



OPEN Critical raw material-free multi-principal alloy design for a net-zero future

Swati Singh¹, Mingwen Bai², Allan Matthews³, Saurav Goel^{1,4,5}✉ & Shrikrishna N. Joshi¹✉

Refractory High-Entropy Alloys (RHEAs), such as NbMoTaW, MoNbTaVW, HfNbTaZr, $\text{Re}_{0.1}\text{Hf}_{0.25}\text{NbTaW}_{0.4}$, $\text{Nb}_{40}\text{Ti}_{25}\text{Al}_{15}\text{V}_{10}\text{Ta}_5\text{Hf}_3\text{W}_2$, $\text{Ti}_x\text{NbMoTaW}$ ($x=0, 0.25, 0.5, 0.75$ and 1), and 3d transition metal HEAs such as $\text{Al}_{10.3}\text{Co}_{17}\text{Cr}_{7.5}\text{Fe}_9\text{Ni}_{48.6}\text{Ti}_{5.8}\text{Ta}_{0.6}\text{Mo}_{0.8}\text{W}_{0.4}$ have demonstrated superior performance compared to traditional superalloys, particularly in high-temperature applications for engine components. However, the development of these alloys often depends on critical raw materials (CRMs) such as Ta, W, Nb, Hf, and others. The reliance on critical raw materials (CRMs) not only generates substantial emissions during recycling processes but also imposes considerable risks across global supply chains, hindering the pursuit of Net-zero ambitions. In this pioneering work, we unveil an inventive approach to inversely predict novel multicomponent alloy compositions, meticulously crafted to eliminate CRMs while achieving hardness levels comparable to those of CRM-containing multi-principal element alloys (MPEAs). A robust machine learning (ML) model was developed using a computational database of 3,608 entries, covering unary and binary materials from the Thermo-Calc 2024a software. Among various ML models, the Extra Trees Regressor (ETR) exhibited superior performance and was integrated with metaheuristic optimization techniques to identify novel MPEA compositions. The Cuckoo Search Optimization (CSO) method produced reduced-CRM MPEAs that closely matched Thermo-Calc predictions, with an error margin below $\pm 20\%$. To assess the efficacy of these reduced-CRM MPEAs, we compared the hardness of newly synthesized MPEA with CRM-containing counterparts reported in the literature, particularly those with high-risk critical raw materials like Niobium (Nb) and Tantalum (Ta). For example, the $\text{CoCrFeNb}_{0.309}\text{Ni}$ alloy, which includes CRMs Nb and Co exhibits a Vickers hardness of 480 HV. In contrast, our proposed composition, $\text{Ti}_{0.01111}\text{NiFe}_{0.4}\text{Cu}_{0.4}$ achieves a comparable hardness of 488 HV without using a CRM. Our objective was not to develop high hardness alloy but to facilitate the development of reduced-CRM multi-principal element alloys (R-CRM-MPEAs). We validated our computational approach through the experimental synthesis of an FCC-phase alloy, $\text{Al}_{6.25}\text{Cu}_{18.75}\text{Fe}_{25}\text{Co}_{25}\text{Ni}_{25}$. Thermo-Calc evaluation and ML model predictions of the Vickers hardness showed excellent agreement with the experimental hardness values, which lends credence to our approach. In conclusion, this study provides a robust framework for accelerating the discovery of novel R-CRM-MPEAs, effectively addressing challenges related to supply chain vulnerabilities, import dependence, and related environmental concerns.

Keywords MPEAs, HEAs, Net zero, Machine learning, CRMs

Abbreviations

ABR	AdaBoost Regressor
ACO	Ant colony optimisation
CCAs	Complex concentrated alloys
CRM	Critical raw materials
CSO	Cuckoo search optimisation
DTR	Decision Tree Regressor
ETR	Extra Tree Regressor (ETR)

¹Department of Mechanical Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, India.

²School of Mechanical Engineering, University of Leeds, Leeds LS2 9JT, UK. ³Henry Royce Institute, The University of Manchester, Manchester M13 9PL, UK. ⁴School of Engineering, London South Bank University, 103 Borough Road, London SE1 0AA, UK. ⁵University of Petroleum and Energy Studies, Dehradun 248007, India. ✉email: GoeLs@Lsbu.ac.uk; snj@iitg.ac.in

GA	Genetic algorithm
GBR	Gradient Boost Regressor
HEAs	High entropy alloys
MPEAs	Multi-principal element alloys
ML	Machine learning
PSO	Particle swarm optimisation
R-CRM-MPEAs	Reduced-CRM multi-principal element alloys
RFR	Random Forest Regressor
TC	Thermo-Calc
WOA	Whale optimization algorithm
XGBR	XGBoost Regressor

High entropy alloys (HEAs), also known as multi-principal element alloys (MPEAs) or complex concentrated alloys (CCAs), are solid solution alloys containing five or more elements in equiatomic or near equiatomic proportions (between 5% and 35% atomic concentration). The increase in configurational entropy of mixing elements overcomes the enthalpies of compound formation, inhibiting intermetallic formation, which stabilises the solid solution in a single phase with high configurational entropy^{1,2}.

Cantor alloy (CrMnFeCoNi) was among the first reported MPEA in 2004 that showed exceptional properties in the race of CCAs³. However, Al_x(CrFeCoNiCu) with varying Al concentration ($x=0$ to 3) was also developed at the same time by Yeh⁴ and it was after this effort that the name high-entropy alloy (HEA) was coined. Since then, various CCAs/MPEAs/HEAs have emerged due to their exceptional mechanical properties over conventional alloys.

Various refractory HEAs (RHEAs) such as NbMoTaW⁵, Ti_{1-x}NbMoTaW ($x=0, 0.25, 0.5, 0.75, 1$)⁵ V_xNbMoTa ($x=0.25, 0.5, 0.75, 1.0$)⁶, Nb₄₀Ti₂₅Al₁₅V₁₀Ta₃Hf₃W₂⁷, NbMoTaW(HfN)_x ($x=0, 0.3, 0.7, 1.0$)⁸, MoNbTaVW⁹, HfNbTaZr⁹, Re_{0.1}Hf_{0.25}NbTaW_{0.4}¹⁰ and some 3d transition metal HEAs such as Al_{10.3}Co₁₇Cr_{7.5}Fe₉Ni_{48.6}Ti_{5.8}Ta_{0.6}Mo_{0.8}W_{0.4}¹¹, Al_{10.2}Co_{16.9}Cr_{7.4}Fe_{8.9}Ni_{47.9}Ti_{5.8}Mo_{0.9}Nb_{1.2}W_{0.4}C_{0.4}¹² have been developed for high-temperature applications in aerospace, gas turbine, and nuclear power plants. Other MPEAs such as CoCrFeNiTa_x ($x=0, 0.1, 0.2, 0.3, 0.4, 0.5, \text{ and } 0.75$)¹³, CoCrFeNiNb_x ($x=0, 0.103, 0.155, 0.206, 0.309 \text{ and } 0.412$)¹⁴ and CoCrFeNiNb_x ($x=0.1, 0.25, 0.5 \text{ and } 0.8$)¹⁵ have also shown a remarkable combination of high strength and ductility for their eutectic counterparts such as CoCrFeNiTa_{0.4} and CrFeCoNiNb_{0.5} respectively. However, it is noticeable that these alloys are reliant on the use of critical raw materials (CRMs) such as Ta, W, Nb and Hf¹⁶. The availability of Hf powder is extremely limited¹⁷. Rizzo et al.¹⁶ have alluded to the importance of having a flawless supply chain of raw materials to maintain a sustainable circular economy. Thus, the exigency of minimizing the use of CRMs to mitigate the excess of imports and reducing the need for excessive mining to accelerate the transition to net zero became the prime focus of this study.

The classification of CRM is an important distinction in the realization of this research. Accordingly, various CRMs identified in the past and present were tabulated and classified under three categories (see Fig. 1).

CRMs that appeared three or more times and remained a current concern were classed as 1st Tier. This includes elements such as Be, Bi, Co, Ga, Ge, Hf, Mg, Nb, Sb, Sc, Si Metal, Ta, V and W which are considered most critical in their usage. The 2nd Tier CRMs include elements such as Al, Cu, He, Li, Mn, Ni, P, Sr, and Ti, which have appeared recently as CRMs for more than once. Materials such as Cr and In are not considered CRMs as they were excluded from the 2023 list of CRMs and therefore are not considered as CRM in this study. An exception was made in the algorithm to include Co and V which are although classed as 1st Tier CRMs were not treated as critical. This is due to the ongoing advances made in the recovery and recycling methods and it is expected that the Co and V will become non-critical over time^{23–26}.

A description of strategies adopted in recovering and recycling Co and V has been discussed in Section “Data collection using Thermo-Calc 2024a”. Henceforth, we considered Co and V as non-CRM for the current investigation and included them in the database for our current investigation. Accordingly, we report:

- Preparation of a fresh database reporting Vickers hardness of unary (pure) and binary-based compositions for non-CRM (Al, Cr, Cu, Fe, Ni, Ti, Mo, Mn, Sn, Zn, Zr) and CRM elements (Co and V- treated as non-criti-

Materials Years	Al	Be	Bi	Co	Cr	Cu	Ga	Ge	He	Hf	In	Li	Mg	Mn	Nb	Ni	P	Sb	Sc	Sr	Si Metal	Ta	Ti	V	W
	2011		█		█			█	█			█		█		█			█			█			
2014		█		█	█						█		█		█			█			█				█
2017		█		█					█		█		█		█			█			█				█
2020	█			█		█				█		█		█		█		█			█				█
2023	█			█		█				█		█		█		█		█			█				█

Severity scale

1st Tier CRM	Appeared =>3 consecutive times and is of current concern
2nd Tier CRM	Appeared 1-2 times lately as CRM
Non-CRMs	Not recognised as CRM as of 2023

Fig. 1. The severity scale of materials listed in the category of CRMs by the EU^{18–22}.

- cal in this study) using Thermo-Calc 2024a property model calculator based on CALPHAD approach. A total of 3,608 instances were recorded and no experimental data was incorporated to keep the database free from experimental uncertainty arising from the manufacturing process.
- Training and testing of various tree-based regression models: Decision Tree Regressor (DTR), Random Forest Regressor (RFR), AdaBoost Regressor (ABR), Gradient Boost Regressor (GBR), XGBoost Regressor (XGBR) and Extra Tree Regressor (ETR) based on the developed database and their performance evaluation based on various regression metrics such as coefficient of determination (R^2 _score), mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) for finding the robust regression model (best predictor) to correctly predict hardness of an unseen instance.
 - Optimisation of the best regression model using metaheuristic optimisation techniques such as Particle Swarm Optimisation (PSO), Genetic Algorithm (GA), Ant Colony Optimisation (ACO), Cuckoo Search Optimisation (CSO) and Whale Optimisation Algorithm (WOA) in search of new multicomponent compositions to generate Reduced-CRM multi-principal element alloys (R-CRM-MPEAs) with target Vickers hardness values. Other optimization algorithms such as Artificial Bees Colony (ABC) and Simulated Annealing (SA) were also optimised for the same objective, but they both failed to generate the desired compositions.
 - The ML-predicted hardness values were benchmarked to the corresponding predictions obtained from the Thermo-Calc calculator to test the robustness of the optimization models and compare the percentage error between ML and Thermo-Calc prediction of hardness value for the same compositions.
 - A thorough evaluation was performed on experimentally synthesized CRM-laden MPEAs from the literature to benchmark the predicted R-CRM-MPEAs compositions generated in this study. This comparison aimed to demonstrate the feasibility of partially or fully replacing CRMs while preserving the hardness of the alloy.
 - As a test case, Vickers hardness of a newly predicted composition ($\text{Al}_{6.25}\text{Cu}_{18.75}\text{Fe}_{25}\text{Co}_{25}\text{Ni}_{25}$) developed by our group^{27–29} was measured experimentally and compared with the corresponding Thermo-Calc and ML predicted values successfully which lends credence to the computational prediction.

