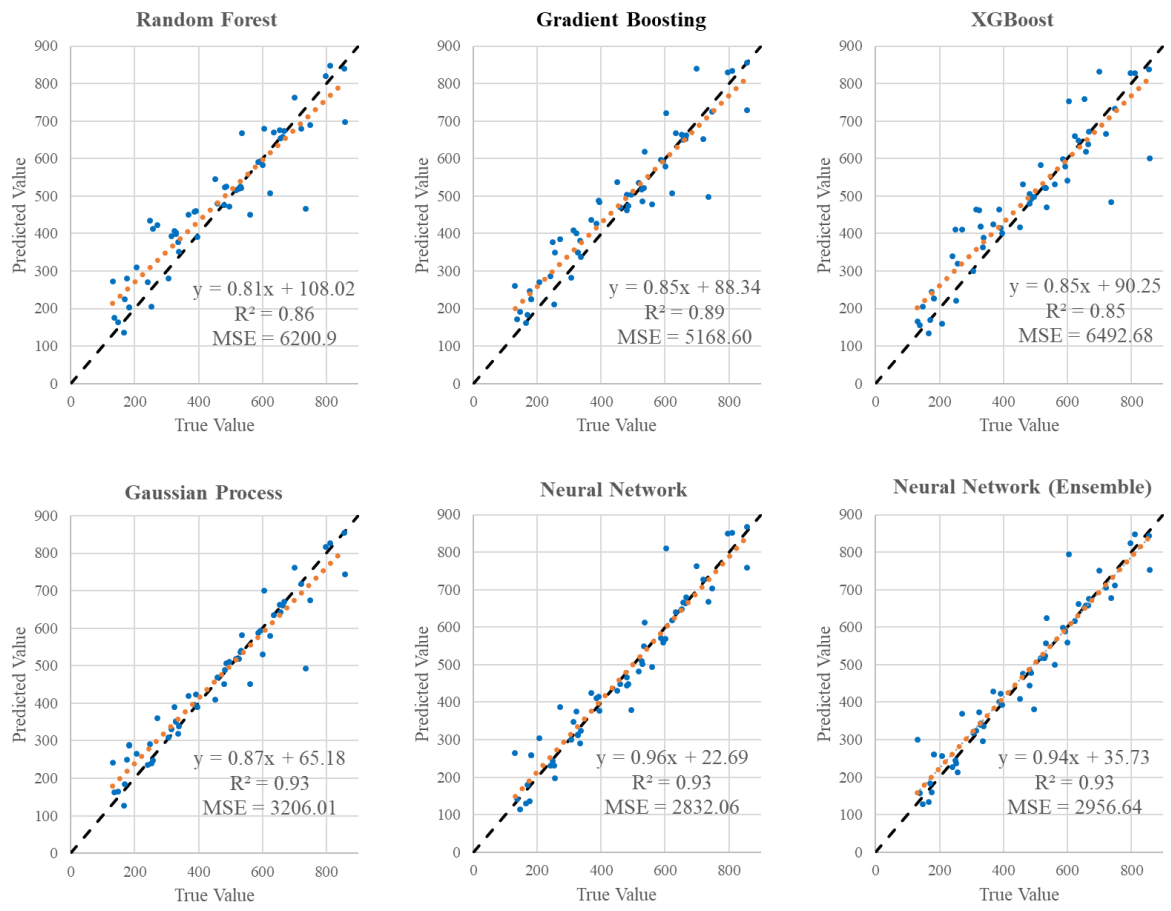

Fig. 3: Scatter plot of holdout testing showing predicted (vertical axis) and true values (horizontal axis) of the following ML models: (a) Random Forest (R^2 = 0.86), (b) Gradient Boosting (R^2 = 0.89), (c) XGBoost (R^2 = 0.85), (d) Gaussian Process (R^2 = 0.93), (e) Single Neural Network (R^2 = 0.93) and (f) Neural Network Ensemble (R^2 = 0.93).

