Design and Selection of High Entropy Alloys for Hardmetal Matrix Applications Using a Coupled Machine Learning and Calculation of Phase Diagrams Methodology

Joshua Berry,* Robert Snell, Magnus Anderson, Lewis R. Owen, Olivier M. D. M. Messé, Iain Todd, and Katerina A. Christofidou*

This study aims to utilize a combined machine learning (ML) and CALculation of PHAse Diagrams (CALPHAD) methodology to design hardmetal matrix phases for metal-forming applications that can serve as the basis for carbide reinforcement. The vast compositional space that high entropy alloys (HEAs) occupy offers a promising avenue to satisfy the application design criteria of wear resistance and ductility. To efficiently explore this space, random forest ML models are constructed and trained from publicly available experimental HEA databases to make phase constitution and hardness predictions. Interrogation of the ML models constructed reveals accuracies >78.7% and a mean absolute error of 66.1 HV for phase and hardness predictions respectively. Six promising alloy compositions, extracted from the ML predictions and CALPHAD calculations, are experimentally fabricated and tested. The hardness predictions are found to be systematically under- and overpredicted depending on the alloy microstructure. In parallel, the phase classification models are found to lack sensitivity toward additional intermetallic phase formation. Despite the discrepancies identified between ML and experimental results, the fabricated compositions show promise for further experimental evaluation. These discrepancies are believed to be directly associated with the available databases but, importantly, have highlighted several avenues for both ML and database development.

1. Introduction

Since their introduction in 2004,[1,2] high entropy alloys (HEAs) have come into prominence particularly due to the opportunities that they offer for alloy design and development, targeting previously unexplored areas of compositional space. Originally, HEAs were defined as a class of alloys containing five or more elements, in equiatomic concentrations.[1] However, this was subsequently expanded to encompass multiprincipal element alloys with elemental concentrations in the range of 5–35 at%,[3,4] thereby increasing the potential for diverse HEA design.[5] HEAs are often interchangeably referred to as multiprincipal element alloys, multicomponent alloys, complex concentrated alloys, and compositionally complex alloys.[6] For this article, the term HEA will be used generally to encompass all classes described above.

The popularity of HEAs in the material science community has been rapidly increasing over the past decade.[3,6–9] Scientific interest in HEAs originates from the number of possible compositions[2,10,11] with the potential to tailor the mechanical, structural, and functional properties, for example, their observed synergy of strength and ductility.[5,12,13] However, within studies of the HEA field exist several key biases. Prior to 2015, HEA studies focused on the investigation of single-phase solid solution microstructures with crystallographically simple phases and exploring microstructural evolution. This has resulted in an emphasis on alloy families containing a limited range of $3d$ transition metals.[6,9] Co, Cr, Fe, and Ni are the most common HEA constituents, appearing in 85% of all HEA compositions published before 2015.[8] Prioritization of single-phase solid solutions contrasts with conventional alloys, which typically utilize strengthening intermetallic or ceramic phases to obtain the optimal balance of strength and damage tolerance.[7] Subsequently, an increasing number of studies, particularly in the multiprincipal element and complex concentrated alloy space, have targeted more technologically relevant microstructures. (e.g., multiphase HEAs,[14,15] with precipitation hardened microstructures,[16,17] and eutectic compositions.[18,19] Nevertheless, around 70% of

J. Berry, R. Snell, L. R. Owen, I. Todd, K. A. Christofidou
Department of Materials Science and Engineering
University of Sheffield
Mappin Street, Sheffield S1 3JD, UK
E-mail: jberry1@sheffield.ac.uk; k.christofidou@sheffield.ac.uk

M. Anderson
Thermo-Calc Software AB
Råsundavägen 18, SE-169 67 Solna, Sweden

O. M. D. M. Messé
Oerlikon AM Europe GmbH
Kapellenstr. 12, 85622 Feldkirchen, Germany

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
ENGINEERING
MATERIALS**

www.aem-journal.com

HEA microstructure studies investigate as-cast alloys, which do not represent the equilibrium state and are not always indicative of industry applications. Annealing tends to result in microstructures consisting of multiple solid solution and intermetallic phases, even in cases where the as-cast microstructure was single phase.[9]

In addition to these challenges and limitations, the great opportunities offered by the emergence of HEAs are further complicated by the lack of robust computational tools able to rapidly explore the compositional space for areas of interest. This in turn offers an opportunity for the application of machine learning (ML) tools to accelerate and automate the compositional exploration tasks. Application of ML in material science is not a novel concept and has become accepted as a useful tool to help automate material discovery.[20,21] Accordingly, a natural relationship has emerged between ML and the HEA field. ML provides an opportunity to explore the vast HEA compositional space,[22] reviewing large amounts of data to discover patterns and trends, and make predictions on unexplored HEA compositions.[23] These predictions can be performed quickly, providing reproducible results, and aiding in alloy design and development, with the capability for future scaling.[24]

Several studies of the application of ML to HEAs have been published, mostly aiming at the design of new compositions for structural applications through the determination of alloy microstructures and properties.[23,25–27] Most studies utilize existing ML algorithms and architectures, including support vector machines, artificial neural networks, gradient boosting, k-nearest neighbors, and random forest (RF).[28–30] ML can be trained from available HEA data to make direct predictions on the phase formation and mechanical properties of HEA compositions.[20,23] New data and features can also be directly added to update and refit the model if they become available.[31] Despite the apparent "black box" nature of ML algorithms, multiple techniques such as permutation importance and SHapley Additive exPlanations (SHAP) can be utilized to explain both the global model methods and local individual predictions respectively, as well as assist in interpretability of feature importance.[24] Future experimental testing can then assess the alloys suitability for application and validate the ML predictions.

Despite the distinct advantages discussed, there are drawbacks to the application of ML in the HEA field. First, the success and development of ML within the HEA field is intrinsically linked to the experimental exploration of the compositional space.[20] Hence, the investigation of HEA systems sampling a wide range of the compositional space is needed for the generation and growth of robust experimental HEA databases. Diverse and expansive databases are required to effectively train the supervised ML algorithms, but HEA databases typically only contain anywhere from a few hundred, to a thousand experimental data points.[22] Hence, the amount of training data is typically insufficient for most algorithms and many HEA ML studies report that, the unavailability of HEA data for training is a major limitation of the models.[32] In addition, the data that are available are typically imbalanced, a direct consequence of the focus on a specific family of elements. Imbalanced data can negatively impact ML predictions by biasing models toward the majority class. Solutions such as oversampling to increase the amounts of the minority class or downsampling to reduce the amount of the majority class often result in additional drawbacks, such as reduction in the useful training dataset. The second major difficulty with the application of ML to HEAs, is the need to produce physically meaningful descriptors to best represent the alloys.[22] ML depends upon the mathematical representation of the materials, their features, which serve as the inputs to the algorithm.[20] Good features should describe the alloys such that the chosen algorithm selects the key information to make predictions from it.[33] Additionally, ML calculations are particularly vulnerable to overfitting. Overfitting occurs when an ML algorithm too closely matches the training dataset and may be unable to effectively make predictions on unseen data.[21]

A common theme that emerges from these studies to address the lack of phase data is the use and integration with CALculation of PHAse Diagrams (CALPHAD) based techniques and predictions.[22,31,34,35] CALPHAD offers a tool that allows the calculations of several microstructural indicators from thermodynamic principles as well as, in some limited cases, and with varying levels of success, the calculation of material properties such as yield strength. These data can be combined with experimental results to enhance the volume of data available for improved ML implementation. Alternatively, CALPHAD outputs are often used as input features for the ML algorithms. However, the accuracy of CALPHAD depends strongly on the reliability of the thermodynamic databases[36] and integrating CALPHAD data with experimental data could bias the outcome of ML predictions. CALPHAD also offers a tool primarily for the prediction of equilibrium phases and as such, in cases where manufacturing methods result in nonequilibrium structures, e.g., for coatings, CALPHAD can yield unreliable data. Furthermore, the generation of a database using the CALPHAD method is both time consuming and computationally expensive.

To explore alternative uses of CALPHAD and ML in alloy selection, this study aims to establish an alloy design methodology that utilizes simple ML architectures trained on experimental databases. CALPHAD is instead integrated as a postprocessing tool to provide further insights into alloy phase formation, enabling alloy downselection.

To establish this novel methodology, a case study within the HEA field was chosen, centering on the development of hardmetal matrix phases for metal-forming applications that could serve as the basis for further carbide reinforcement to derive improved performance. This case study emphasizes hardness for improved wear resistance and a face-centered cubic (FCC) phase to maximize ductility as body-centered cubic (BCC) phases are typically observed to display a higher hardness,[37] but exhibit lower ductility compared to FCC phases.[7,12,38] For the purposes of this case study, single-phase solid solutions were deemed most beneficial to reduce the number of interfaces present. This case study was simplified to predictions of only phase formation and hardness to enable the easy assesment and demonstration of ML performance.