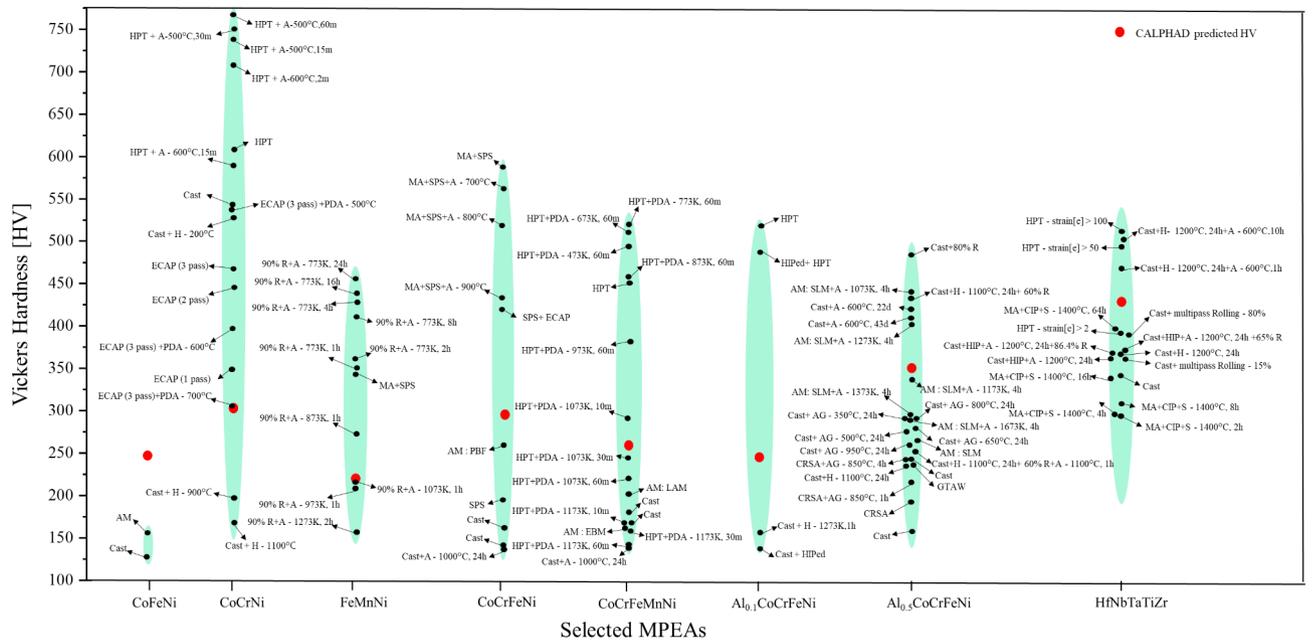
Methodology

Most recent HEAs developed have been obtained by adjusting the composition percentages or substituting one element from an already established HEA. For example, various research articles built the new composition based on cantor alloy (CoCrFeMnNi) either by replacement of (CrFeCoNiCu ³⁰ and TiCrFeCoNi ³¹), variation ($\text{CrMnFeCo}_x\text{Ni}$ and CrMnFeCoNi_x with $x=0-2$ ³²), or by addition (CrMnFeCoNiCu ³, CrMnFeCoNiAl_x ³³) of an element². Additionally, reduction-based alloys have spawned lower-order systems such as binaries, ternaries and quaternary alloys- termed low and medium-entropy alloys². Ten binaries, ten ternaries and five quaternaries' compositions can be made from a cantor alloy. Among two of the ten possible binaries (FeNi and CoNi), five of the ten possible ternaries (CoFeNi , CrFeNi , FeMnNi , CoCrNi and CoMnNi), and three of the five possible quaternary (CoCrFeNi , CoFeMnNi and CoCrMnNi) are single phase FCC solid solution^{34,35}. Interestingly, equiatomic CoCrNi medium entropy alloy shows better mechanical properties than the CoCrFeMnNi HEA, which demonstrates that configurational entropy (increasing the number of elements in an alloy) does not necessarily improve the mechanical property of an alloy³⁶. Moreover, numerous research articles employed various state-of-the-art ML strategies in designing and developing novel MPEAs^{37–40}.

Apart from these compositions, various MPEAs have been developed by utilising elements such as Al, Cu, Cr, Ti, V, W, Ta, Hf, Nb, Mo, Zn, Zr, Si extensively, while certain precious metal HEAs incorporate elements such as Ag, Pt, Au, Ru, Rh, Pd. However, many of these elements (Hf, Nb, Ta, Pt, Pd, Ru, Rh, W) have been marked as critical and have reached an alarming stage⁴¹. Consequently, when considering MPEAs comprised solely of non-CRMs elements, the available experimental data in the literature is scarce which limits the training and testing of the machine learning models. Therefore, we extracted a fresh database using Thermo-Calc 2024a software, which is based on a CALculation of PHase Diagram (CALPHAD) approach. This database contains Vickers hardness values for unary (pure) and binary compositions of materials extracted from Thermo-Calc software using the TCHEA7 database. The work aimed to discover R-CRM-MPEAs compositions (from unary and binary composition databases) with mechanical properties comparable to CRM-laden MPEA compositions.

In recent years, CALPHAD has played a crucial role in designing transition alloys from a completely serendipitous process to a well-established method seeking a thermodynamic rationale³⁶. CALPHAD has been extensively utilised in literature for phase prediction and rapid screening of potential alloys by estimating their compositional and microstructural properties which are validated experimentally^{36,42}. However, no study can be seen in the literature with a focus on mechanical property prediction solely from the CALPHAD method. This is because predicting mechanical properties is not as straightforward as phase prediction. Phase prediction relies solely on the Gibbs free energy for lower-order compositions. For more complex or higher-order compositions, phases are predicted by extrapolating Gibbs free energy from the lower-order systems⁴³. Unlike phase prediction, mechanical property prediction requires rigorous research into the manufacturing (processing) routes, processing parameters, post-processing treatments, testing parameters and extensive knowledge or expertise in the field. Thermo-Calc 2024a offers a property model calculator, which allows the prediction of yield strength and hardness of a composition based on the phases present at a particular temperature⁴⁴.

Currently, Thermo-Calc does not account for factors such as processing history, parameters, time and other crucial variables for accurately predicting mechanical properties under specific experimental conditions. However, its hardness prediction tool still provides a solid foundation for making informed estimations. To assess the discrepancy between experimentally obtained mechanical properties and Thermo-Calc predictions, we extracted the Vickers hardness values of various alloys, including medium- and high-entropy alloys, from experimental literature across different processing methods (casting, additive manufacturing, powder metallurgy, rolling, and severe plastic deformation techniques such as High-Pressure Torsion (HPT) and Equal Channel Angular Pressing (ECAP)). These were then compared to Thermo-Calc predictions, as shown in Fig. 2.



A: Annealing; AM: Additive Manufactured; AG: Aging; CIP: Cold isostatically pressing; CRSA: Casting+70% R +A-1200°C, 1h + Water quenching; d: days; EBM: Electron Beam melting; ECAP: Equal Channel Angular Pressing; GTAW: Gas-tungsten-arc-welding; H: Homogenization; HIP: Hot isostatic pressing; h: hours; High Pressure Torsion; LAM: Laser additive manufacturing; MA: Mechanical Alloying; PBF: Powder bed fusion; PDA: Post deformation annealing; R: Rolling; S: Sintering; SPS: Spark plasma sintering; SLM: Selective Laser Melting

Fig. 2. Vickers hardness comparison for selective MPEAs based on the experimental results obtained from various manufacturing methods (black dots) vs. CALPHAD predicted values (in red dots).

A detailed comparison is provided in tabular form in supplementary Table 1s. Thermo-Calc predictions were seen to be insensitive to the strain rate applied during the manufacturing process which can lead to different hardness values based on the manufacturing process. For a few compositions, the CALPHAD hardness value matched the experimental values, however, it is difficult to generalize which experiments led to the values that are closest to the CALPHAD predictions. For instance, the hardness values obtained from CALPHAD for CoCrNi match with the ECAP processed (for 3 passes) and post-deformation annealed (at 700 °C) samples. As for the FeMnNi medium entropy alloy, its CALPHAD value was closest to the alloy processed via rolling (90% rolled) and then annealed at 1073 K for 1 h (see Fig. 2). Thus, generalizing which experimental processing route leads to hardness values that closely match those predicted by CALPHAD is arduous.

Data collection using Thermo-Calc 2024a

Despite cobalt's inclusion in the 1st tier CRM category, its production is not a concern. The Democratic Republic of Congo (DRC) is the world's largest producer of Co, with its production projected to increase from 11,000 MT in the year 2000 to 98,000 MT by 2020⁴⁵. However, the market of Co presents a considerable risk due to supply-chain complexities. China, which has limited domestic cobalt production has significantly increased its imports from the Democratic Republic of the Congo (DRC) and controls cobalt processing in the region through various Chinese firms. This initiative by China was aimed at securing a competitive advantage in regulating the electric vehicle market.

A potential remedy here would be to recover Co from waste batteries. Suriyanarayanan et al.²⁴ recently introduced an innovative and efficient approach for Co extraction with an extraction efficiency >97%, using a nonionic deep eutectic solvent (ni-DES) comprised of N-methylurea and acetamide. Zhang et al.²⁵ developed a supercritical fluid extraction process using supercritical CO₂ solvent with tributyl phosphate–nitric acid and hydrogen peroxide adduct to recover Li, Co, Mn and Ni with a 90% extraction efficiency. Moreover, Yang et al.²⁶ in 2024 estimated the sales volume of new energy passenger vehicles (NEPV) from 2023 to 2035 based on the historical NEPV sales data from 2013 to 2022. Utilizing Weibull distribution to analyze different sales scenarios, they estimated the potential of recycling Co for maintaining a balance between supply and demand. Their analysis predicted the peak potential of recycling of Co to be about 0.167 MT with an economic value ranging from 49.01 billion to 94.60 billion RMB in 2035. Consequently, they concluded that recycling Co is necessary to alleviate the supply risk pressure and take Co off of the CRM list.