Table 2: Iteration 2 model performance (R^2) comparison

Model type	Random Forest	Gradient Boosting	XGBoost	Gaussian Process	Neural Networks	
					Single	Ensemble (n = 5)
Baseline	0.84	0.78	0.78	0.81	0.56	—
Hyperparameter-tuned	0.86	0.89	0.85	0.93	0.93	0.93

Iteration 3 focused on reducing the computational time required to search the alloy virtual compositional space. With skopt's Bayesian Optimization routine [45], computational time is significantly reduced from 14 hours (through generating synthetic compositions using exhaustive search of almost 2 million compositions) to approximately 2 hours. Samples 3-1 to 3-5 were generated as a result whereby Samples 3-1 and 3-2 are exploration points, and Samples 3-3 to 3-5 were exploitation points. Sample 3-3 ($Al_{47.3}Fe_{23.8}Ti_{28.9}$) was predicted to have an effective specific hardness of 197.2 HV/g.cm^{-3} and its experimental effective specific hardness was found to be 187.6 HV/g.cm^{-3} (Table 1). Its experimental density (measured via Archimedes' method) was 4.5 g.cm^{-3}, which is close to the theoretical density of 4.3 g.cm^{-3}. Our work has, in only three iterations, obtained a sample with an experimental effective specific hardness of 187.6 HV/g.cm^{-3}, which is 8.6% higher than the database maximum of 172.8 HV/g.cm^{-3} (Table 3).

Table 3: Comparison of this work's best results with the database.

Alloy	Effective Specific Hardness (HV/g.cm^{-3})	Improvement (%)	Reference
Sample 3-3 ($Al_{47.3}Fe_{23.8}Ti_{28.9}$)	**187.6**	**+8.6%**	This work
Sample 2-5	176.4	+2.0%	This work
$Al_{47}Co_{20}Cr_{18}Cu_5Fe_5Ni_5$	172.8	—	[21]
$FeCoNiCrAlSiCuTiMoB_{0.5}$	171.6	—	[46]