2. Aim of Work

The aim of this work was to develop a robust methodology, presented in **Figure 1**, for the investigation of HEA compositional space, for the design of structural alloys. The methodology was applied to the design of hardmetal matrix compositions as a case
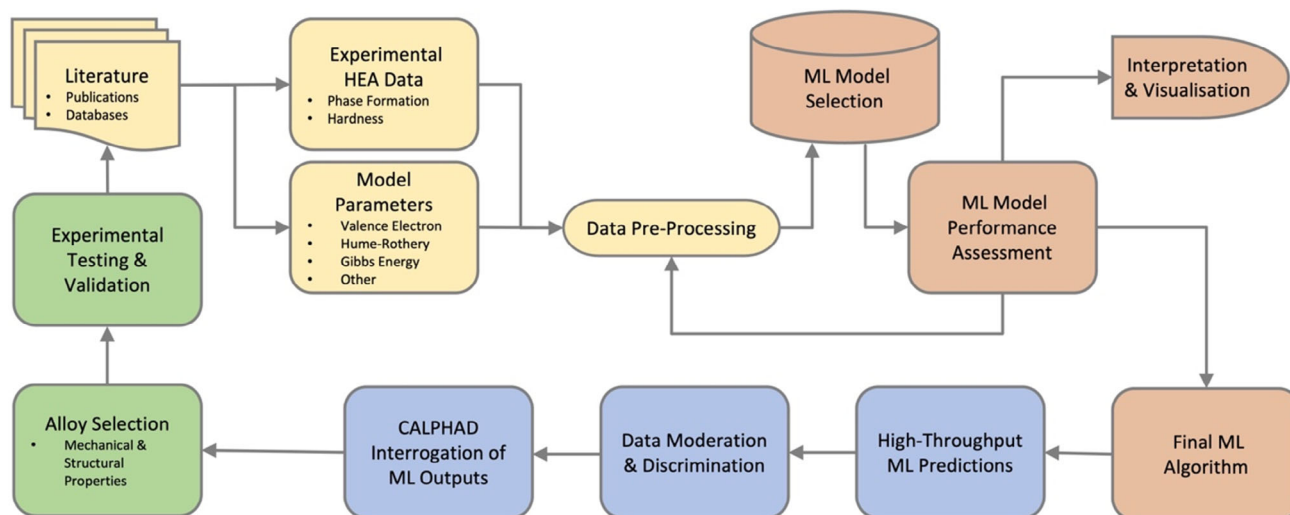
Figure 1. A schematic of the alloy design methodology used in this study, utilizing both ML and CALPHAD for compositional downselection.

study for the development of phase and hardness prediction models. An RF ML architecture was trained from available experimental HEA phase formation and mechanical property data to make high-throughput predictions on the phase formation and hardness of potentially suitable unseen HEA compositions. Additionally, the CALPHAD method was used as a postprocessing technique to interrogate the ML predictions of phase formation for the downselection of novel and potentially suitable HEA compositions for experimental assessment. Consequently, six alloy compositions were selected from the high-throughput analysis for experimental evaluation to assess their suitability for the use case selected, validate the ML methodology, and explore further modifications to enable increased complexity to be built into the models. Albeit a derivative first step to allow the development of algorithms and architectures, this methodology presents an opportunity for the design and development of novel alloy compositions for a host of different structural and functional applications.

3. Machine Learning

Several studies utilizing ML for the design of HEAs have been reported in the literature and the reader is referred to the comprehensive reviews by Liu et al.[39,40] Arróyave[41] also provided a detailed discussion on ML for phase prediction and stability. The majority of early ML studies focus on the prediction of a single property, either phase formation or hardness. For example, Huang et al.[21] trialed three different ML models, k-nearest neighbors, support vector machines, and a neural network to predict phase formation from a database of 401 alloy compositions. Good performance was achieved, with the neural network producing a prediction accuracy in excess of 80% on average. However, the phase prediction in this study is simplified to classification into three classes consisting of solid solution, intermetallic, or solid solution plus intermetallic. Similarly, Kaufmann et al.[31] applied an RF model to predict phase formation from a database of 1798 alloy compositions. A novel model confidence

measure was applied to assess model performance, achieving 75% confidence in model predictions. But, only 134 of the data points were experimentally determined, utilizing DFT to supplement the available data with a further 1664 compositions. For prediction of HEA hardness, Yang et al.[42] again trialed several ML algorithms and determined support vector machines to be the best model on their dataset of 370 data points. High model prediction performance with a root mean square error (RMSE) of 75 and coefficient of determination of 0.94 was reported. Beniwal et al.[43] apply an ensemble of 165 artificial neural networks to predict hardness of HEA systems from a dataset of 218 data points constructed by Gorsse et al.[44] Good model performance was reported with a mean absolute error (MAE) of 82.8 HV and the nature of model predictions was probed further, investigating the impact of each feature on predicted hardness over continuous composition variations.

In addition to the models constructed with a single target output described above, many studies have begun to utilize ML to predict multiple outputs of microstructural and mechanical properties for the design of new structural HEAs. For example, Huang et al.[45] compared five common ML algorithms for hardness and solid-solution single-phase formation, with RF producing the best performance on a dataset of 106 alloy compositions. Good model performance was achieved and ten alloys were experimentally fabricated to validate the model predictions. However, this study limited the experimental search to two HEA systems to simplify the alloy fabrication process. Furthermore, both Jain et al.[46] and Shen et al.[47] utilize ML to predict both phase formation and hardness. Jain et al.[46] utilized two different models, an extra trees classifier and artificial neural network, trained on 1120 and 99 data points to predict phase formation and hardness, respectively. In contrast, Shen et al.[47] employed an XGBoost model to predict both phase formation and hardness from a dataset containing over 500 data points. Good model performance was achieved by both with phase prediction accuracies around 90% and higher, and MAEs less than 35 HV.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
ENGINEERING
MATERIALS**

www.aem-journal.com

The literature outlined above indicates that successful ML models, comprising architectures from support vector machines to artificial neural networks, on a range of data and input features, can be constructed to predict microstructural and mechanical properties of HEAs. The purpose of the study herein was twofold. First, to compare and evaluate two models constructed with the same ML algorithms from two distinct experimental HEA databases for phase prediction. Second, utilizing the same data source and model architecture to compare phase and hardness prediction models. Interpretability metrics were extracted to understand the equivalence of feature importance between the two models.

3.1. Database Selection

Two independent experimental HEA databases were selected to train and test the RF model. The database produced by Gorsse et al.[44] "Database on mechanical properties of high entropy alloys and complex concentrated alloys," contains experimental data collected from studies on 370 alloys and was chosen for its unique inclusion of mechanical property information. A corrigendum was published for this database with a new dataset that both corrects errors and includes previously omitted data.[48] The updated dataset was used in this study. The second database, produced by Machaka et al.[49] "Machine learning-based prediction of phases in high-entropy alloys: A data article," was selected as at the time of this study it was the most recently published HEA dataset and contains data on 1360 alloys with extensive phase formation information, albeit lacking in property data.

3.2. Data Cleaning and Processing

The database produced by Gorsse et al.[44] contained 123 compositions with missing hardness data. Following the approach by Huang et al.[45] and Wen et al.[33] to maximize the amount of training data available for hardness predictions, an empirical relationship, given by Equation (1), was utilized to extrapolate hardness values from the yield strength provided in the database, where σ_y denotes the yield strength and HV denotes the Vickers hardness. This relationship between yield strength and hardness has been shown to be approximately obeyed by BCC HEAs.[33,45,50] Despite application of this relationship, 28 compositions still had missing hardness data. The average of the hardness values available in the dataset was imputed for these compositions. The Machaka et al.[49] database contained extensive phase formation data for every composition, but seventeen compositions had to be omitted from the data used to train the ML model. Data were omitted where the compositions were outside the compositional space considered for the case study herein.

The Gorsse database is important as it contains both mechanical property data and a richer description of the constituent phases, but contains no processing information; while the Machaka database contains sporadic processing information, but no mechanical property data and a more simplified description of the phases present. Given these limitations, in the current study it was deemed appropriate to include all the available data. Ideally, a subset of the data accounting for processing history would be chosen to train the ML. However, the databases available and used in this study are insufficiently complete to enable such a reduction in training data. This highlights a deficiency in the databases currently available. Development and expansion of future databases should be focused on including a more holistic processing history, enabling manufacturing to be taken into account in the implementation of ML.

$$\sigma_y(\mathrm{MPa}) \approx \frac{1}{3}H(\mathrm{MPa}) = \frac{9.81}{3}\mathrm{HV}(\mathrm{Hv}) \qquad (1)$$

3.3. Machine Learning Input Features

The training of the ML models and their use in the prediction of hardness and phase formation of new HEA compositions were based upon the physical properties of the constituent elements of the alloys. Properties such as atomic radii and valence electrons and bulk elemental properties, such as melting temperature and Young's modulus, were used. These physical properties were transformed into features that mathematically describe the alloy and were thought to be relevant to phase formation and hardness within HEAs. Feature selection is critical in ML studies in materials science to produce meaningful and interpretable predictions. Hence, ten independent features were selected from the relevant literature, detailed in **Table 1**, calculated through an assortment of Hume–Rothery rules, Gibbs free energy rules, and valence electron criteria. These features have previously been shown to be useful in identifying key areas of compositional space. For example, Guo et al.[51] demonstrated that valence electron concentration effectively discriminated between FCC and BCC phase formation within HEAs. As CALPHAD is being used as a downselection method, features derived from CALPHAD calculations, commonly used in ML HEA studies, have been excluded in this work, e.g., the solidus and liquidus temperatures of the alloys.[31] A subset of these features was trialed in different combinations to assess their impact on the predictive capability of the models. It was found that utilizing the full feature space detailed in Table 1, resulted in improved prediction accuracy, while importantly, the correlation matrix (discussed later in Section 4) did not reveal excessive correlations between features that may result in overfitting. Consequently, in the models trained herein, all features were retained. In subsequent sections, the importance and correlation between the individual features are discussed.