Similarly, Petranikova et al.²³ summarized the efforts in the recovery of Vanadium by selecting more sustainable technologies with lower generation of harmful by-products. They highlighted the importance of combining hydrometallurgical and pyrometallurgical approaches to increase the material recovery rates. With ongoing strategic advancements in recovery and recycling, Cobalt (Co) and Vanadium (V) are expected to transition from their current status as CRMs to non-CRMs. As these management methods evolve, the associated risks related to these materials are anticipated to gradually decrease.

In this context, a large dataset of Vickers hardness values for unary and binary element-based compositions using the property model calculator in Thermo-Calc 2024a (Version 2024.1.132110-55) was compiled by focusing on elements Al, Cr, Cu, Co, Fe, Ni, Ti, V, Mo, Mn, Sn, Zn and Zr.

While extracting the data, it was observed that the property predictor module relies on certain assumptions. Those assumptions are: (i) the material is homogeneous i.e., with no imperfections or defects (ii) it considers only local equilibrium and neglects long-range diffusion. Therefore, to calculate the mechanical properties, it uses simplified theoretical models and databases (containing thermodynamic and kinetic data), which do not capture the complexities related to processing (as-cast, heat-treated, homogenized, or severely deformed samples) and testing conditions (the amount of load and time required in hardness testing). Thermo-Calc predictions can therefore be expected to carry a certain percentage of error when compared to the experimentally synthesized specimen based on its processing history.

For determining the hardness value of a particular composition using Thermo-Calc, the phases present in the system were first identified using their corresponding phase diagram observed at a range of temperatures, computed using an equilibrium calculator. Based on the available features within the software, by considering the system size of 1 mol at a temperature of 300 K and 1 bar pressure, the hardness value was estimated. In some cases, the identified phases in the phase diagram could not be marked while calculating Vickers hardness due to their unavailability in the property model calculator, highlighting one of the several limitations of Thermo-Calc 2024a that needs to be improved.

Consequently, a total of 3,608 instances of different compositions were extracted. This database contains only compositional information and the hardness value of each instance. The highest hardness value obtained was in the range of 400–405 HV. Some of the interesting compositions with higher Vickers hardness values in the database were $\text{Cr}_{42}\text{Ti}_{58}$, $\text{Ti}_{74}\text{Zn}_{26}$, $\text{Ni}_{24}\text{Ti}_{76}$, $\text{Cu}_{16}\text{Ti}_{84}$ with a hardness value of 405HV, 404 HV, 403 HV and 402 HV respectively. The complete dataset is provided as supplementary data. It's worth noting that no experimental data was considered in this database to avoid mixing synthetic data into the prediction model.

Data sorting

The database was first checked to avoid repetition. The screened dataset was divided into 80:20 (2,886 and 722 instances) ratio for training, evaluation and verification.

Tree-based ML model evaluation

Various tree-based regression algorithms such as Decision Tree Regressor (DTR), Random Forest Regressor (RFR), AdaBoost Regressor (ABR), Gradient Boost Regressor (GBR), XGBoost Regressor (XGBR) and Extra Tree Regressor (ETR) were employed. A detailed description of each algorithm with its flowchart is provided in Section “Methodology” as Supplementary information.

The performance of a machine learning (ML) model is influenced by hyperparameters, which are adjustable settings that govern various aspects of the model's learning process, such as complexity, regularisation, and convergence. Examples of hyperparameters include the maximum depth of trees in decision tree models, the number of trees in ensemble models like Random Forest and Gradient Boost, as well as the number of hidden layers in neural networks and the penalty term used in support vector machines. Proper tuning of these hyperparameters is essential for achieving an optimal balance between model accuracy and generalisation.

To optimize the hyperparameters of the selected algorithms, Random Search CV was employed. Unlike Grid Search, which systematically evaluates all possible combinations of hyperparameter values, Random Search, samples a fixed number of hyperparameter combinations from a defined distribution. This approach significantly reduces computational costs while still allowing the exploration of a diverse range of values, making it particularly advantageous for complex models with many hyperparameters.

Each hyperparameter configuration was evaluated using a 5-fold cross-validation approach, where the dataset was divided into five subsets. Each subset was used once as a validation set while the others served as the training set, ensuring a thorough and unbiased assessment of the model's generalization capabilities. The coefficient of determination (R^2 score), mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE), as detailed in Eq. 1 to 4, were used as performance indicators to quantify the difference between predicted values and observed outcomes. The optimized hyperparameters for each regressor model are listed in Table 1. This methodology helped in identifying the best-performing hyperparameter settings, ensuring that the model would perform consistently across different data splits. The working principle of this investigation is illustrated in Fig. 3.

$$R^2_Score = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \mu)^2} \quad (1)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|(y_i - \hat{y}_i)|}{(y_i)} \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

The results obtained from various algorithms are presented in Fig. 4a. The Extra Trees Regressor (ETR) demonstrated a superior performance and achieved an R^2 score of 0.82 and Mean Absolute Percentage Error

Regression algorithm	Hyperparameters
Decision Tree Regressor (DTR)	min_samples_split: 4; min_sample_leaf: 3; max_features: None
Random Forest regressor (RFR)	n_estimators: 651; min_samples_split: 2; min_sample_leaf: 1; max_depth: 70; bootstrap: True max_features: 'sqrt'
Gradient Boost Regressor (GBR)	Subsample: 0.97; n_estimators: 609; min_samples_split: 9; min_sample_leaf: 3; max_depth: 8; learning_rate: 0.042 max_features: 'log2'
AdaBoost Regressor (ABR)	learning_rate: 0.78; loss: linear; n_estimators: 280
XGBoost (XGB)	colsample_bytree: 0.87; gamma: 4.7; learning_rate: 0.22; max_depth: 7; min_child_weight: 5; n_estimators: 1738; subsample: 0.94
Extra Trees Regressor (ETR)	n_estimators: 1301; min_samples_split: 4; min_sample_leaf: 1; max_depth: None; bootstrap: False max_features: 'log2'

Table 1. Summary of the optimized hyperparameters for the chosen ML algorithms.

(MAPE) of 0.17, utilizing optimized hyperparameters determined through Random Search Cross-Validation (see the scatter plot in Fig. 4b). It is important to note that while several studies report even higher R^2 scores, those models typically incorporate numerous descriptors, such as atomic size difference (δ), electronegativity difference ($\Delta\chi$), valence electron concentration (VEC), mixing enthalpy (ΔH_{mix}), mixing entropy (ΔS_{mix}), melting temperature (ΔT_m), Young's modulus (E), shear modulus (G), differences in shear modulus (δG), lattice distortion energy (μ), the Peierls-Nabarro factor (F), and other parameters (Ω -parameter, ϕ -parameter, and γ -parameter), which together enhance performance metrics. In contrast, our approach relies solely on compositional information and hardness values, achieving an R^2 score of 0.82. This underscores the model's robustness to predict hardness values without requiring any additional descriptors.

To explore novel multi-component compositions, we employed a metaheuristic optimization strategy aimed at generating alloys with superior hardness compared to those in the training and testing datasets, which consisted exclusively of unary and binary compositions. The optimization techniques that facilitate the exploration of more complex and high-performance MPEAs are detailed in the subsequent sections. The hardness values of the newly optimized MPEAs were then compared against CALPHAD-based predictions, and these findings are discussed comprehensively in Section "Results and discussions".

Optimisation techniques

The classification of various optimisation models followed in this work is shown in Fig. 3c with red fonts highlighting the algorithms used in this investigation. Traditional optimisation techniques, which include general methods and non-general or specified methods tailored for specific types of problems, have certain limitations such as the requirement of the objective function to be differentiable and lack of ability to obtain a globally optimum solution. Some of the popular traditional optimisation techniques such as Newton Raphson, Successive Quadratic Programming algorithm, Steepest Descent Algorithm, Stochastic Newton optimisation method and Sequential Unconstrained Minimization technique^{46,47} are well known. Recently, some non-traditional methods of optimisation popularly known as meta-heuristic optimisation techniques have gained increasing popularity in solving complex problems. The term metaheuristic combines meta and heuristic, both originated from Greek. Meta symbolizes higher or beyond, and heuristic signifies intelligent guesswork based on past experience or intuitive solution of a problem. Therefore, metaheuristic optimisation can be considered as something beyond intuitive, combined with certain mathematical rules or higher-level frameworks. It can broadly be classified into two categories: single trajectory-based and population-based optimisation. Single-trajectory based optimisation (such as Hill Climbing, Gradient Descent, Tabu Search, Random Search etc.) starts with a single solution at each iteration, and the current solution is replaced by another best solution found in the neighbourhood for that iteration. Contrarily, population-based optimisation techniques are inspired by natural-selection and biological evolution, where a set of solutions are randomly initialized and updated through an iterative process. Genetic algorithm, Differential evolution and others belong to evolutionary optimisation techniques, while Particle swarm optimisation, Ant colony optimisation, Cuckoo search and others are examples