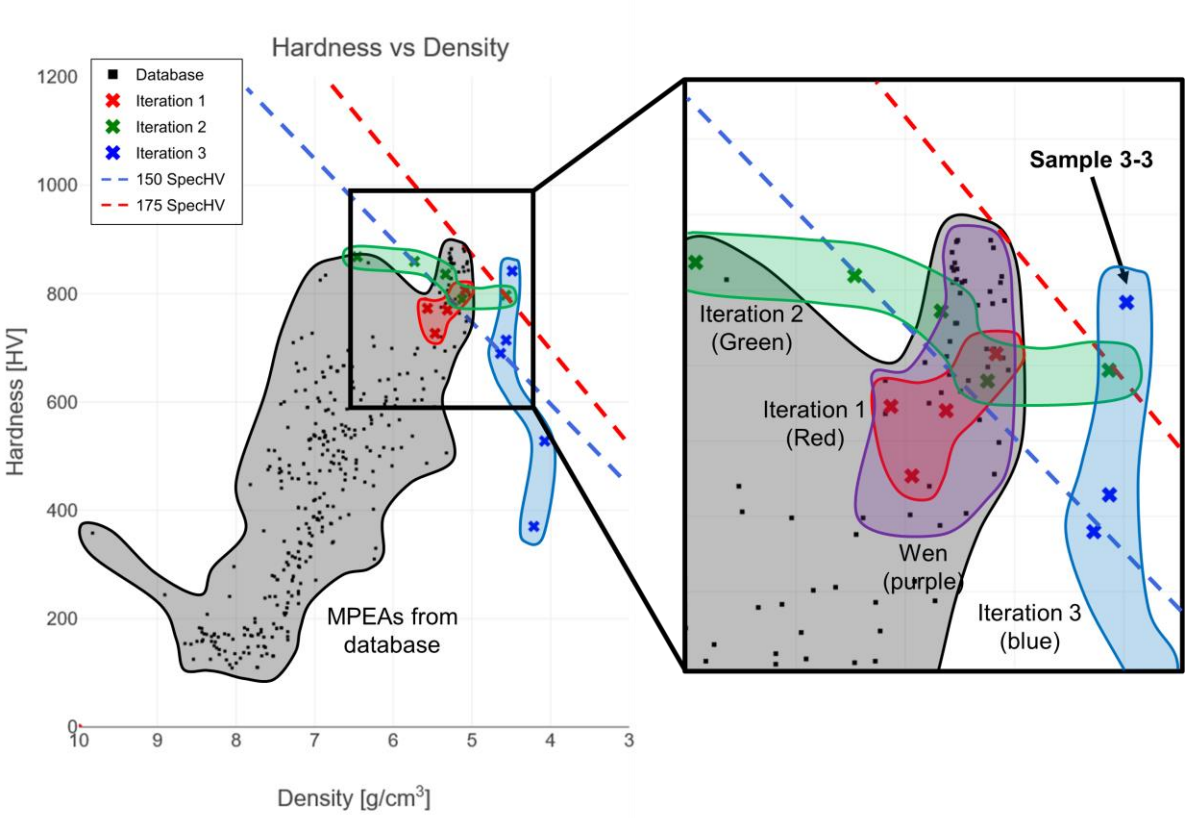

Fig. 4: Hardness vs Density plot of the data in this work, inclusive of results from all iterations. Inset on the right shows the high effective specific hardness area of the plot. Gray shaded region represents literature compiled data. Red, green, and blue shaded region represent experimental validation results from iteration 1, iteration 2, and iteration 3, respectively. A subset of training data from Wen *et al.* [21], is enclosed with purple shade on the inset. Blue and red dashed line represent effective specific hardness points of 150 and 175 HV/g.cm^{-3}, respectively.

Fig. 4 shows a plot of Hardness (vertical axis) versus Density (horizontal axis). The 150 HV/g.cm^{-3} blue dotted line benchmarks the top 5% of effective specific hardness in the database, and there is no data in the database which exceeds the red dotted line of 175 HV/g.cm^{-3}. Between the iterations, it is apparent that using PCA guided sampling in iterations 2 and 3 ensures a wide variation in composition that allows for a balanced approach to exploration of the search space. Therefore, alloys from iteration 2 and 3 have a wide range of results compared

to iteration 1. Another key factor in this work is the implementation of BO as an optimization strategy. Wen *et al.* [21] utilised a material design strategy that combines a ML surrogate mode with design algorithms to optimize for HEA hardness and managed to exceed the maximum hardness of their database by 10% in 7 experiments. In Fig. 4's inset, we highlight the alloys from Wen (in purple) and note that the visual representation of their optimization routine is largely vertical in nature as the optimization focused mainly on increasing the hardness of the alloy. Comparatively, our iteration 3 (denoted in blue on the same graph) has a diagonal shift due to the optimization of both hardness and density simultaneously. The BO specifically targeted reducing alloy density as a way of improving effective specific hardness for the output compositions (all of which were less than 5 $g.cm^{-3}$).

3.2 Feature analysis using SHAP

ML models are typically seen as 'black-boxes' where the input-output relationship is typically unclear and unpredictable. Shapley values, introduced in 1951 by Lloyd Shapley, are used as predictive rules in cooperative game theory. In 2017, Lundberg and Lee [47] introduced a model-agnostic method of approximating feature importance through the use of SHAP (SHapley Additive exPlanations) values, which are Shapley values of a conditional expectation function of the original model. The SHAP values function as an amalgamated measure of feature importance. In Fig. 5, SHAP was used to assess and explain the effects of features in iteration 3's NNE model. The horizontal axis denotes the positive (or negative) impact of a feature, the vertical axis lists the distinctive features, each dot represents a datapoint from the database, and the colour of the datapoint denotes the value of the feature. Datapoints with high feature values are denoted in red whereas low feature values are denoted in blue. For individual features, the wider the plot range of SHAP values, the greater the influence on the output. Fig. 5 shows that Al and Ti have the highest impact on effective specific hardness which does not

come as a surprise as these elements have been known for their high strength-to-density ratios [48, 49], whereas Cu is known to be a soft and malleable metal with higher density. We postulate that the strengthening effects of Ni is outweighed by its high density (8.9 $g.cm^{-3}$ as compared to Al's 2.7 $g.cm^{-3}$). V has been shown to provide strengthening in FCC and BCC MPEAs and alloys containing V have significantly higher strength over a large temperature range [50].