3.4. Machine Learning Model Selection

The critical factor in the model selection process is the availability of data[52] and it is common practice for ML studies to compare multiple models to find the optimal model for the available data.[46,47] For example, Bundela and Rahul[53] investigated the performance of a number of different ML models for the prediction of mechanical properties on the same database, produced by Gorsse et al.[44] utilized in this study. Concluding that an artificial neural network performs well on the experimental data, but ultimately finding that an XGBoost model performs best. In this study, an RF architecture was ultimately selected because it was found to outperform the XGBoost proposed by Bundela and Rahul, in addition to other models, such as simple linear regression and support vector machines, on the chosen datasets.

Table 1. A table of the features used to mathematically describe the alloy compositions, enabling the ML models to make predictions on their phase formation and hardness.

Parameter symbol[a]	Parameter name	Equation	References
δ	Atomic size difference	$\delta = 100\sqrt{\sum_{i=1}^{n} c_i \left(1 - \frac{r_i}{\bar{r}}\right)^2}$	[72,75,76]
γ	Atomic packing parameter	$\gamma = \omega_S / \omega_L,$ $\omega_x = 1 - \sqrt{\frac{(r_x + \bar{r})^2 - \bar{r}^2}{(r_x + \bar{r})^2}}$	[75]
$\Delta\chi$	Electronegativity difference	$\Delta\chi = \sqrt{\sum_{i=1}^{n} c_i (\chi_i - \bar{\chi})^2}$	[77]
VEC	Valence electron concentration	$\text{VEC} = \sum_{i=1}^{n} c_i (\text{VEC})_i$	[51,78]
e/a	Number of itinerant electrons per atom	$\frac{e}{a} = \sum_{i=1}^{n} c_i \left(\frac{e}{a}\right)_i$	[25,51,78]
ΔH_{mix}	Enthalpy of mixing	$\Delta H_{mix} = 4 \sum_{i=1, i<j}^{n} c_i c_j \Omega_{ij}$	[79–81]
ΔS_{mix}	Entropy of mixing	$\Delta S_{mix} = -R \sum_{i=1}^{n} c_i \ln c_i$	[7,80]
T_m	Weighted melting temperature	$T_m = \sum_{i=1}^{n} c_i (T_m)_i$	[82]
Ω	Yang parameter	$\Omega = \frac{T_m S_{mix}}{|H_{mix}|}$	[82]
$\Delta\varepsilon$	Young's modulus asymmetry	$\varepsilon = \sqrt{\sum_{i=1}^{n} c_i \left(1 - \frac{E_i}{\bar{E}}\right)^2}$	[83]

[a] Quantities denoted with a bar indicate that it is the average value, while quantities with an i indicate the i'th element. c represents the atomic fraction of the element, r denotes the atomic radii, and E represents the Youngs modulus. R is the molar gas constant. For γ, ω_x denotes the solid angles around the largest and smallest atoms, represented by subscript L and S, respectively. In ΔH_{mix}, Ω_{ij} is the enthalpy coefficient for elements i and j, respectively.

This is likely caused by the different methodology of data cleaning and the chosen input features used in this study. Deep learning and neural networks were not selected, due to the limited number of data available being unsuitable for training and validation of these model architectures,[52] and their increased complexity compared to RF for little gain in prediction accuracy.[54] The RF was also chosen for its ease of construction, interpretability,[55] ability to perform both classification and regression tasks,[56] and its previous successful application in material science studies.[31,45,57] The capability to perform both classification and regression tasks was crucial to both classifying phase formation of compositions and predicting alloy hardness through regression analysis. RF combines many individual and uncorrelated decision trees that are constructed and run in parallel with no interaction. Each tree generates an output for the prediction. In regression analysis, the vote of each decision tree is averaged to produce the final prediction of the model. In contrast, for classification analysis, the majority vote from all the decision trees is taken as the final output of the model.[24,58] The RF model used in this study was created using the scikit-learn ML toolkit in Python.[59] Fivefold shuffled cross-validation was also implemented to enable optimization of model hyperparameters and provide a greater insight into model performance. Optimizing model hyperparameters to, for example, control the growth of trees in the forest is particularly useful for small databases to minimize the likelihood of overfitting. Additionally, a certainty metric for RF predictions, proposed by Kaufmann et al.[31] was employed to provide further insight into model prediction confidence.

3.5. Random Forest Hyperparameters

Hyperparameters are settings that control the learning process of the ML model during training. These parameters can be finely tuned and optimized to improve the performance of the model. Each type of ML algorithm has different hyperparameters that impact the learning process in different ways. In the case of RFs, for example, these include the number of decision trees in the forest, how deep each tree is, and the maximum number of features considered when performing a split inside the tree. To determine the optimal hyperparameters, a randomized hyperparameter fivefold cross-validation grid search was performed across a range of values. The algorithm selects a new combination of hyperparameter values on each iteration to construct the RF. To evaluate the random hyperparameter search, the best model produced by the random search was compared to the base model to see if there was an improvement in performance. If there was no improvement in performance the range of hyperparameters was adjusted and the random grid search run again. When a satisfactory improvement in performance was found, then these hyperparameters were implemented into the final ML algorithm.

3.6. Model Training Process

Three individual RF models were produced and trained from the two available databases. From the Gorsse et al.[44] database, a classification model for phase prediction and a regression model for hardness prediction were developed, denoted model X and model Y, respectively. From the Machaka et al.[49] database, another classification model for phase prediction was developed, denoted model Z. Based on the information available in both databases, model X considers 15 different phase outcomes whereas model Z considers 7 different phase outcomes, detailed in Table S1, Supporting Information.[44,49] Unlike the majority of phase prediction models presented in the literature that only consider binary or tertiary classification tasks, the models trained

herein have an increased number of outcomes. It is therefore anticipated that the classification models from this work will result in lower prediction accuracies in comparison to binary and tertiary classification tasks.[29,31,45,54,56,60]

Cross-validation was implemented to provide an accurate assessment of the performance of the RF models. The data were split into five equal and randomized folds, with each being used once as the testing set while the others were used for training. After cross-validation and assessments of the model's performance, the models were retrained using the whole database, enabling training on the maximum amount of data available to improve their performance. A measure of the prediction confidence, developed by Kaufmann et al.[31] for RF classification algorithms, was applied in this study. The model's confidence in its prediction was calculated as the ratio of the number of decision trees inside the forest that voted for the final phase prediction of that composition, against the total number of trees. Hence, the more trees that vote for a phase, the more confident the model is in its prediction of that phase. This confidence measure was not applied to the hardness prediction model Y, as it is not suitable for an RF regression model. Instead, the coefficients of determination, R^2 and the RMSE, were used to assess the error and confidence in the regression model.

3.7. Generation of Virtual Candidate Search Space

After the model training process, a virtual candidate search space was created for the ML to make predictions on the phase formation and hardness of a large number of HEA compositions. Elements were selected that were included in the chosen databases based on domain knowledge to promote FCC phase formation, as defined in the design criteria. Additionally, the elements were selected for minimizing cost and maximizing raw element abundance, not considering recycled sources. The process for generating the candidate search space followed the sequence described below: 1) Definition of key elements relevant to the case study and design application: Fe, Ni, Co, Al, Ti, W, Cr, Mn, Hf, Nb, Mo, and Ta. 2) Consideration of every possible equiatomic permutation of these elements in a five-element system without repetition, in this case for a total of 792 compositions. 3) Creation of every possible compositional permutation of these five element systems between a minimum of 5 at% and a maximum of 45 at% elemental weighting, with a granularity of 5 at%, yielding a total of 2 238 193 compositions.

4. Machine Learning Performance, Outputs, and Discussion

Fivefold cross-validation assessments of the performance of the RF models after the training process were conducted to understand how the model would perform on the training data. Accuracy, precision, recall, and F_1 scores were all calculated to assess model performance for the phase classification task, the results of which are shown in **Table 2**.[61,62] Performance scores >78% and >82% for classification models X and Z, respectively, indicate that the models successfully predict phase formation on the validation data from the dataset. Furthermore, it suggests that the models do not suffer from overfitting,

Table 2. A table of the metrics calculated to assess performance of the classification models at phase prediction. Model X was trained on the Gorsse et al.[44] database and model Z on the Machaka et al.[49] database.