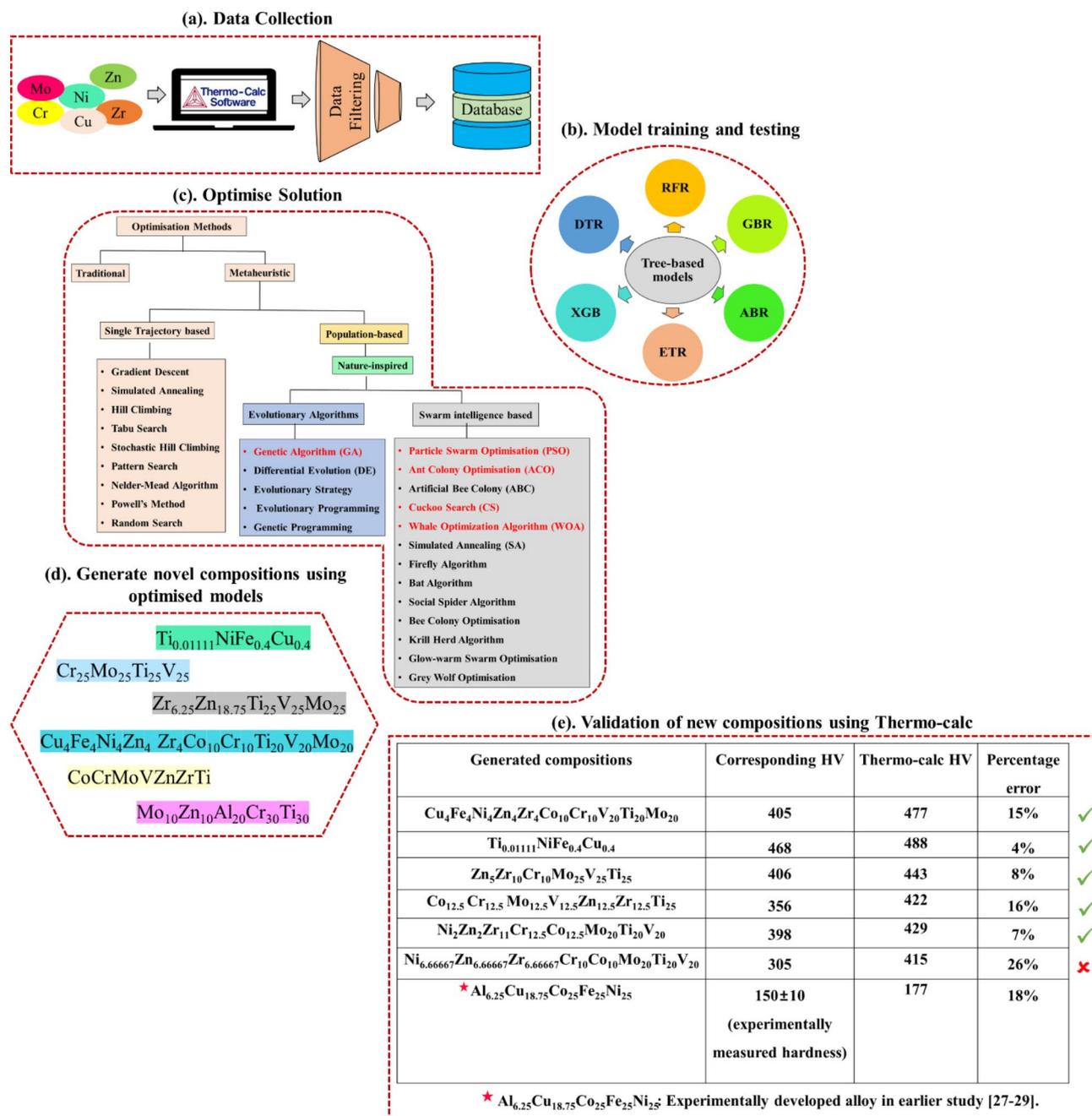


Fig. 3. Workflow of the current study: (a). data collection and processing; (b). training, testing and evaluation of tree-based models; (c). classification of optimization models (full-length image is provided in Section “Optimisation techniques”) and selection of best optimization model (models in red font were selected for this study) for obtaining our objective; (d). generated novel compositions; (e). comparison of the predicted hardness with Thermo-Calc ones for validation; hardness of novel HEA ($Al_{6.25}Cu_{18.75}Co_{25}Fe_{25}Ni_{25}$) prepared in the earlier study was measured and compared with Thermo-Calc one for reinforcing the present study.

of swarm-intelligence-based optimisation techniques^{47,48}. These techniques are exploration-and-exploitation oriented, which introduces diversification in the search space, resulting in the attainment of global optimum solutions by avoiding local optimum solutions for complex real-world problem⁴⁹.

Among all available various nature-inspired metaheuristic optimisation techniques, genetic algorithms, particle swarm optimisation, Cuckoo search and a few others have been proven to successfully solve a wide range of complex real-world problems. However, some of the recently introduced metaheuristic optimisation techniques such as Ant colony optimisation, Artificial Bee Colony optimisation, Spotted Hyena optimisation, whale optimisation need substantiation for their convergence. Rao⁵⁰ in his book chapter has highlighted that the fundamental idea of these methods is the same while naming them differently. In this work, we compared which optimisation technique works best for our ETR model to discover new multicomponent compositions

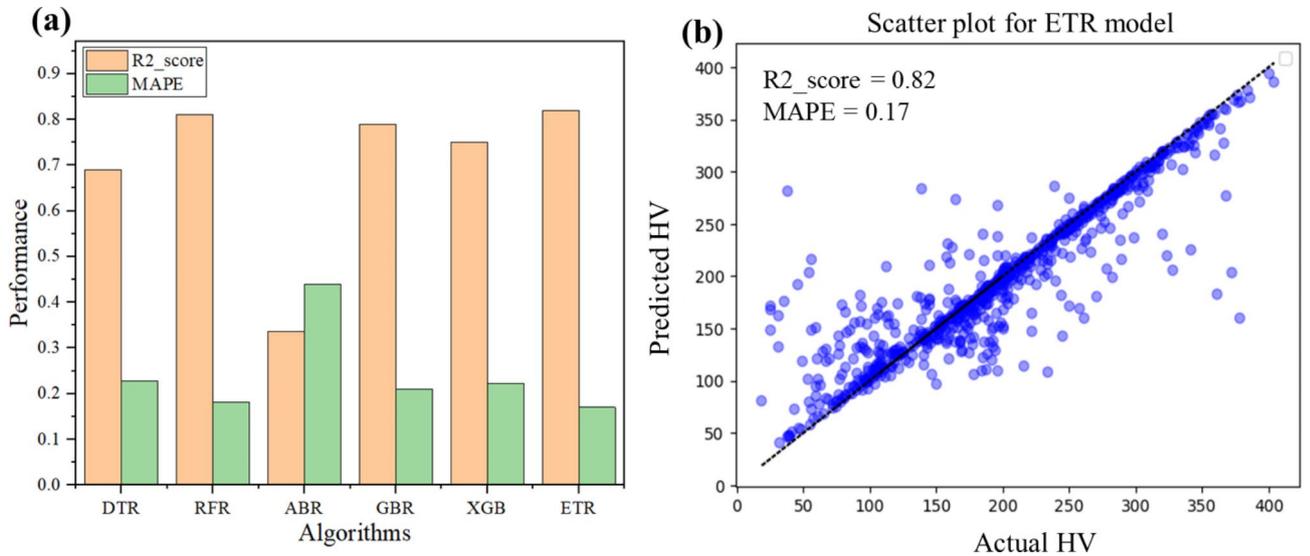


Fig. 4. (a) Performance of various tree-based ML models: R2_score (in light orange), MAPE (in green), (b) actual vs. predicted hardness of test data for the ETR model.

Possible compositions from 13 elements, while considering ternary to decenary based compositions, using combination theory formula:

$$C(n, k) = n! / (k! * (n-k)!)$$

- C (13, 3) = 286,
- C (13, 4) = 715,
- C (13, 5) = 1287,
- C (13, 6) = 1716,
- C (13, 7) = 1716,
- C (13, 8) = 1287,
- C (13, 9) = 715,
- C (13, 10) = 286.

Total sum of all combinations = 8,008 possible combinations.

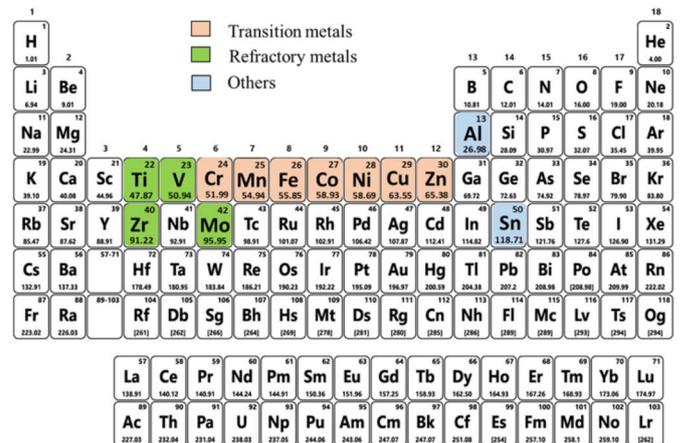


Fig. 5. Total possible compositions from selected 13 elements.

from the unexplored compositional space of nearly 8,008 types (see Fig. 5) as per combination theory while only ternary to decenary-based compositions were considered for 13 elements. Figure 6 describes five metaheuristic optimization techniques used in this study. A detailed description of each optimization technique with its working principle is elaborated in section 3s of the supplementary information.

Results and discussions

Evaluation of several reduced-CRM-MPEAs

Optimisation function was set up to obtain novel multi-principal element alloys with Vickers hardness > 400 HV, by enforcing composition constraint such that the sum of 3, 4, 5, 6, 7, 8, 9, or 10 elements become 100, while keeping the proportion of each element equal or near equal to generate multicomponent compositions.

Among all the optimisation models, cuckoo search optimisation (CSO) provided predictions near the Thermo-Calc predictions. Recent literature suggests that the cuckoo search optimisation (CSO) performs better than PSO, GA, ACO, ABC and WOA^{51–54}. Gandomi et al.⁵¹ provided an extensive comparison and concluded that CSO performs better than GA and PSO, as GA requires a higher number of iterations and its implementation is computationally expensive⁵³. On the other hand, PSO requires less computational effort but considerable execution time to find a solution from a large space for a complex optimisation problem. Civicioglu and Desdo⁸

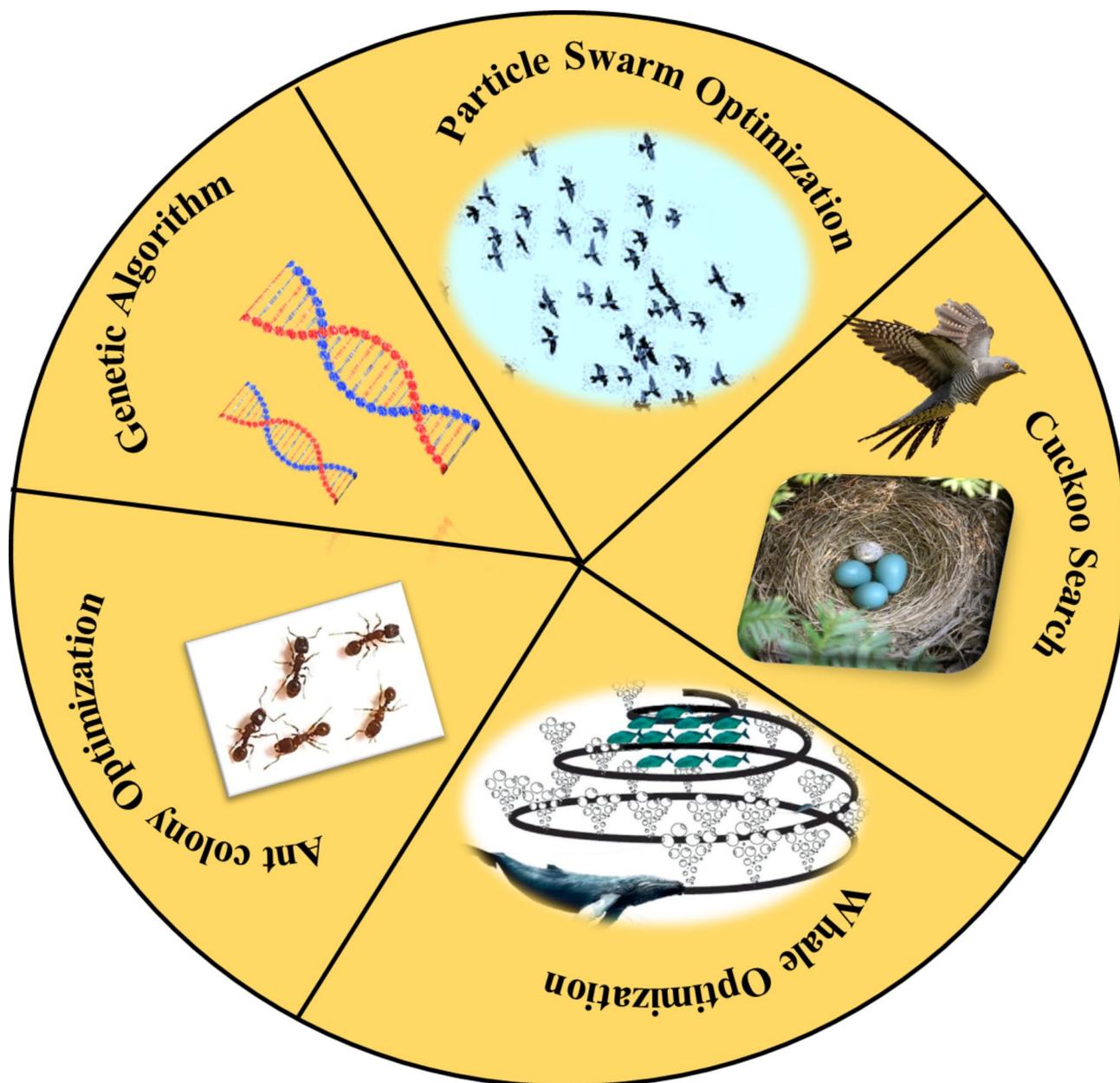


Fig. 6. Various metaheuristic optimisation models applied in this work: (1). Genetic Algorithm (GA), (2). Particle Swarm Optimisation (PSO), (3). Cuckoo Search Optimisation (CSO), (4). Whale Optimisation Algorithm (WOA) (5). Ant Colony Optimisation (ACO).

suggested that CSO provides more robust results than PSO and ABC. Bhargava et al.⁵⁵ showed that CSO offers a reliable method for solving thermodynamic calculations for complex phase equilibrium applications.