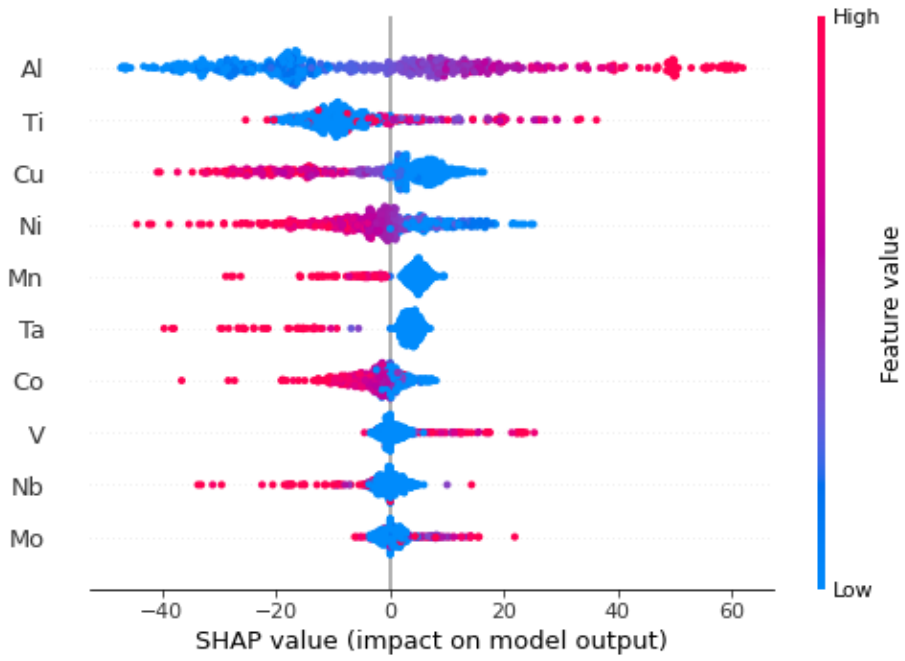

Fig. 5: Analysis of top ten features obtained from SHAP value distribution based on the NNE model.

3.3 Characterization of sample 3-3 ($Al_{47.3}Fe_{23.8}Ti_{28.9}$)

Further characterization of Sample 3-3 ($Al_{47.3}Fe_{23.8}Ti_{28.9}$) using XRD, SEM imaging, EDX elemental distribution mapping and EBSD phase mapping (Fig. 6) show that the alloy exhibits a dual-phase microstructure comprised of an FCC ($Mn_{23}Th_6$ type) and a tetragonal phase. As shown in Fig. 6b, the alloy exhibits a dual-phase microstructure consisting of grey primary dendritic phases and dark grey inter-dendritic phases. Through XRD, EDS, and EBSD, it has

been revealed that the inter-dendritic region is Al-rich and of τ_2* phase, while the dendrites are Ti-rich with τ_2 phase (Fig. 6c). The detailed elemental compositions of dendrites (region A) and inter-dendritic region (region B) are listed in Table 6. The Al–Fe–Ti system is a relatively well-established ternary alloy system with its microstructure [51-54] and properties [55, 56] widely investigated. When its composition ranges from $Al_{24.6}Fe_{24.5}Ti_{50.9}$ to $Al_{47.8}Fe_{21.4}Ti_{30.8}$, the system tends to form FCC structures known as τ_2, while with lower Ti contents, i.e., from $Al_{49.0}Fe_{27.0}Ti_{24.0}$ to $Al_{53.6}Fe_{25.1}Ti_{21.3}$, the expected phase is tetragonal τ_2* [57]. Palm [57] also confirmed the presence of "Al_2FeTi" (τ_2) in various reported ternary compounds. Given that the composition of our designed Sample 3-3 ($Al_{47.3}Fe_{23.8}Ti_{28.9}$) lies between the two regions, it is not surprising that both τ_2 and τ_2* phases are observed. The high hardness of the alloy can also be attributed to these hard intermetallic τ_2 and τ_2* phases [58, 59].

Table 6: Elemental concentrations of the different phases determined by EDX analysis. Refer to Fig. 6b for areas A and B.

Phase/Elements(at%)	Al	Ti	Fe
τ_2 (Area A)	43.9	32.0	24.1
τ_2* (Area B)	50.7	25.1	24.2