Performance metric[a]	Equation	Model X	Model Z	Model Y
Accuracy	$A = \frac{TP+TN}{TP+TN+FP+FN}$	78.7%	82.2%	–
Precision	$P = \frac{TP}{TP+FP}$	80.9%	82.9%	–
Recall	$R = \frac{TP}{TP+FN}$	78.7%	82.2%	–
F_1 Score	$F_1 = 2 \times \frac{P \times R}{P+R}$	76.9%	81.8%	–
MAE	$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$	–	–	66.1 HV
Coefficient of determination	$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$	–	–	73.7%
RMSE	$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}$	–	–	92.6 HV

[a]True positives (TP) are the data points that are correctly predicted by the model for a class. False negatives (FN) are data points that are incorrectly predicted as a different class by the model. False positives (FP) are data points that are a different class but are predicted to be the class under consideration. True negatives (TN) are data points that are a different class to the one under consideration and are predicted to be a different class, thus the class being considered is not involved. y denotes the true value, $\hat{y}$ represents the predicted value, $\bar{y}$ represents the mean of the true data, and N is the number of data points.

enabling potential generalizability to predictions on unseen compositions within the virtual candidate search space. In contrast to the phase classification models, numerically and physically meaningful values for accuracy of the hardness regression model (Model Y) could be obtained. The MAE, R^2, and RMSE were utilized to evaluate the error between the predicted value produced by model Y and known value of hardness from the database. The results of these performance assessments, determined for the hardness predictions of model Y are also displayed in Table 2. These indicate a good fit and correlation between the predictions and experimental data within the databases used for training the models.

Figure 2 depicts the correlation between the predicted hardness and database hardness in the testing dataset. The majority
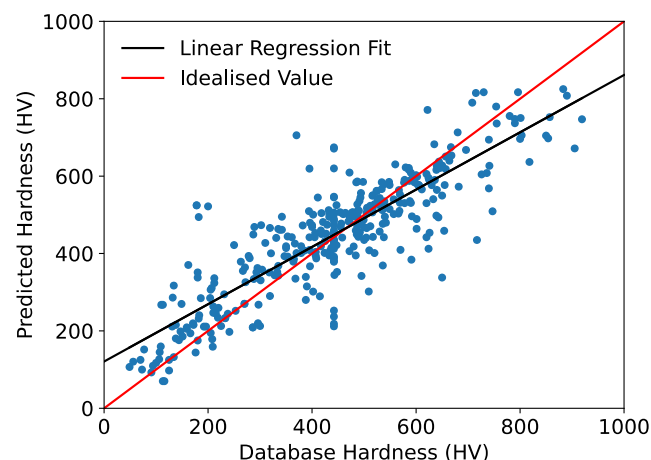
Figure 2. A plot of the known alloy hardness in HV from the Gorsse et al.[44] database against the hardness predicted by model Y across all folds of the cross-validation testing.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
ENGINEERING
MATERIALS**

www.aem-journal.com

of points lie close to the diagonal red solid line that represents perfect agreement, again showing the good correlation between predicted and database hardness values. A linear fit of the predicted hardness against the database hardness is denoted by the black solid line. The closer the diagonal red solid line is to the linear fit of the black solid line, the lower the systematic error of the model.

The Pearson's correlation coefficient (PCC) of the input features, denoted by r, Equation (2), was also calculated for both databases to measure the linear correlation between any two of the features, highlighting any interdependencies

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \qquad (2)$$

where x and y denote two of the features and $\overline{x}$ and $\overline{y}$ represent the mean of the two features, respectively. PCC values can range from $+1$ to -1, with positive values indicating a positive relationship between the variables and negative values indicating a negative relationship. Commonly in correlation analysis, if the correlation coefficient between two features is >0.80, then this is considered a very strong correlation[63,64] and the feature that ranks the highest in the feature importance is retained, while the other feature is eliminated from the model.[65,66] However, ML studies in the HEA space typically allow for much stronger feature correlations before eliminating features from the model, such as $r > 0.95$.[21,26,33,45,54,66] When analyzing the results of correlation analysis between features utilized by the ML model, context is critical to understand the extent and impact of the correlation. Domain knowledge is useful in helping understand and interpret correlations in a more meaningful way. It can be seen from the PCC matrices in **Figure 3**, that the features used in this study are correlated to varying degrees. In both correlation matrices, the pair of features having one of the highest degrees of correlation are δ and γ with values of 0.77 and 0.62 for models based on the databases by Gorsse et al.[44] and Machaka et al.[49] respectively. This was anticipated as both features relate to the distribution of atomic radii of the alloys' constituent elements. However, based on the levels of correlation observed it was determined suitable to retain both features for training of the ML model. $\Delta\chi$ and $\Delta\varepsilon$ were the next strongest positively correlated features from the two databases, 0.72 and 0.65 across the Gorsse et al.[44] and Machaka et al.[49] databases, respectively. Interestingly, comparing across the literature, a broad range of correlation values have been reported for these features. Huang et al.[45] reported a value comparable to this study of 0.71 for the correlation of $\Delta\chi$ and bulk modulus asymmetry (a similar correlation would be expected for $\Delta\varepsilon$[67]). In contrast, Chen et al.[64] reported a correlation value of 0.11, although this may be the result of considering a single HEA system. The strongest negative correlation occurs between valence electron concentration (VEC) and $\Delta\varepsilon$, -0.68 and -0.62 across the Gorsse et al.[44] and Machaka et al.[49] databases, respectively. This observed correlation agrees with the negative correlation of -0.71 reported by Chen et al.[64] The negative correlation observed between VEC and e/a was unexpected as both features describe electronic structure and hence, they would be expected to positively correlate. However, both VEC[51] and e/a[25] have been shown to be effective in the prediction of HEA phase

formation, thus they were both retained within the models. The feature correlations observed across the two databases are comparable and hence it would be expected that the features would have similar impacts on prediction of phase formation and hardness across the ML models constructed from these databases.

To interpret and understand the ML, permutation feature importance was utilized. Permutation feature importance is a global ML interpretation technique that describes the average behavior of the model to show general mechanisms and trends, by measuring the increase in prediction error as the model's parameters are permutated. A feature is considered important if randomly shuffling its values increases the error of the model, as in this case, the model relies upon this feature to make its predictions. In contrast, a feature is considered unimportant if shuffling its values does not significantly impact the model's error.[24] Hence, permutation importance analysis highlights the key features influencing phase and hardness predictions in HEAs.

The results of the permutation feature importance assessment in this study are shown in **Figure 4**. VEC is considered the most important feature for phase prediction in both classification models, X and Z. In addition, VEC is also an important feature in model Y, for hardness predictions, an encouraging result due to the dependence of hardness on underlying phase constitution. This agrees strongly with literature, with several reports finding VEC to be the most important feature in the determination of phase formation in HEAs.[28,33,56,60] Furthermore, the importance of VEC in phase formation has previously been established by Hume–Rothery, who found that similar crystal structures are formed if the VEC of two intermetallic compounds are comparable.[68,69] In addition, Guo et al.[51] demonstrated that higher values of VEC lead to FCC phase formation and lower values of VEC lead to BCC phase formation. Notably, model X is found to be more strongly affected by VEC than model Z, likely reflecting the experimental databases used. The Gorsse et al.[44] database used for model X includes more complete phase information, whereas the Machaka et al.[49] database used for model Z, often groups intermetallic phases and HCP solid solutions together into simpler phase classes, as described in Table S1, Supporting Information. Furthermore, the relative size of the databases, with the Machaka et al.[49] database being significantly larger than the Gorsse et al.[44] database, will likely reduce the individual feature dependence of the derived models. In contrast to VEC, $\Delta\chi$, is found to have little impact on the predictions of all three models, shown last in all permutation feature importance assessments. For the hardness predictions of regression model Y, $\Delta\varepsilon$ is shown to be significantly the most important feature, but in contrast, it is not considered important by the two-phase classification models X and Z.