The ternary to decenary multicomponent compositions generated by the various algorithms, including GA, PSO, WOA, ACO and CSO, are presented in Tables 2, 3, 4, 5 and 6, respectively. These compositions were subsequently validated using the property model calculator in Thermo-Calc (TC) software. By comparing the hardness values of the newly generated compositions to those predicted by TC, it is evident that the CSO algorithm produced the most reliable multicomponent compositions, with a prediction error of less than $\pm 20\%$. Furthermore, compositions generated by CSO exhibited superior hardness compared to those derived from other optimization methods.

For instance, CSO successfully generated compositions such as $\text{Cu}_4\text{Fe}_4\text{Ni}_4\text{Zn}_4\text{Zr}_4\text{Co}_{10}\text{Cr}_{10}\text{V}_{20}\text{Ti}_{20}\text{Mo}_{20}$ (10 elements), $\text{Zn}_3\text{Cr}_{10}\text{Zr}_{10}\text{Mo}_{25}\text{Ti}_{25}\text{V}_{25}$ (6 elements), $\text{Zr}_{6.25}\text{Zn}_{18.75}\text{Ti}_{25}\text{V}_{25}\text{Mo}_{25}$ (5 elements), $\text{Ti}_{0.011111}\text{NiFe}_{0.4}\text{Cu}_{0.4}$ (4 elements), which achieved hardness values of 477 HV, 443 HV, 434 HV and 488 HV respectively. In contrast, while GA, PSO, and WOA demonstrated consistency in their predictions for several compositions, they often failed to accurately predict compositions with higher hardness values. Notably, these algorithms became increasingly erroneous with the inclusion of more elements in the alloy design.

No. of elements	Composition suggested from GA	Corresponding HV	Thermo-calc HV	Percentage error (%)
3	Mn _{23.41682} Sn _{23.41682} Ti _{53.16636}	387	421	8%
3	Cu _{9.10794325} Mo _{9.10794325} Ti _{81.7841135}	392	364	8%
4	Cu _{6.11253929} Ni _{6.11253929} Cr _{32.1722748} Ti _{55.60264661}	399	425	6%
4	Cr _{11.111111} Co _{22.222222} Ti _{22.222222} Mo _{22.222222} V _{22.222222}	377	423	11%
5	Ni _{5.79651873} Zn _{5.79651873} Mo _{5.79651873} Mn _{30.09710402} Ti _{53.2668911}	388	391	0.8%
5	Mo ₂₀ Mn ₂₀ Ti ₂₀ Sn ₂₀ V ₂₀	399	350	14%
6	Ti _{6.666667} Cr _{16.666667} Co _{16.666667} Mo ₂₀ V ₂₀ Zr ₂₀	381	414	8%
6	Cr ₁₀ Zr ₁₀ Mo ₂₀ Ti ₂₀ Sn ₂₀ V ₂₀	368	397	7%
7	Cu ₅ Zr ₅ Co ₁₅ Cr ₁₅ Ti ₂₀ Mo ₂₀ V ₂₀	382	393	3%
7	Zn ₅ Zr ₅ Co ₁₀ Ti ₂₀ Cr ₂₀ Mo ₂₀ V ₂₀	324	414	22%
8	Ni ₅ Zr ₅ Cr ₁₀ Co ₁₀ Zn ₁₀ Ti ₂₀ V ₂₀ Mo ₂₀	310	413	25%
8	Co ₁₀ Sn ₁₀ V ₁₀ Zn ₁₀ Zr ₁₀ Cr _{12.5} Ti _{17.5} Mo ₂₀	398	370	8%
9	Cu ₅ Ni ₅ Zn ₅ Zr ₅ Co ₁₂ Cr ₁₂ Mo ₁₃ V ₁₃ Ti ₃₀	367	398	8%
9	Cu ₅ Co ₅ Ni ₅ Zr ₅ Cr ₈ Mo ₁₀ Ti ₁₀ V ₁₀ Sn ₄₂	391	308	27%
10	Mn ₂ Sn ₂ Zr ₄ Cu ₆ Zn ₆ Cr ₁₀ Co ₁₀ Ti ₂₀ V ₂₀ Mo ₂₀	319	409	22%
10	Cu ₃ Ni ₃ Fe ₃ Zn ₃ Zr ₃ Co ₅ Cr ₅ Mo ₂₅ Ti ₂₅ V ₂₅	350	444	21%

Table 2. Compositions generated using GA optimisation.

No. of elements	Composition suggested from PSO	HV predicted by PSO	Thermo-calc HV	Percentage error (%)
4	Mo _{5.555556} Zn _{5.555556} Mn _{44.44444} V _{44.44444}	306	339	10
4	Fe _{6.25} Ni _{6.25} Cr _{43.75} Zn _{43.75}	303	397	24
5	Ni _{5.88235} Zr _{5.88235} Mn _{5.88235} Cr _{41.17647} Ti _{41.17647}	330	394	16
5	Cr _{4.76190} Ni _{4.76190} Co _{4.76190} Mn _{42.85714} Ti _{42.85714}	335	372	10
6	Cu _{4.54545455} Mn _{4.54545455} Sn _{4.54545455} Fe _{4.54545455} Cr _{40.90909091} Ti _{40.90909091}	349	308	13
6	V _{4.166667} Ni _{4.166667} Fe _{4.166667} Mo _{4.166667} Cr _{41.666667} Ti _{41.666667}	326	395	17
7	Cu ₈ Zr ₈ Co ₁₂ Cr ₁₂ Ti ₂₀ Mo ₂₀ V ₂₀	304	394	23
7	Co ₄ Cu ₄ Fe ₄ Mn ₄ Ni ₄ Cr ₄₀ Ti ₄₀	340	310	10
7	Co ₅ V ₅ Zn ₅ Zr ₅ Cr ₁₀ Mo ₂₀ Ti ₅₀	311	364	14
8	Ni _{6.666667} Zn _{6.666667} Zr _{6.666667} Cr ₁₀ Co ₁₀ Mo ₂₀ Ti ₂₀ V ₂₀	305	415	26
8	Cu ₃ Zn ₃ Zr ₃ Co ₁₂ Cr ₁₂ Mo ₁₃ V ₁₃ Ti ₃₅	298	378	21
9	Cu ₅ Ni ₅ Co ₅ Zn ₅ Ti ₁₀ V ₁₀ Cr ₁₀ Mo ₁₀ Sn ₄₀	316	325	3
9	Cu ₅ Ni ₅ Zn ₅ Zr ₅ Co ₁₀ Cr ₁₀ Mo ₁₅ V ₁₅ Ti ₃₀	290	411	29
10	Al _{4.16666667} Cu _{4.16666667} Cr _{4.16666667} Fe _{4.16666667} Mn _{4.16666667} Ti _{4.16666667} Zn _{4.16666667} Zr _{4.16666667} Mo _{33.33333333} V _{33.33333333}	315	424	26
10	Co ₁₀ Cr ₁₀ Cu ₁₀ Fe ₁₀ Ni ₁₀ Mo ₁₀ Ti ₁₀ V ₁₀ Zr ₁₀ Zn ₁₀	343	296	16

Table 3. Compositions generated using PSO optimisation.

ACO method proved to be reliable for ternary and quaternary compositions, yielding relatively low prediction errors. However, as the complexity escalated from quinary to decenary compositions, the percentage error substantially increased, and ACO was unable to generate viable compositions with more than seven elements.

As a result, CSO emerged as the most reliable algorithm among those tested, capable of generating multicomponent compositions with reduced CRM content and enhanced hardness values. Figure 7 illustrates the hardness values generated by each algorithm in comparison to those evaluated by Thermo-Calc for ternary to decenary compositions, providing a comprehensive overview of algorithm performance across varying levels of compositional complexity.

We further compared the hardness values of the newly generated R-CRM-MPEAs with those of CRM-containing MPEAs reported in the literature to assess the feasibility of partially substituting CRMs. Figure 8 provides a comparison of the mechanical properties between experimentally synthesised CRM-laden alloys reported in the literature and the newly predicted R-CRM-MPEAs in this work.

The results indicate that the proposed approach can yield comparable hardness. For example, CoCrFeNb_{0.309}Ni containing two 1st-Tier CRMs exhibits a hardness of 480 HV, whereas the newly predicted composition, Ti_{0.0111}NiFe_{0.4}Cu_{0.4}, achieves a superior hardness of 488 HV without including any 1st-Tier CRMs. The computational approach employed using ML in this work has great potential to design R-CRM-MPEA compositions. This approach can be leveraged to eliminate the use of CRMs in diverse applications such as catalysis, semiconductors, transportation and other carbon-intensive sectors.