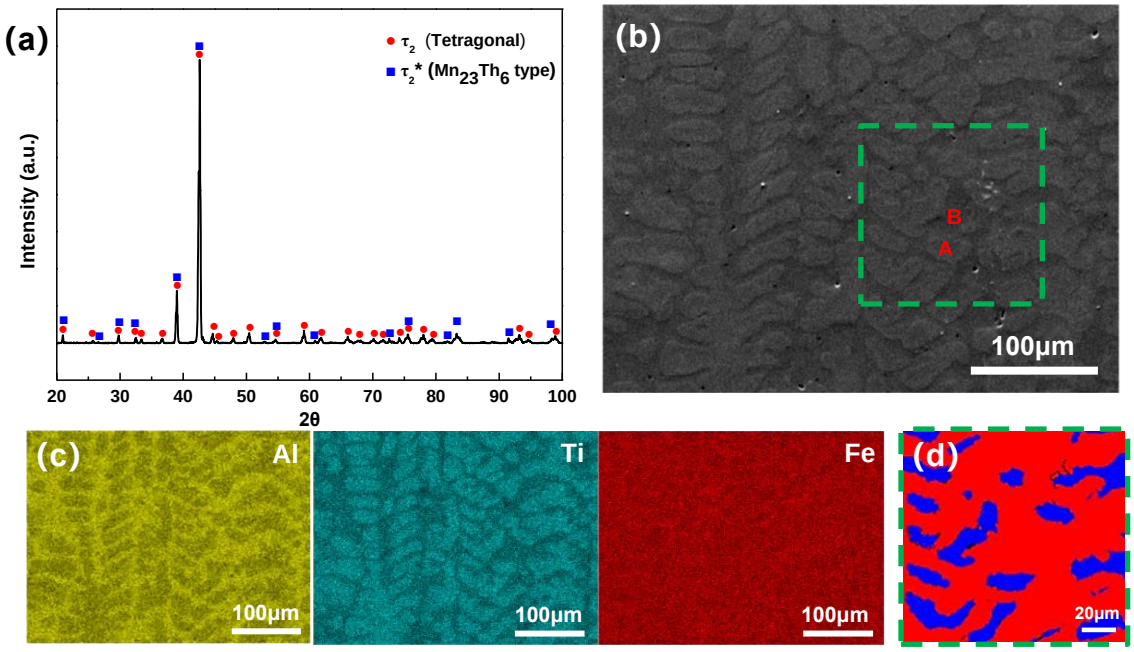

Fig. 6: Characterization of Sample 3-3 ($Al_{47.3}Fe_{23.8}Ti_{28.9}$). (a) The XRD results showing $Mn_{23}Th_6$ type complex FCC (τ_2) phase and tetragonal (τ_2*) phase. (b) SEM image of Sample 3-3, showing the dendritic region A and inter-dendritic region B (c) Element distribution map showing the inverse distribution of Al and Ti, respectively and the uniformity of Fe. (d) EBSD phase mapping of Sample 3-3, blue and red represent τ_2 and τ_2* phases, respectively.

4 Conclusion

In summary, we leverage on a composition-driven machine learning approach to unearth high effective specific hardness MPEAs. Our approach utilized a combination of surrogate models, Bayesian Optimization (BO) and principal component analysis (PCA) to explore and exploit the vast compositional space efficiently, enabling the rapid screening and multi-objective optimization of alloy properties. Through three iterative cycles, we identified the $Al_{47.3}Fe_{23.8}Ti_{28.9}$ alloy, which demonstrated an exceptional effective specific hardness of 187.6 HV/g.cm^3—surpassing the previous maximum in the training database by 8.6%. Importantly, this hardness value was experimentally validated, confirming the robustness of our predictive

model. This framework is composition-only, eliminating the need for additional feature derivation, and highlights a significant advancement in the search for high-performance materials. The successful application of machine learning underscores a shift from traditional trial-and-error methods toward data-driven approaches, offering the potential to develop new alloys that could outperform established superalloys, providing stronger and lighter materials for various industries.

Acknowledgements

This work acknowledges the support of the National Research Foundation Singapore and DSO National Laboratories under the AI Singapore Programme (AISG Award No: AISG2-GC-2023-010), Agency for Science, Technology and Research (A*STAR) under Structural Metal Alloy Program (A18b1B0061) and Manufacturing, Trade and Connectivity (MTC) Industry Alignment Fund- Pre-Positioning (IAF-PP) (Award M22K7a0047), the National Natural Science Foundation of China (52277021) and Beijing Municipal Natural Science Foundation (2212025). M. Li acknowledges the support of the China National Scholarship Council.

Data availability

The data that support the findings of this study are available upon request.