SHAP is a local interpretation method that can be used to explain individual predictions by determining the contribution of each feature.[70] SHAP summary plots combine feature importance and feature effects[24] and are shown in **Figure 5** for each model, respectively. Each point in the plots represents a SHAP value for a feature and prediction for an individual composition. Overlapping points are jittered in the y-axis to provide an illustration of the distribution of SHAP values per feature. A positive/negative SHAP value indicates a positive/negative impact on model predictions, associated with the likelihood to predict FCC phase formation and values of hardness, both for the
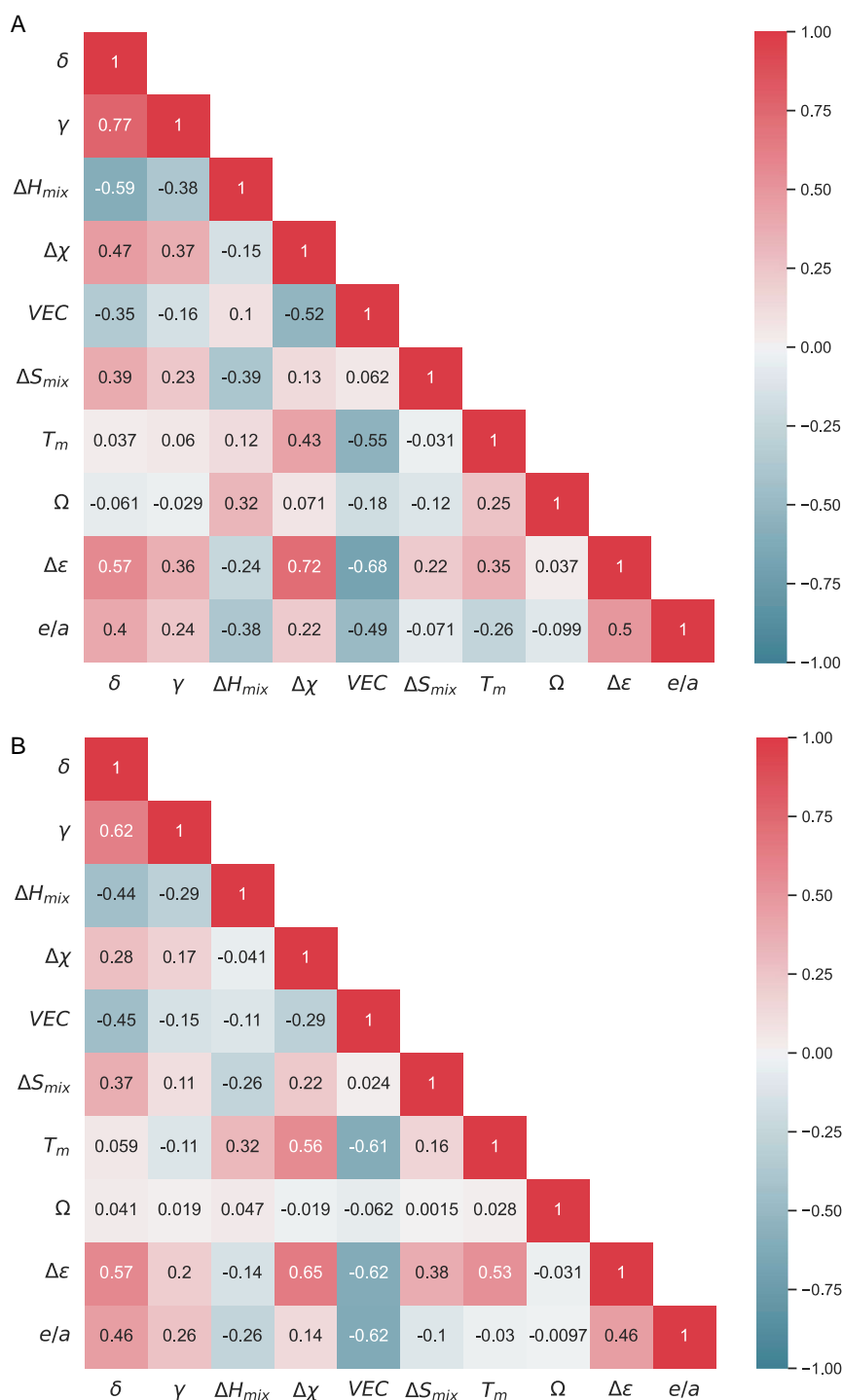
Figure 3. A correlation map between the ten features used as mathematical descriptors for the ML models. The value in the grid shows the PCC between the different features and the color intensity is proportional to the magnitude of the PCCs. A) From the Gorsse et al.[44] database; B) From the Machaka et al.[49] database.

classification and regression tasks, respectively.[42,71] Hence, the wider the horizontal distribution of points, the greater the influence of that feature on the models' predictions. Additionally, features are ordered on the y-axis by their importance for predictions. The color of the points denotes the value of the

feature in question. Red indicates a larger feature value, while blue indicates a smaller feature value.[24]

Equivalent conclusions to the PCC and permutation importance in Figure 3 and 4, respectively, can be drawn from the SHAP summary plots shown in Figure 5. All three models place
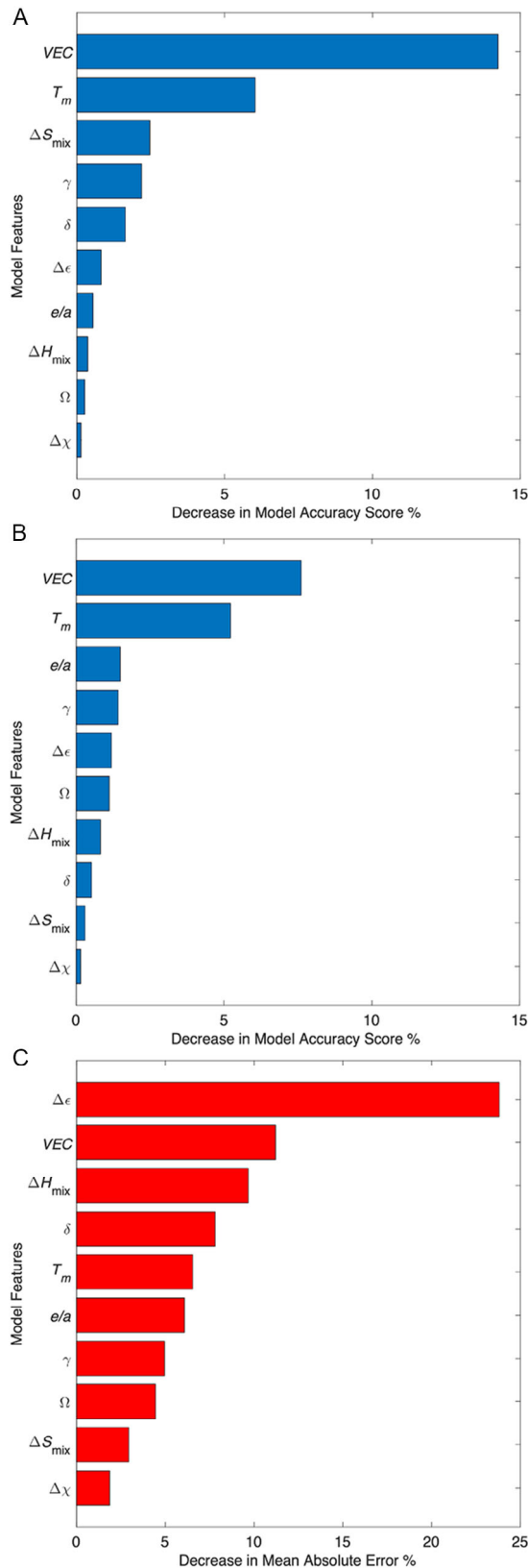
**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
ENGINEERING
MATERIALS**

www.aem-journal.com

Figure 4. Plots of the permutation importance of the ten features used as mathematical descriptors for the ML models. A) From classification model X; B) From classification model Z; C) From regression model Y.
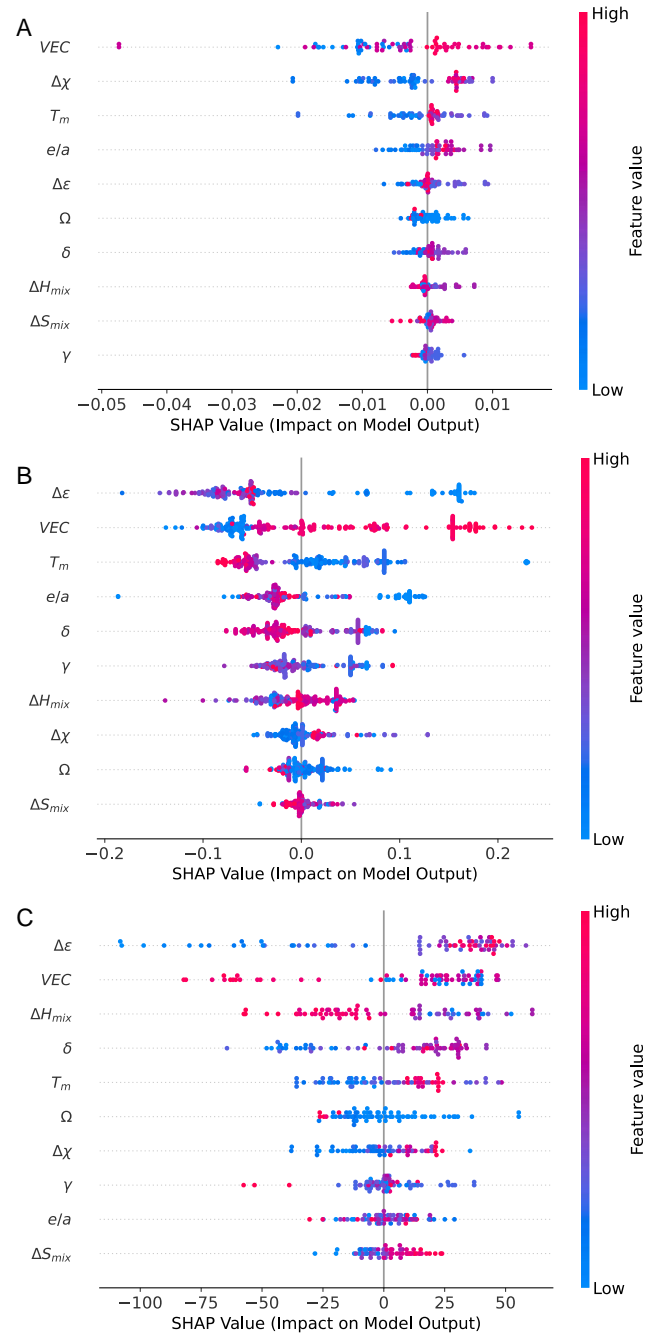