No. of elements	Composition suggested from WOA	HV predicted by WOA	Thermo-Calc HV	Percentage error (%)
3	Cu _{11.2} Mn _{36.8} Ti ₅₂	391	419	7
4	Mo _{5.92105} Al _{13.81579} Co _{17.10526} Ti _{63.15789}	363	192	89
4	Fe ₁₀ Zn ₂₀ Cr ₃₀ Ti ₄₀	355	325	9
5	Ni _{3.41880} Mn _{4.27350} Zn _{10.25641} V _{29.05983} Mo _{52.99145}	346	375	8
5	Cr _{9.09091} Mn _{9.09091} Zr _{9.09091} Mo _{36.36364} V _{36.36364}	324	400	19
6	Fe _{5.55556} Mn _{5.55556} V _{5.55556} Zn _{5.55556} Ti _{38.88889} Cr _{38.88889}	333	382	13
6	Cr _{7.14286} Mn _{7.14286} Ti _{7.14286} Zr _{7.14286} Mo _{35.71429} V _{35.71429}	322	445	28
7	Co _{4.34783} Cr _{4.34783} Ni _{4.34783} Ti _{4.34783} Zn _{4.34783} Mo _{39.13043} V _{39.13043}	337	414	18
7	Co ₅ Cr ₁₀ V ₁₀ Zn ₁₀ Zr ₁₀ Mo ₂₀ Ti ₃₅	323	409	21
8	Al ₅ Co ₅ Cu ₅ Fe ₅ Mn ₅ Ni ₅ Mo ₃₅ V ₃₅	319	390	18
8	Al _{12.5} Co _{12.5} Cr _{12.5} Mo _{12.5} Ti _{12.5} V _{12.5} Zn _{12.5} Zr _{12.5}	312	379	18
9	Fe _{4.34783} Ni _{4.34783} Mo _{4.34783} Sn _{4.34783} Ti _{4.34783} Zn _{4.34783} Zr _{4.34783} Cu _{34.78261} V _{34.78261}	281	296	5
9	Cu ₅ Ni ₅ Zn ₅ Zr ₅ Co ₆ Mo ₉ V ₉ Ti ₅₀	326	413	21
10	Al _{3.57143} Co _{3.57143} Cr _{3.57143} Cu _{3.57143} Fe _{3.57143} Ni _{3.57143} Ti _{3.57143} Zr _{3.57143} Mo _{35.71429} V _{35.71429}	314	420	25
10	Al ₅ Co ₅ Cr ₅ Fe ₅ Mn ₅ Ti ₅ Zr ₅ Zn ₅ Mo ₃₀ V ₃₀	315	365	14

Table 4. Compositions generated using WOA optimisation.

No. of elements	Composition suggested from ACO	HV predicted by ACO	Thermo-calc HV	Percentage error (%)
3	Co ₁₃ Al ₂₀ Cu ₆₇	168	196	14
4	Fe ₁₀ Zn ₂₀ Cr ₃₀ Ti ₄₀	278	325	14
5	Fe ₄ Co ₁₅ V ₂₄ Sn ₂₈ Cr ₂₉	273	352	22
6	Sn ₄ Mn ₇ Ti ₁₃ Ni ₂₁ Al ₂₆ Zr ₂₉	150	235	36
6	Cr ₈ Cu ₈ Mn ₉ Sn ₁₆ Al ₁₈ Zn ₄₁	167	272	39
6	Cu ₆ Mo ₆ Ni ₆ Zr ₂₀ Sn ₃₀ Cr ₃₂	258	332	22
7	Fe ₃ Mo ₄ Mn ₈ Cu ₉ Zn ₁₉ Cr ₂₅ Co ₃₂	269	348	23
7	Ti ₂ V ₆ Fe ₁₄ Mn ₁₅ Cr ₁₇ Zr ₁₈ Mo ₂₈	244	376	35

Table 5. Compositions generated using ACO optimisation.

No. of elements	Composition suggested from CSO	HV predicted by CSO	Thermo-calc HV	Percentage Error (%)
4	Mo ₂₅ Sn ₂₅ Ti ₂₅ V ₂₅	397	383	4
4	Cr ₂₅ Mo ₂₅ Ti ₂₅ V ₂₅	401	424	5
4	Ti _{0.01111} NiFe _{0.4} Cu _{0.4}	468	488	4
5	Co _{12.5} Cr _{12.5} Mo ₂₅ Ti ₂₅ V ₂₅	398	416	4
5	Mo ₁₀ Zn ₁₀ Al ₂₀ Cr ₃₀ Ti ₃₀	400	398	0.5
5	Co _{11.11111} Cr _{11.11111} Mo _{22.22222} V _{22.22222} Ti _{33.33333}	404	430	6
5	Zn _{6.25} Cr _{18.75} Mo ₂₅ Ti ₂₅ V ₂₅ (similar to novel one)	402	415	3
5	Zr _{6.25} Zn _{18.75} Ti ₂₅ V ₂₅ Mo ₂₅ (similar to novel one)	404	434	7
6	Zn ₅ Cr ₁₀ Zr ₁₀ Mo ₂₅ Ti ₂₅ V ₂₅	406	443	8
6	Co _{16.66667} Cr _{16.66667} Mo _{16.66667} Ti _{16.66667} V _{16.66667} Zr _{16.66667}	395	427	7
7	CoCrMoVZnZrTi	356	422	16
7	Cu ₅ Zr ₅ Co ₁₂ Cr ₁₂ Mo ₁₃ V ₁₃ Ti ₄₀	382	411	7
8	Co ₁₀ Mn ₁₀ V ₁₀ Zn ₁₀ Zr ₁₀ Cr _{12.5} Ti _{17.5} Mo ₂₀	396	416	5
8	Co _{12.5} Cu _{12.5} Cr _{12.5} Mo _{12.5} Ti _{12.5} V _{12.5} Zn _{12.5} Zr _{12.5}	385	427	10
8	Ni ₂ Zn ₂ Zr ₁₁ Cr _{12.5} Co _{12.5} Mo ₂₀ Ti ₂₀ V ₂₀	398	429	7
9	Cu ₅ Ni ₅ Zn ₅ Zr ₅ Co ₈ Cr ₈ Mo ₁₀ V ₁₀ Ti ₄₄	403	432	7
10	Co _{3.57143} Cr _{3.57143} Cu _{3.57143} Fe _{3.57143} Ni _{3.57143} Zn _{3.57143} Zr _{3.57143} Mo ₂₅ Ti ₂₅ V ₂₅ OR (CoCrCuFeNiZnZr) _{0.1428572} MoTiV	404	448	10
10	Cu ₄ Fe ₄ Ni ₄ Zn ₄ Zr ₄ Co ₁₀ Cr ₁₀ Ti ₂₀ V ₂₀ Mo ₂₀	405	477	15

Table 6. Compositions generated using CSO optimisation. Significant values are in bold.

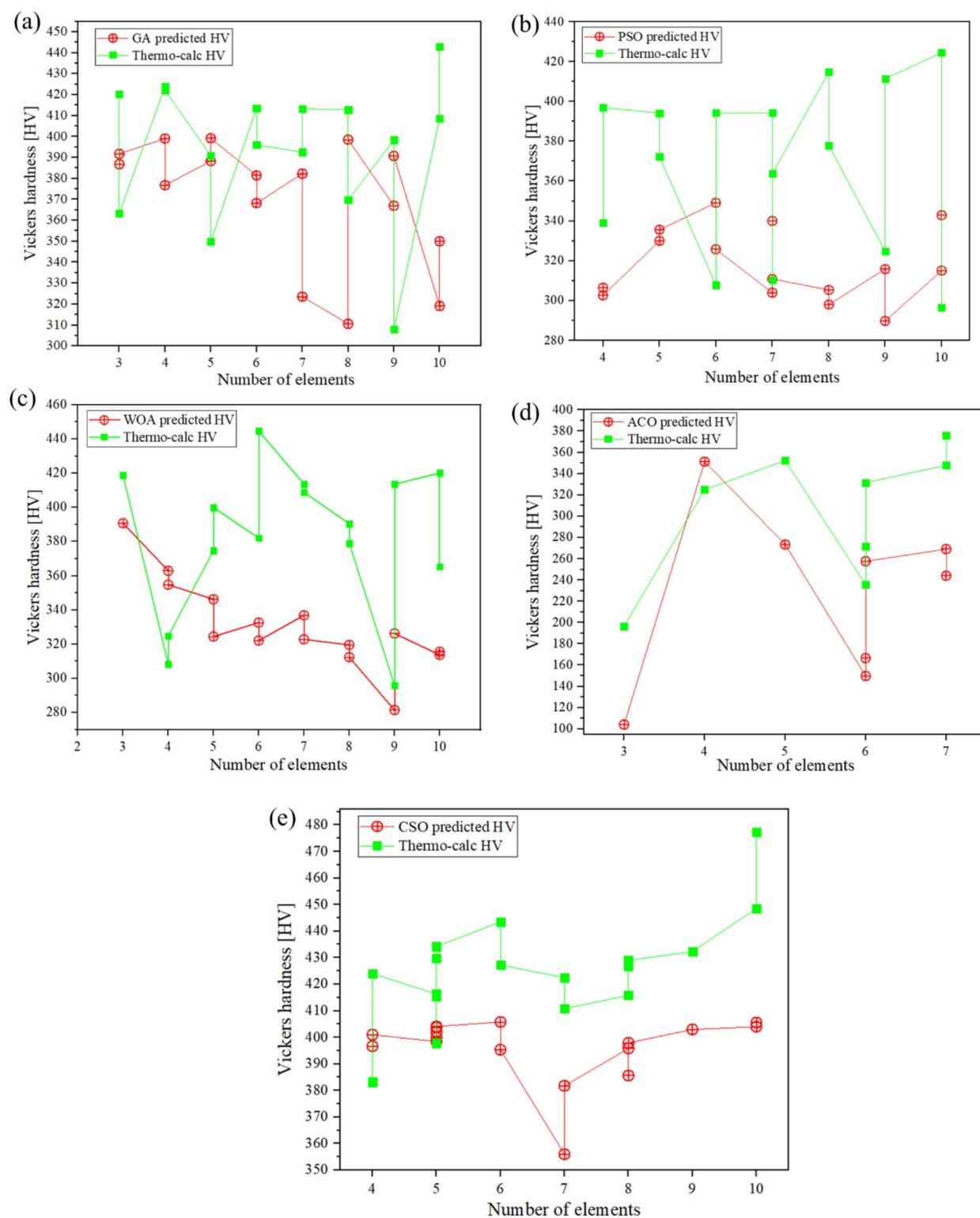


Fig. 7. Comparison of hardness value for newly generated MPEA compositions compared for ML and Thermo-Calc predictions using various techniques: (a) Genetic algorithm (GA), (b) Particle swarm optimisation (PSO), (c) Whale optimization (WOA), (d) Ant colony optimization (ACO), (e) Cuckoo search optimization (CSO).