References

[1] J.W. Yeh, S.K. Chen, S.J. Lin, J.Y. Gan, T.S. Chin, T.T. Shun, C.H. Tsau, S.Y. Chang, Nanostructured High-Entropy Alloys with Multiple Principal Elements: Novel Alloy Design Concepts and Outcomes, Adv. Eng. Mater. 6(5) (2004) 299-303.
[2] 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.
[3] Y. Qiu, M.A. Gibson, H.L. Fraser, N. Birbilis, Corrosion characteristics of high entropy alloys, Mater. Sci. Technol. 31(10) (2015) 1235-1243.
[4] Y. Qiu, S. Thomas, M.A. Gibson, H.L. Fraser, N. Birbilis, Corrosion of high entropy alloys, npj Mater. Degrad. 1(1) (2017) 1-18.
[5] K.F. Quiambao, S.J. McDonnell, D.K. Schreiber, A.Y. Gerard, K.M. Freedy, P. Lu, J.E. Saal, G.S. Frankel, J.R. Scully, Passivation of a corrosion resistant high entropy alloy in non-oxidizing sulfate solutions, Acta Mater. 164 (2019) 362-376.
[6] Z. Li, K.G. Pradeep, Y. Deng, D. Raabe, C.C. Tasan, Metastable high-entropy dual-phase alloys overcome the strength–ductility trade-off, Nature 534(7606) (2016) 227-230.
[7] Y. Qiu, S. Thomas, D. Fabijanic, A.J. Barlow, H.L. Fraser, N. Birbilis, Microstructural evolution, electrochemical and corrosion properties of AlxCoCrFeNiTiy high entropy alloys, Mater. Des. 170 (2019) 107698.
[8] C.K.H. Borg, C. Frey, J. Moh, T.M. Pollock, S. Gorsse, D.B. Miracle, O.N. Senkov, B. Meredig, J.E. Saal, Expanded dataset of mechanical properties and observed phases of multi-principal element alloys, Sci. Data 7(1) (2020) 430.
[9] J.M. Rickman, H.M. Chan, M.P. Harmer, J.A. Smeltzer, C.J. Marvel, A. Roy, G. Balasubramanian, Materials informatics for the screening of multi-principal elements and high-entropy alloys, Nat. Commun. 10(1) (2019) 2618.
[10] W. Huang, P. Martin, H.L. Zhuang, Machine-learning phase prediction of high-entropy alloys, Acta Mater. 169 (2019) 225-236.
[11] Y. Zeng, M. Man, C.K. Ng, D. Wuu, J.J. Lee, F. Wei, P. Wang, K. Bai, D.C. Cheh Tan, Y.-W. Zhang, Machine learning-based inverse design for single-phase high entropy alloys, APL Mater. 10(10) (2022).
[12] R. Jain, S.K. Dewangan, V. Kumar, S. Samal, Artificial neural network approach for microhardness prediction of eight component FeCoNiCrMnVAlNb eutectic high entropy alloys, Mater. Sci. Eng. A 797 (2020) 140059.
[13] Y.-J. Chang, C.-Y. Jui, W.-J. Lee, A.-C. Yeh, Prediction of the Composition and Hardness of High-Entropy Alloys by Machine Learning, JOM 71(10) (2019) 3433-3442.
[14] L. Qiao, Z. Lai, Y. Liu, A. Bao, J. Zhu, Modelling and prediction of hardness in multi-component alloys: A combined machine learning, first principles and experimental study, J. Alloys Compd. 853 (2021) 156959.
[15] C. Yang, C. Ren, Y. Jia, G. Wang, M. Li, W. Lu, A machine learning-based alloy design system to facilitate the rational design of high entropy alloys with enhanced hardness, Acta Mater. 222 (2022) 117431.
[16] M.S. Ozerdem, S. Kolukisa, Artificial neural network approach to predict the mechanical properties of Cu–Sn–Pb–Zn–Ni cast alloys, Mater. Des. 30(3) (2009) 764-769.
[17] G. Kim, H. Diao, C. Lee, A.T. Samaei, T. Phan, M. de Jong, K. An, D. Ma, P.K. Liaw, W. Chen, First-principles and machine learning predictions of elasticity in severely lattice-distorted high-entropy alloys with experimental validation, Acta Mater. 181 (2019) 124-138.
[18] C.Y.J. Lim, I.M. R, Z.H.J. Khoo, C.K. Ng, Y. Bai, J. Wang, G. Yang, A.D. Handoko, Y.F. Lim, Machine learning-assisted optimization of multi-metal hydroxide electrocatalysts for overall water splitting, Mater. Horiz. 10(11) (2023) 5022-5031.