Figure 5. SHAP value distribution plots of different compositions, showing the importance and effects of different features. A) From model X on FCC phase formation; B) From model Z on FCC phase formation; C) From model Y on hardness predictions.

a high importance on VEC. Large VEC values produce large positive SHAP values for FCC phase prediction and large negative SHAP values for hardness prediction. Indicating that for compositions with a larger VEC value, the ML is more likely to predict FCC phase formation and lower values of hardness. This shows strong agreement with the formation of FCC phases at higher values of VEC, resulting in lower values of hardness, as harder BCC and intermetallic phases are less likely to form.[22,51,54,72]

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
ENGINEERING
MATERIALS**

www.aem-journal.com

This is encouraging that the ML models are producing outputs with sensible dependencies based on empirical physics understanding. In contrast, ΔS_{mix} is placed toward the bottom of all three SHAP summary plots. Model Y places low importance on the configurational entropy in the prediction of alloy hardness, producing a 0.30% decrease in model performance under permutation analysis, as shown in Figure 4b. In contrast, for models X and Z, it is shown to have a slightly more significant and also comparable impact. A 2.47% and 2.94% decrease in model performance under permutation analysis are shown in Figure 4a,c. Models X and Z predicting phase formation are trained from different databases but perform comparably. Both databases[44,49] used in this study have comparable numbers of compositions forming solid solutions, around 60%. In this case, the calculation of the configurational or mixing entropy is based upon ideal randomly distributed solid solutions and is often not representative of the total entropy. Excess vibrational, magnetic moment, and electronic effect terms can contribute significantly to the total entropy and impact the role of entropy on phase selection within HEAs.[9]

4.1. High-throughput Machine Learning Predictions and CALPHAD Analysis

All three RF models were used to make high-throughput predictions on the phase and hardness of a series of unseen HEA compositions, with the goal of downselecting a small sample of compositions for experimental testing, to further validate and refine the models. The results of these high-throughput predictions were collated and subsequently abridged. To satisfy the design criteria outlined for the case study considered herein, CALPHAD analysis was employed to enable further downselection of alloys for experimental fabrication from the ML outputs.

To meet the design criteria of the chosen case study, as a first step, all compositions where classification models X and Z did not **both** predict FCC phase formation were eliminated. A total of 24 613 HEA compositions were predicted to form an FCC phase by both models X and Z. CALPHAD calculations of the equilibrium phase formation of the remaining HEA compositions at 1000 °C were also obtained using the TCNi8 database within the Thermo-Calc software package and automated using TQ-Fortran. CALPHAD calculations are known to be less accurate at "low" temperatures, predicting often kinetically inhibited or unrealistic intermetallic phase formation. In addition, calculations at temperatures over 1000 °C may eliminate a number of compositions if the solidus temperature of the material is low. 1000 °C was therefore chosen as a compromise temperature between accuracy of CALPHAD predictions and practicability of the chosen case study, with 1000 °C representing a sensible operating limit. TCNi8 was utilized in this study as an FCC crystal structure was targeted and hence it was considered to be the most relevant database, providing more holistic information for the systems under consideration. Compositions were sorted by FCC phase fraction formation and subsequently, ML-predicted hardness. The twelve hardest FCC-predicted compositions (by all models and CALPHAD) are shown in **Table 3**. Due to compositional similarities, the compositions highlighted in bold in Table 3 and labeled A–C, were chosen as the most interesting

Table 3. Table of the hardest alloy compositions according to model Y, also predicted to display single-phase FCC formation by models X and Z, and CALPHAD equilibrium phase fraction calculations. The compositions highlighted in bold and labeled A–C were chosen as the most interesting to experimentally investigate.

Sample index	Alloy	ML confidence model X [%]	ML confidence model Z [%]	ML hardness model Y [HV]
A	**$Cr_{20}Mn_{20}Fe_{10}Ni_{45}Hf_5$**	47.0	41.5	483
–	$Cr_{20}Mn_{15}Fe_{15}Ni_{45}Hf_5$	43.0	42.5	479
B	**$Ti_{10}Cr_{10}Mn_{25}Fe_{20}Ni_{35}$**	57.0	37.4	476
–	$Ti_{20}Mn_{20}Fe_5Co_{10}Ni_{45}$	43.0	38.6	474
–	$Ti_{20}Mn_{25}Fe_5Co_5Ni_{45}$	39.0	39.5	473
–	$Cr_{15}Mn_{20}Fe_{15}Ni_{45}Hf_5$	49.0	41.6	471
–	$Cr_{15}Mn_{25}Fe_{10}Ni_{45}Hf_5$	48.0	45.5	471
–	$Cr_{15}Mn_{15}Fe_{20}Ni_{45}Hf_5$	48.0	41.6	468
–	$Cr_{20}Mn_{10}Fe_{20}Ni_{45}Hf_5$	45.0	40.5	461
–	$Cr_{10}Mn_{25}Fe_{15}Ni_{45}Hf_5$	50.0	46.4	460
–	$Cr_{10}Mn_{20}Fe_{20}Ni_{45}Hf_5$	50.0	46.9	452
C	**$Ti_{10}Mn_{45}Fe_{15}Ni_{25}Mo_5$**	32.0	37.5	451

to experimentally investigate. In addition, the three highest hardness alloys, as shown in **Table 4** and labeled D–F, irrespective of CALPHAD-predicted phase formation, were selected for experimental testing to enable an improved understanding of the performance of the regression model constructed. It is worth noting that the higher hardness as predicted by model Y in Table 4 is likely due to the alloys not being FCC based, as indicated by the CALPHAD predictions, despite being projected to form a single-phase solid solution FCC by ML models X and Z.

This methodology, first applying ML and subsequently downselecting using CALPHAD, has a number of key benefits.[62] Performing CALPHAD analysis of over 2 million compositions considered by the ML in this study is computationally impractical and time intensive. Initial application of ML dramatically reduces the number of compositions that need to be considered and hence the computational time. Applying CALPHAD data as an input to the ML model raises the issue of training the model on calculated data as opposed to experimentally determined data. While CALPHAD offers a remarkable tool for alloy design, it is well documented that calculation accuracies are lower for compositions away from established alloy systems. This is due to the construction of the databases being built from binary and ternary systems as well as the lack of full characterization of such systems across the relevant compositional space. Furthermore, CALPHAD provides calculations of equilibrium phases only. Therefore, caution is necessary as for manufacturing purposes reaching equilibrium phases can be unrealistic. Hence, CALPHAD can bias the results against compositions that are kinetically sluggish, but would otherwise be suitable for application. While this latter point is not applicable to the chosen case study, the use of ML is envisaged to progressively evolve for increasingly more complicated case studies, where equilibrium and indeed the use of CALPHAD may in fact be inadvisable. Therefore, in this case study CALPHAD was used simply as a downselection tool only.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
ENGINEERING
MATERIALS**

www.aem-journal.com

Table 4. Table of the hardest alloy compositions according to model Y, also predicted to display single-phase FCC formation by models X and Z, irrespective of CALPHAD equilibrium phase fraction calculations.

Sample index	Alloy	ML confidence model X [%]	ML confidence model Z [%]	ML hardness model Y [HV]	CALPHAD phase	CALPHAD phase fraction
D	$Al_{20}Ti_5Co_{30}Ni_{30}Nb_{15}$	26.0	26.0	587	C14 Laves, BCC, and FCC	42.5%, 41.0% and 16.5%
E	$Al_{30}Ti_{20}Fe_{30}Ni_{15}Mo_5$	40.0	39.0	584	BCC and C14 Laves	60.8% and 39.2%
F	$Al_{30}Ti_{20}Cr_{10}Fe_5Ni_{35}$	47.0	41.0	583	BCC	100%

Thus, the ML architecture can predict compositional hardness from experimental data, save significant time over the CALPHAD approach and can be quickly updated with new data, while providing reproducible and accurate predictions. Utilizing CALPHAD as a postprocessing tool to interrogate a small subset of the ML predictions provides further detailed microstructural information for alloy downselection, but importantly retains separation and allows critical interrogation of predictions. In the absence of further experimental data, CALPHAD provides a robust tool to enhance the phase predictions obtained from the ML models to target experimentation into both areas of agreement where technological benefits can be derived, as well as areas of discrepancies where improved experimental data can be of great importance.