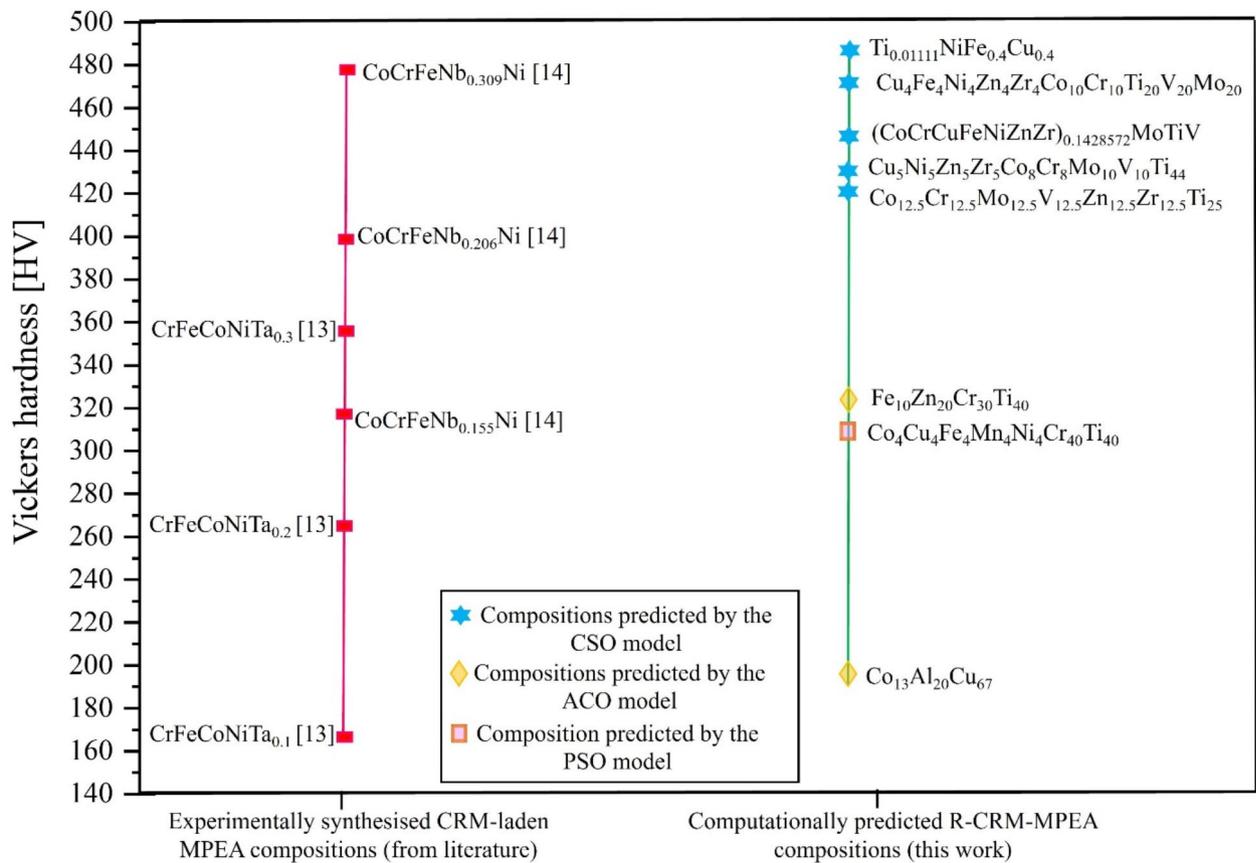


Fig. 8. Comparison of R-CRM-MPEA with CRM-laden MPEAs to demonstrate the feasibility of partially substituting CRMs with readily available elements while attaining comparable mechanical properties.

Alloy	Experimentally evaluated [HV _{0.1}]	Predicted [TC and ML]		Percentage error (%)
		TC-predicted	ML-predicted	
Al _{6.25} Cu _{18.75} Fe ₂₅ Co ₂₅ Ni ₂₅	Mean	177 HV	179 HV	18
	150 ± 10 HV			19

Table 7. Comparison of experimentally measured hardness with TC and ML predicted hardness.

Benchmarking of the newly synthesised alloy Al_{6.25}Cu_{18.75}Fe₂₅Co₂₅Ni₂₅

To test our predictions, a random and a new FCC phase alloy, Al_{6.25}Cu_{18.75}Fe₂₅Co₂₅Ni₂₅^{27–29} was selected for experimental validation. Vickers hardness testing was performed utilising a Wilson hardness testing apparatus, with a load of 0.1 kgf. The Vickers hardness value was determined by averaging the measurements from nine indents on a polished surface. The experimentally measured Vickers hardness of Al_{6.25}Cu_{18.75}Fe₂₅Co₂₅Ni₂₅ was subsequently compared with the hardness values evaluated by Thermo-Calc (TC) and our ML model (refer to Table 7). It was observed that both the TC-evaluated and ML-predicted hardness values were in strong agreement with the experimentally measured hardness with an error of less than 20%.

While numerous studies in the literature on MPEAs have used machine learning, a majority of these remain concentrated on developing algorithms for phase classification^{56–58} or predicting mechanical properties, such as hardness, yield strength, or elastic modulus^{59–63}. Relatively few investigations have aimed to generate or optimize novel MPEA compositions. Most of these efforts have concentrated on achieving high hardness values which differs from the focus of this investigation where the main objective was to develop substitute alloys to eliminate the use of CRMs and hardness was used as an indicator to demonstrate that comparable properties can still be obtained.

Ren et al.⁶⁴ used a dataset of 205 HEA samples featuring 19 characteristics commonly employed in HEA property prediction. They implemented a tree-based machine learning model to predict hardness and integrated it with Particle Swarm Optimization (PSO) for component optimization. Due to the limited availability of real experimental data, they resorted to synthetic data through random oversampling to improve the performance of their Component Optimization Model (COM), which raises concerns about the reliability of their prediction. Their database primarily included Al, Co, Cr, Cu, Fe and Ni.

Similarly, Wen et al.⁶⁵ focussed on the same Al-Co-Cr-Cu-Fe-Ni HEA family and its subgroups but worked with a limited dataset of only 155 samples. They proposed a property-oriented materials design strategy that combined machine learning with the Design of Experiments (DOE) to discover alloys with high hardness within this HEA system. Their resulting alloy exhibited a hardness approximately 10% higher than the best value found in the original training set.

In contrast, Yang et al.⁶⁶ used a dataset of 370 HEAs, including compositions such as Al-Co-Cr-Cu-Fe-Ni, Al-Co-Cr-Fe-Mn-Ni, and their derivatives, along with vanadium-containing alloys, which are recognised for their high hardness. They applied techniques such as Inverse Projection (IP) and High-Throughput Screening (HTS) and encountered substantial prediction errors of up to 58% due to the risks associated with extrapolation beyond the boundaries of the training dataset and insufficient data diversity.

To effectively navigate the design space for alloys with high hardness, it is essential to enrich the training datasets with refractory alloys, given their inherent high-temperature stability and robust mechanical properties. However, a major impediment to the reliable application of ML in materials science continues to be the scarcity of relevant data, particularly for HEAs/MPEAs. Roy et al.⁶⁷ addressed this issue by employing a generative adversarial network (GAN) to explore an 18-dimensional design space involving Co-Fe-Ni-Si-Al-Cr-Mo-Ti-Nb-V-Zr-Mn-Cu-Sn-Ta-Hf-W-Zn MPEAs using a limited dataset of 241 alloys. They successfully designed two new MPEAs with hardness values exceeding 941 HV.

In contrast to these studies, our research emphasizes sustainable materials design by assessing whether compositions with reduced or no CRMs can achieve competitive hardness values. This objective addresses a critical gap in the current alloy design landscape—the need for environmentally sustainable materials that minimize reliance on CRMs while maintaining desirable mechanical properties. This approach not only advances sustainability goals but also enhances supply chain resilience, representing a significant step forward in alloy innovation.

Conclusion

This study represents a significant effort to support Net-Zero initiatives by developing new compositions with reduced critical raw materials (CRMs). The research relies on a computational framework that involves sourcing a dataset of Vickers hardness values for unary (pure) and binary material compositions from Thermo-Calc 2024a and the TCHEA7 database. This dataset was used to build machine learning models to identify complex compositions of alloys with reduced-CRM without negating the mechanical properties.

Among all regression models, the Extra Trees Regressor (ETR) demonstrated superior performance, achieving an R^2 score of 0.82 and a MAPE of 0.17 for the test data. Various metaheuristic optimization techniques were subsequently employed to inversely predict novel multicomponent alloy compositions free of critical raw materials (CRMs) but with hardness comparable to CRM-containing multi-principal element alloys (MPEAs). Of all the optimization models, Cuckoo Search Optimization (CSO) demonstrated a high level of concordance with Thermo-Calc predictions, with an average deviation of $\pm 20\%$. A literature-sourced CRM-laden composition, $\text{CoCrFeNb}_{0.309}\text{Ni}$, containing two 1st-Tier CRMs (Co and Nb), showed a hardness value of 480 HV and a new alloy was generated using a machine learning method namely, $\text{Ti}_{0.01111}\text{NiFe}_{0.4}\text{Cu}_{0.4}$, with a hardness of 488 HV that showed great opportunity to eliminate CRMs in developing MPEAs.

The validity of this study was reinforced by comparing computational predictions—derived from our machine learning methodology and Thermo-Calc evaluations—with the experimentally measured Vickers hardness of a test alloy $\text{Al}_{6.25}\text{Cu}_{18.75}\text{Fe}_{25}\text{Co}_{25}\text{Ni}_{25}$, which contains a single 1st-Tier CRM, cobalt (Co). Therefore, this investigation offers valuable insights into the potential for designing novel MPEAs with reduced or even no CRMs, significantly contributing to sustainable materials innovation to support Net Zero in the metal sector. Future research will focus on further experimental validation to corroborate the findings for the newly generated compositions.

Data availability

The database can be accessed at <https://gitfront.io/r/user-6296136/13cNmHofQt3/Thermo-calc-database>. Additional details about the capabilities of Thermo-Calc software can be found in the following resources: (1) Thermo-Calc Product Overview - Property Model Calculator, (2) Brochure: Properties that Thermo-Calc can Calculate.

Received: 26 June 2024; Accepted: 22 January 2025

Published online: 24 January 2025

References

1. Cantor, B. Multicomponent and high entropy alloys. *Entropy* **16**(9), 4749–4768 (2014).
2. Cantor, B. Multicomponent high-entropy Cantor alloys. *Prog. Mater. Sci.* **120**, 100754 (2021).
3. Cantor, B. et al. Microstructural development in equiatomic multicomponent alloys. *Mater. Sci. Eng., A* **375**, 213–218 (2004).
4. Yeh, J. W. et al. Nanostructured high-entropy alloys with multiple principal elements: Novel alloy design concepts and outcomes. *Adv. Eng. Mater.* **6**(5), 299–303 (2004).
5. Han, Z. et al. Microstructures and mechanical properties of $\text{Ti}_x\text{NbMoTaW}$ refractory high-entropy alloys. *Mater. Sci. Eng., A* **712**, 380–385 (2018).
6. Wang, M. et al. Designing VxNbMoTa refractory high-entropy alloys with improved properties for high-temperature applications. *Scripta Mater.* **191**, 131–136 (2021).
7. Pang, J. et al. A ductile $\text{Nb}_{10}\text{Ti}_{25}\text{Al}_{15}\text{V}_{10}\text{Ta}_5\text{Hf}_5\text{W}_5$ refractory high entropy alloy with high specific strength for high-temperature applications. *Mater. Sci. Eng., A* **831**, 142290 (2022).
8. Wan, Y., et al., A nitride-reinforced NbMoTaW refractory high-entropy alloy with potential ultra-high-temperature engineering applications. *Engineering*, (2023).