[19] Y. Zhang, Y.J. Zhou, J.P. Lin, G.L. Chen, P.K. Liaw, Solid-Solution Phase Formation Rules for Multi-component Alloys, Adv. Eng. Mater. 10(6) (2008) 534-538.
[20] X. Huang, C. Jin, C. Zhang, H. Zhang, H. Fu, Machine learning assisted modelling and design of solid solution hardened high entropy alloys, Mater. Des. 211 (2021) 110177.
[21] 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.
[22] R. Feng, C. Zhang, M.C. Gao, Z. Pei, F. Zhang, Y. Chen, D. Ma, K. An, J.D. Poplawsky, L. Ouyang, Y. Ren, J.A. Hawk, M. Widom, P.K. Liaw, High-throughput design of high-performance lightweight high-entropy alloys, Nat Commun 12(1) (2021) 4329.
[23] S.K. Dewangan, S. Samal, V. Kumar, Microstructure exploration and an artificial neural network approach for hardness prediction in AlCrFeMnNiWx High-Entropy Alloys, J. Alloys Compd. 823 (2020) 153766.
[24] E. Menou, F. Tancret, I. Toda-Caraballo, G. Ramstein, P. Castany, E. Bertrand, N. Gautier, P.E.J. Rivera Díaz-Del-Castillo, Computational design of light and strong high entropy alloys (HEA): Obtainment of an extremely high specific solid solution hardening, Scr. Mater. 156 (2018) 120-123.
[25] H. Zhang, H. Fu, S. Zhu, W. Yong, J. Xie, Machine learning assisted composition effective design for precipitation strengthened copper alloys, Acta Mater. 215 (2021) 117118.
[26] B. Shahriari, K. Swersky, Z. Wang, R.P. Adams, N. de Freitas, Taking the Human Out of the Loop: A Review of Bayesian Optimization, Proc. IEEE 104(1) (2016) 148-175.
[27] C.E. Rasmussen, K.I.W. Christopher, Gaussian process for machine learning, The MIT Press, London, England, 2006.
[28] A.K.Y. Low, E. Vissol-Gaudin, Y.-F. Lim, K. Hippalgaonkar, Mapping pareto fronts for efficient multi-objective materials discovery, J. Mater. Inf. 3(2) (2023).
[29] J. Wu, S. Toscano-Palmerin, P. Frazier, A. Wilson, Practical Multi-fidelity Bayesian Optimization for Hyperparameter Tuning, 35th Conference on Uncertainty in Artificial Intelligence, UAI 2019, 2019.
[30] S. Daulton, M. Balandat, E. Bakshy, Parallel Bayesian Optimization of Multiple Noisy Objectives with Expected Hypervolume Improvement, Adv. Neural Inf. Process. Syst. 34 (2021) 2187-2200.
[31] L. Andre KY, M.-B. Flore, O. Aleksandr, X. Jiaxun, V.-G. Eleonore, L. Yee-Fun, G. Abhishek, L. Qianxiao, O. Yew Soon, K. Saif A, H. Kedar, Evolution-guided Bayesian optimization for constrained multi-objective optimization in self-driving labs, ChemRxiv (2023).
[32] S. Gorsse, M.H. Nguyen, O.N. Senkov, D.B. Miracle, Database on the mechanical properties of high entropy alloys and complex concentrated alloys, Data Br. 21 (2018) 2664-2678.
[33] M.A. Abdelkareem, B. Soudan, M.S. Mahmoud, E.T. Sayed, M.N. AlMallahi, A. Inayat, M.A. Radi, A.G. Olabi, Progress of artificial neural networks applications in hydrogen production, Chem. Eng. Res. Des. 182 (2022) 66-86.
[34] O.E. Gundersen, S. Shamsaliei, R.J. Isdahl, Do machine learning platforms provide out-of-the-box reproducibility?, Future Gener. Comput. Syst. 126 (2022) 34-47.
[35] T.G. Dietterich, Ensemble Methods in Machine Learning, Springer, 2000, pp. 1-15.
[36] T. Hastie, R. Tibshirani, J. Friedman, Neural Networks, in: T. Hastie, R. Tibshirani, J. Friedman (Eds.), The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer New York, New York, NY, 2009, pp. 389-416.
[37] B. Charles, C. Julien, K. Koray, W. Daan, Weight Uncertainty in Neural Network, PMLR, pp. 1613-1622.