5. Experimental Section

To validate the ML model and assess the suitability for the design use case of the downselected alloy compositions in Tables 3 and 4, the HEA systems were synthesized via arc melting in an inert Ar atmosphere. Button ingots of 30 g were produced for each composition from high purity (>99.5% by mass) bulk raw elements. To achieve alloy melt homogeneity, the ingot was inverted and remelted for a minimum of seven melting cycles with the same arc current intensity. The ingots were melted and solidified in a water-cooled Cu crucible, resulting in faster cooling rates than other typical casting methodologies. Due to the low evaporation point of Mn in comparison to the melting point of some of the refractory elements included in the alloy compositions, extra Mn (in the range 10–20%) was added to account for the anticipated losses during the manufacturing process. Differential scanning calorimetry (DSC) measurements of the as-cast alloys were conducted using a TA Instruments SDT Q600, to determine suitable temperatures for homogenization heat treatments. Samples were heated at a rate of 20 °C min^{-1} to a maximum temperature of 1450 °C and held at this temperature for two minutes before a controlled cool at the same rate to 400 °C, holding for two minutes. Two cycles of this heating and cooling regime were performed. Plots of the DSC measurements can be found in Figure S1, Supporting Information. Two separate homogenization regimes were identified based upon the DSC data, 1000 °C and 1150 °C for 48 h, for alloys A–C and D–F, respectively, followed by air cooling. These temperatures were selected to be as close to the solidus temperature of the alloys as possible without the risk of incipient melting.

Specimens from each as-cast and heat-treated condition were prepared for microstructural characterization by grinding with P400 – P2500 grades of SiC paper and polishing to a 1 μm finish using a diamond. Imaging and compositional characterization

were performed using scanning electron microscopy (SEM) coupled with energy-dispersive X-ray spectroscopy (EDX) on an FEI Inspect F50 and JEOL 7900 F operated at 20 keV and equipped with a Bruker XFlash 6 solid-state EDX detector. Additional phase characterization of each as-cast and heat-treated condition in bulk alloy form was performed using X-ray diffraction (XRD) on a Panalytical Aeris diffractometer using Ni-filtered Cu Kα radiation. Patterns were recorded in the $10° < 2\theta < 100°$ range at 0.02° increments and were analyzed using the full-pattern Pawley fitting procedure[73] in TOPAS-academic. Characterization of alloy hardness in both as-cast and homogenized conditions was performed using microindentation hardness on a Durascan 70 G5 by applying a 1 kgf load with a 15 s dwell time for a series of ten indentations, in accordance with ASTM E384.[74]

6. Experimental Results and Discussion

The bulk elemental compositions following the arc-melting fabrication procedure were confirmed by averaging large-area EDX scans. The results of this and comparison to the nominal compositions from the ML virtual candidate search space are detailed in **Table 5**. In the majority of cases, the bulk elemental compositions are within ±4% of the target concentrations for each element. However, for some elements, in particular for the Mo, Cr, and Mn, there is often a significant difference between the target and the fabricated compositions. This is due to the higher melting points of Mo and Cr and the low evaporation temperature of Mn compared to the other elements in the alloy systems leading to evaporation of Mn and possible lack of melting of Mo and Cr. The phase formation in the experimentally fabricated compositions was also evaluated through both the ML and CALPHAD and was found to produce the same predictions as the nominal compositions. Therefore, it was deemed suitable to use these alloy compositions for further experimental evaluation to assess the fidelity of the models utilized.

6.1. Microstructural Analysis of As-cast and Homogenized Material

Microstructural analysis of the alloy compositions was performed by SEM to determine the phases present and their relative chemistry. Backscattered electron (BSE) micrographs of the alloy compositions in the as-cast state are presented in **Figure 6**. Following the rapid solidification from the arc-melting process, all alloys exhibited a large-grained microstructure with high contrast due to both crystal orientation and compositional variation. SEM-based EDX maps, Figure S2–S13, Supporting

ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED
ENGINEERING
MATERIALS
www.aem-journal.com

Table 5. Table comparing the downselected alloys nominal composition generated as part of the virtual candidate search space to the experimentally fabricated composition.

Sample index[a]	Nominal alloy composition	Experimental alloy composition	ML confidence model X [%]	ML confidence model Z [%]	ML hardness model Y [HV]
A	$Cr_{20}Mn_{20}Fe_{10}Ni_{45}Hf_5$	$Cr_{22.3}Mn_{17.9}Fe_{10.2}Ni_{44.8}Hf_{4.8}$	42.0	41.5	477
B	$Ti_{10}Cr_{10}Mn_{25}Fe_{20}Ni_{35}$	$Ti_{10.0}Cr_{9.6}Mn_{27.0}Fe_{19.6}Ni_{33.8}$	60.0	43.5	478
C	$Ti_{10}Mn_{45}Fe_{15}Ni_{25}Mo_5$	$Ti_{10.3}Mn_{45.9}Fe_{15.5}Ni_{26.0}Mo_{2.3}$	37.0	48.0	429
D	$Al_{20}Ti_5Co_{30}Ni_{30}Nb_{15}$	$Al_{19.9}Ti_{5.0}Co_{30.4}Ni_{30.1}Nb_{14.6}$	24.9	30.9	586
E	$Al_{30}Ti_{20}Fe_{30}Ni_{15}Mo_5$	$Al_{29.7}Ti_{20.6}Fe_{30.3}Ni_{14.6}Mo_{4.8}$	40.7	40.0	586
F	$Al_{30}Ti_{20}Cr_{10}Fe_5Ni_{35}$	$Al_{29.6}Ti_{20.5}Cr_{11.4}Fe_{5.2}Ni_{33.3}$	42.7	41.0	590

[a]ML confidence and hardness predictions shown are performed using the experimental alloy compositions.
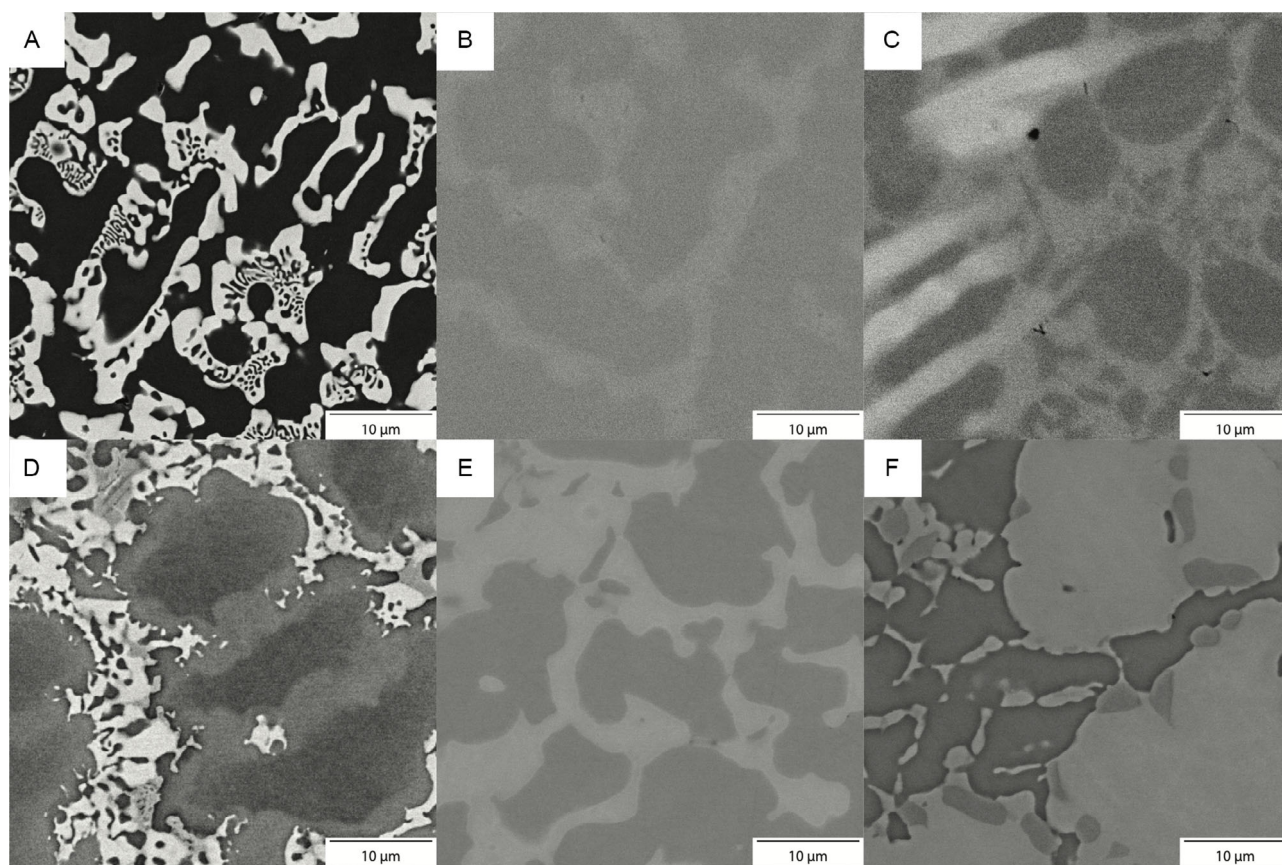

Figure 6. BSE micrographs of samples in the as-cast state showing two and three-phase dendritic microstructures. Samples are labeled A–F) according to Table 5.