9. Dixit, S. et al. Refractory high-entropy alloy coatings for high-temperature aerospace and energy applications. *J. Therm. Spray Technol.* **31**(4), 1021–1031 (2022).
10. He, H. et al. Carbide-reinforced $\text{Re}_{0.1}\text{Hf}_{0.25}\text{NbTaW}_{0.4}$ refractory high-entropy alloy with excellent room and elevated temperature mechanical properties. *Int. J. Refractory Metals Hard Mater.* **116**, 106349 (2023).
11. Tsao, T. K. et al. On the superior high temperature hardness of precipitation strengthened high entropy Ni-based alloys. *Adv. Eng. Mater.* **19**(1), 1600475 (2017).
12. Tsao, T.-K. et al. The high temperature tensile and creep behaviors of high entropy superalloy. *Sci. Rep.* **7**(1), 12658 (2017).
13. Jiang, H. et al. Effects of Ta addition on the microstructures and mechanical properties of CoCrFeNi high entropy alloy. *Mater. Chem. Phys.* **210**, 43–48 (2018).
14. Liu, W. et al. Effects of Nb additions on the microstructure and mechanical property of CoCrFeNi high-entropy alloys. *Intermetallics* **60**, 1–8 (2015).
15. He, F. et al. Designing eutectic high entropy alloys of CoCrFeNiNb. *J. Alloys Compd.* **656**, 284–289 (2016).
16. Rizzo, A. et al. The critical raw materials in cutting tools for machining applications: A review. *Materials* **13**(6), 1377 (2020).
17. Dobbstein, H. et al. Laser metal deposition of refractory high-entropy alloys for high-throughput synthesis and structure-property characterization. *Int. J. Extreme Manufact.* **3**(1), 015201 (2020).
18. Commission, E. *Tackling the challenges in commodity markets and on raw materials* (European Commission, 2011).
19. Commission, E., On the review of the list of critical raw materials for the eu and the implementation of the raw materials initiative. In *Ad-hoc working group on defining critical raw minerals of the raw materials supply group*, (2014).
20. Commission, E., *List of critical raw materials for the EU*. Online verfügbar unter, (2017).
21. Commission, E., Critical Raw materials resilience: charting a path towards greater security and sustainability. In *Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions*, (2020).
22. Grohol, M. & Veeh, C. *Study on the Critical Raw Materials for the EU 2023*. 2023: Publications Office of the European Union.
23. Petranikova, M. et al. Vanadium sustainability in the context of innovative recycling and sourcing development. *Waste Manag.* **113**, 521–544 (2020).
24. Suriyanarayanan, S. et al. Highly efficient recovery and recycling of cobalt from spent lithium-ion batteries using an N-Methylurea–Acetamide nonionic deep eutectic solvent. *ACS Omega* **8**(7), 6959–6967 (2023).
25. Zhang, J. & Azimi, G. Recycling of lithium, cobalt, nickel, and manganese from end-of-life lithium-ion battery of an electric vehicle using supercritical carbon dioxide. *Resour. Conserv. Recycling* **187**, 106628 (2022).
26. Yang, L., et al., *Recycling potential of cobalt metal from end-of-life new energy passenger vehicles in China*. Waste Management & Research, 2024; p. 0734242X231219650.
27. Singh, S. et al. Phase prediction and experimental realisation of a new high entropy alloy using machine learning. *Sci. Rep.* **13**(1), 4811 (2023).
28. Fan, P. et al. Anisotropic plasticity mechanisms in a newly synthesised high entropy alloy investigated using atomic simulations and nanoindentation experiments. *J. Alloys Compd.* **970**, 172541 (2024).
29. Fan, P., et al., Uniaxial pulling and nano-scratching of a newly synthesized high entropy alloy. *APL Materials*, **10**(11), (2022).
30. Qin, G. et al. Microstructures and mechanical properties of Nb-alloyed CoCrCuFeNi high-entropy alloys. *J. Mater. Sci. Technol.* **34**(2), 365–369 (2018).
31. Zhang, K. et al. Microstructure and mechanical properties of CoCrFeNiTiAlx high-entropy alloys. *Mater. Sci. Eng.: A* **508**(1–2), 214–219 (2009).
32. Zhu, Z. et al. Compositional dependence of phase formation and mechanical properties in three CoCrFeNi-(Mn/Al/Cu) high entropy alloys. *Intermetallics* **79**, 1–11 (2016).
33. He, J. et al. Effects of Al addition on structural evolution and tensile properties of the FeCoNiCrMn high-entropy alloy system. *Acta Materialia* **62**, 105–113 (2014).
34. Wu, Z. et al. Temperature dependence of the mechanical properties of equiatomic solid solution alloys with face-centered cubic crystal structures. *Acta Materialia* **81**, 428–441 (2014).
35. Xu, D. et al. A critical review of the mechanical properties of CoCrNi-based medium-entropy alloys. *Microstructures* **2**(1), 2022001 (2022).
36. Agarwal, R. et al. Understanding the deformation behavior of CoCuFeMnNi high entropy alloy by investigating mechanical properties of binary ternary and quaternary alloy subsets. *Mater. Design* **157**, 539–550 (2018).
37. Wang, J. et al. A neural network model for high entropy alloy design. *Npj Comput. Mater.* **9**(1), 60 (2023).
38. Jaiswal, U. K. et al. Machine learning-enabled identification of new medium to high entropy alloys with solid solution phases. *Comput. Mater. Sci.* **197**, 110623 (2021).
39. Zhang, J., et al., *Composition design of high-entropy alloys with deep sets learning*. npj Computational Materials, 2022. **8**(1): p. 89.
40. Liu, X., Zhang, J. & Pei, Z. Machine learning for high-entropy alloys: Progress, challenges and opportunities. *Progress Mater. Sci.*, 101018, (2022).
41. Rivera-Diaz-del-Castillo, P. E. & Fu, H. Strengthening mechanisms in high-entropy alloys: Perspectives for alloy design. *J. Mater. Res.* **33**(19), 2970–2982 (2018).
42. Mahmoud, E. R. et al. Phase prediction, microstructure and mechanical properties of Fe–Mn–Ni–Cr–Al–Si high entropy alloys. *Metals* **12**(7), 1164 (2022).
43. Lukas, H., Fries, S. G. & Sundman, B. *Computational thermodynamics*. 104, (2007).
44. Walbrühl, M. et al. Modelling of solid solution strengthening in multicomponent alloys. *Mater. Sci. Eng.: A* **700**, 301–311 (2017).
45. Gulley, A. L. China, the democratic republic of the Congo, and artisanal cobalt mining from 2000 through 2020. *Proc. Natl. Acad. Sci.* **120**(26), e2212037120 (2023).
46. Ahmed, Z. E. et al. Optimizing energy consumption for cloud internet of things. *Front. Phys.* **8**, 358 (2020).
47. Janga Reddy, M. & Nagesh Kumar, D. Evolutionary algorithms, swarm intelligence methods, and their applications in water resources engineering: A state-of-the-art review. *h2oj* **3**(1), 135–188 (2020).
48. Kumar, A. & Bawa, S. A comparative review of meta-heuristic approaches to optimize the SLA violation costs for dynamic execution of cloud services. *Soft Comput.* **24**(6), 3909–3922 (2020).
49. Dhiman, G. & Kaur, A. Optimizing the design of airfoil and optical buffer problems using spotted hyena optimizer. *Designs* **2**(3), 28 (2018).
50. Rao, S. S. *Engineering optimization: Theory and practice* (John Wiley & Sons, 2019).
51. Gandomi, A. H., Yang, X.-S. & Alavi, A. H. Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems. *Eng. Comput.* **29**, 17–35 (2013).
52. Gandomi, A. H. et al. Design optimization of truss structures using cuckoo search algorithm. *Struct. Design Tall Special Build.* **22**(17), 1330–1349 (2013).
53. Katiyar, S. A comparative study of genetic algorithm and the particle swarm optimization. *Int. J. Technol.* **2**(2), 21–24 (2010).
54. Yang, X.-S. & Deb, S. Cuckoo search: Recent advances and applications. *Neural Comput. Appl.* **24**, 169–174 (2014).
55. Bhargava, V., Fateen, S.-E.K. & Bonilla-Petriciolet, A. Cuckoo search: A new nature-inspired optimization method for phase equilibrium calculations. *Fluid Phase Equilibria* **337**, 191–200 (2013).
56. Zhang, Y. et al. Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models. *Acta Materialia* **185**, 528–539 (2020).

57. Huang, W., Martin, P. & Zhuang, H. L. Machine-learning phase prediction of high-entropy alloys. *Acta Materialia* **169**, 225–236 (2019).
58. Machaka, R., et al., Machine learning-based prediction of phases in high-entropy alloys: A data article. *Data in Brief*, **38**, (2021).
59. Bhandari, U. et al. Yield strength prediction of high-entropy alloys using machine learning. *Mater. Today Commun.* **26**, 101871 (2021).
60. Bhandari, U. et al. Deep learning-based hardness prediction of novel refractory high-entropy alloys with experimental validation. *Crystals* **11**(1), 46 (2021).
61. Roy, A. et al. Machine learned feature identification for predicting phase and Young's modulus of low-, medium-and high-entropy alloys. *Scripta Materialia* **185**, 152–158 (2020).
62. He, J., et al., Machine learning-assisted design of high-entropy alloys with superior mechanical properties. *J. Mater. Res. Technol.*, (2024).
63. Liu, S., Lee, K. & Balachandran, P. V. Integrating machine learning with mechanistic models for predicting the yield strength of high entropy alloys. *J. Appl. Phys.*, **132**(10), (2022)
64. Ren, W. et al. Prediction and design of high hardness high entropy alloy through machine learning. *Mater. Design* **235**, 112454 (2023).
65. Wen, C. et al. Machine learning assisted design of high entropy alloys with desired property. *Acta Materialia* **170**, 109–117 (2019).
66. Yang, C. et al. A machine learning-based alloy design system to facilitate the rational design of high entropy alloys with enhanced hardness. *Acta Materialia* **222**, 117431 (2022).
67. Roy, A. et al. Rapid discovery of high hardness multi-principal-element alloys using a generative adversarial network model. *Acta Materialia* **257**, 119177 (2023).

Acknowledgements

SNJ would like to thank the Department of Science and Technology (DST) India. SG would like to acknowledge the financial support from the UKRI via Grant No. EP/T024607/1 and the International exchange Cost Share award by the Royal Society (IEC\NSFC\223536). We also acknowledge Royal Society Research Grant (RGS\R2\222304) for funding the purchase of Thermo-Calc software and database (TCHEA6 and MOBHEA2 for High Entropy Alloys).

Author contributions

SS: Performing research and Creating the Draft MB: Providing the software AM: Providing access to the resources SG: Supervision and editing SNJ: Supervision and editing.

Declarations

Competing interests

The authors declare no competing interests.

Content of the publication

During the preparation of this work, the author(s) used AI assistance to improve the readability and language of the manuscript. The authors reviewed and edited the final content and took full responsibility for the content of the published article.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-87784-0>.

Correspondence and requests for materials should be addressed to S.G. or S.N.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025