[38] T. Desautels, A. Krause, J.W. Burdick, Parallelizing Exploration-Exploitation Tradeoffs in Gaussian Process Bandit Optimization, J. Mach. Learn. Res. 15(119) (2014) 4053-4103.
[39] D.R. Jones, A Taxonomy of Global Optimization Methods Based on Response Surfaces, J. Glob. Optim. 21(4) (2001) 345-383.
[40] P.I. Frazier, A Tutorial on Bayesian Optimization, 2018, p. arXiv:1807.02811.
[41] W.J. Dunn, D.R. Scott, W.G. Glen, Principal components analysis and partial least squares regression, Tetrahedron Comput. Methodol. 2(6) (1989) 349-376.
[42] L.P. Kaelbling, M.L. Littman, A.W. Moore, Reinforcement Learning: A Survey, J. Artif. Intell. Res. 4 (1996) 237-285.
[43] M. Ali, PyCaret: An open source, low-code machine learning library in Python, 2020.
[44] Y.-F. Lim, C.K. Ng, U.S. Vaitesswar, K. Hippalgaonkar, Extrapolative Bayesian Optimization with Gaussian Process and Neural Network Ensemble Surrogate Models, Advanced Intelligent Systems 3(11) (2021) 2100101.
[45] G. Louppe, M. Kumar, Bayesian optimization with skopt. https://scikit-optimize.github.io/stable/auto_examples/bayesian-optimization.html, 2016).
[46] H. Zhang, Y. He, Y. Pan, Enhanced hardness and fracture toughness of the laser-solidified FeCoNiCrCuTiMoAlSiB0.5 high-entropy alloy by martensite strengthening, Scr. Mater. 69(4) (2013) 342-345.
[47] S.M. Lundberg, S.-I. Lee, A Unified Approach to Interpreting Model Predictions, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.) Curran Associates, Inc., 2017.
[48] E. Georgantzia, M. Gkantou, G.S. Kamaris, Aluminium alloys as structural material: A review of research, Eng. Struct. 227 (2021) 111372.
[49] E.O. Ezugwu, Z.M. Wang, Titanium alloys and their machinability—a review, J. Mater. Process. Technol. 68(3) (1997) 262-274.
[50] B. Yin, F. Maresca, W.A. Curtin, Vanadium is an optimal element for strengthening in both fcc and bcc high-entropy alloys, Acta Mater. 188 (2020) 486-491.
[51] V. Raghavan, Al-Fe-Ti (aluminum-iron-titanium), J. Phase Equilib. 23(4) (2002) 367.
[52] A. Grytsiv, P. Rogl, V. Pomjakushin, Structural transition with loss of symmetry in Ti–M–Al based G-phases (MFe and Co), Intermetallics 14(7) (2006) 784-791.
[53] M. Palm, The Al–Cr–Fe system–Phases and phase equilibria in the Al-rich corner, J. Alloys Compd. 252(1) (1997) 192-200.
[54] M. Palm, G. Inden, N. Thomas, The Fe-A1-Ti system, J. Phase Equilib. 16(3) (1995) 209-222.
[55] M.C.J. Marker, L.I. Duarte, C. Leinenbach, K.W. Richter, Characterization of the Fe-rich corner of Al–Fe–Si–Ti, Intermetallics 39 (2013) 38-49.
[56] U. Prakash, G. Sauthoff, Structure and properties of Fe–Al–Ti intermetallic alloys, Intermetallics 9(2) (2001) 107-112.
[57] M. Palm, J. Lacaze, Assessment of the Al–Fe–Ti system, Intermetallics 14(10) (2006) 1291-1303.
[58] O.E. Matts, S.Y. Tarasov, B. Domenichini, D.V. Lazurenko, A.V. Filippov, V.A. Bataev, M.V. Rashkovets, I.K. Chakin, K.I. Emurlaev, Tribo-oxidation of Ti-Al-Fe and Ti-Al-Mn cladding layers obtained by non-vacuum electron beam treatment, Surf. Coat. Technol. 421 (2021) 127442.
[59] M. Palm, G. Sauthoff, Deformation behaviour and oxidation resistance of single-phase and two-phase L21-ordered Fe–Al–Ti alloys, Intermetallics 12(12) (2004) 1345-1359.