Information, highlight clear elemental segregation within the grains, indicating dendritic solidification in all alloys. In all cases, no evidence of single-phase solid-solution formation was found, with all alloys displaying a multiphase microstructure.

To remove or mitigate the microsegregation observed due to the rapid solidification of the casting process, alloys A–C were homogenized at 1000 °C for 48 h and alloys D–F were homogenized at 1150 °C for 48 h, followed by air cooling. BSE micrographs of the alloy compositions in the homogenized condition are presented in **Figure 7**. The SEM-based EDX maps,

Figures S2–S13, Supporting Information, highlight clear elemental segregation within the grains between the matrix and precipitating phases. As with the as-cast state, there is no evidence of single-phase FCC solid solution formation in any of the fabricated compositions. However, there is a clear change in microstructure due to the homogenization process for all alloys. In all cases, the homogenization process was sufficient to remove segregation within the solidifying phases and ensure chemical equilibration of the composition of each phase. The skeletal interdendritic solidification in alloys in the as-cast state was also
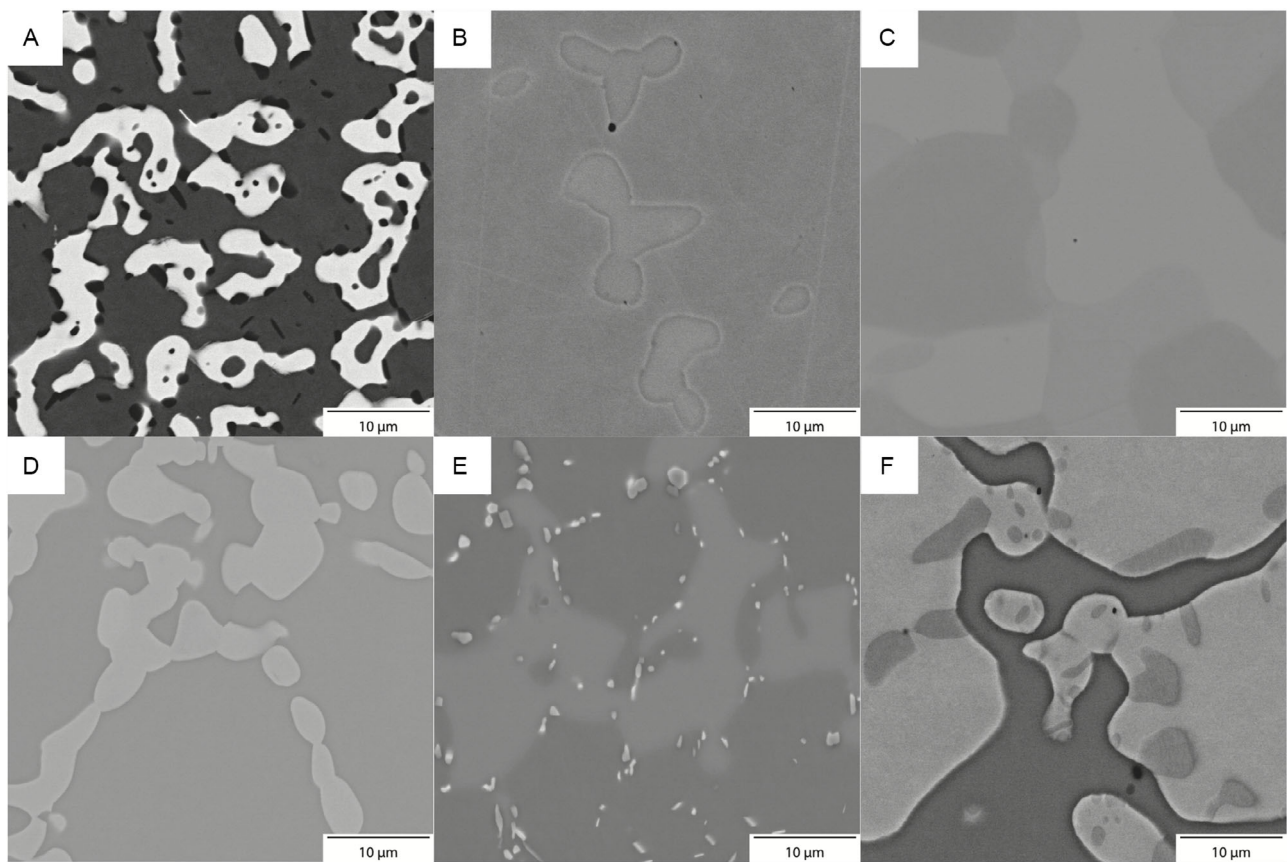
Figure 7. BSE micrographs of samples in the homogenized state according to the heat treatments outlined in Section 5, showing two and three-phase dendritic microstructures. Samples are labeled A–F) according to Table 5.

observed to break down in most cases in the as-homogenized samples in favor of discrete particles. Importantly, all compositions were found to be multiphase even after the homogenization/solution heat treatments.

Elemental partitioning information was obtained, and phase identification was performed through EDX and XRD techniques respectively. A summary of the data extracted from the EDX and XRD analysis is given in **Table 6**. Results presented in Table 6 agree with the SEM observations as discussed above.

6.2. Mechanical Property Analysis of As-cast and Homogenized Material

Mechanical properties of the alloys were assessed by microindentation hardness testing, as outlined in Section 5. Results of the microindentation hardness assessments of the alloys in both the as-cast and homogenized condition are presented in **Figure 8**. In all cases for alloys A–C, predicted to form single-phase FCC solid-solutions by both ML and CALPHAD, but experimentally

Table 6. Summary of the EDX and XRD analysis of the alloy systems in the as-cast and homogenized state. EDX elemental maps and an example XRD pattern overlaid with a Pawley fit can be found in the Supporting Information.

Sample Index	Experimental Alloy Composition	Condition	EDX Analysis	XRD Analysis
A	$Cr_{22.3}Mn_{17.9}Fe_{10.2}Ni_{44.8}Hf_{4.8}$	As-Cast	Light Contrast = Ni, Hf rich Dark Contrast = Cr, Fe rich	$FCC + HfNi_5 + BCC$
		Homogenised	Light Contrast = Ni, Hf rich Dark Contrast = Cr, Mn, Fe rich Dark Precipitates = Cr rich	$FCC + HfNi_5 + BCC$
B	$Ti_{10.0}Cr_{9.6}Mn_{27.0}Fe_{19.6}Ni_{33.8}$	As-Cast	Light Contrast = Ni and Ti rich* Dark Contrast = Cr, Mn, Fe rich	FCC
		Homogenised	Dark Contrast = Cr, Mn, Fe, Ti Light Precipitates = Ti rich*	$FCC + BCC + Ni_3Ti$

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
ENGINEERING
MATERIALS**

www.aem-journal.com

Table 6. Continued.

| Sample Index	Experimental Alloy Composition	Condition	EDX Analysis	XRD Analysis
C	$Ti_{10.3}Mn_{45.9}Fe_{15.5}Ni_{26.0}Mo_{2.3}$	As-Cast	Light Contrast = Fe, Mo rich Mid Contrast = Ni, Ti rich Dark Contrast = Mn, Fe rich	FCC + Ni_3Ti + Fe_2Mo (C14 Laves)
		Homogenised	Light Contrast = Ti, Fe, Mo Mid Contrast = Ti rich Dark Contrast = Ni, Mn	FCC + Fe_2Mo (C14 Laves)
D	$Al_{19.9}Ti_{5.0}Co_{30.4}Ni_{30.1}Nb_{14.6}$	As-Cast	Light Contrast = Co, Ni, Nb rich Mid Contrast = Al, Co, Ni, Nb Dark Contrast = Al and Ti rich	FCC + TiAl (B2) + $(CoNi)_2Nb$ (C14 Laves)
		Homogenised	Light Contrast = Co, Nb, Ni Dark Contrast = All Additional Co, Nb rich phase	Ni(TiAl) (B2)
E	$Al_{29.7}Ti_{20.6}Fe_{30.3}Ni_{14.6}Mo_{4.8}$	As-Cast	Light Contrast = Al, Ti, Mo rich Dark Contrast = Ni rich	BCC or B2 + Fe_2Ti (C14 Laves)
		Homogenised	Light Contrast = Al, Ti, Mo rich Dark Contrast = Al, Ti, Fe, Ni Light Precipitates = Mo rich	BCC or B2 + Fe_2Ti (C14 Laves) + $AlMo_3$
F	$Al_{29.6}Ti_{20.5}Cr_{11.4}Fe_{5.2}Ni_{33.3}$	As-Cast	Light Contrast = Al, Ti, Cr Mid Contrast = Cr rich Dark Contrast = Al, Ti rich	BCC or B2 + TiCrAl (C14 Laves) + $AlCr_2$
		Homogenised	Light Contrast = Al, Ti, Ni rich Mid Contrast = Cr rich Dark Contrast = Al, Ti rich	BCC or B2 + TiCrAl (C14 Laves) + $AlCr_2$
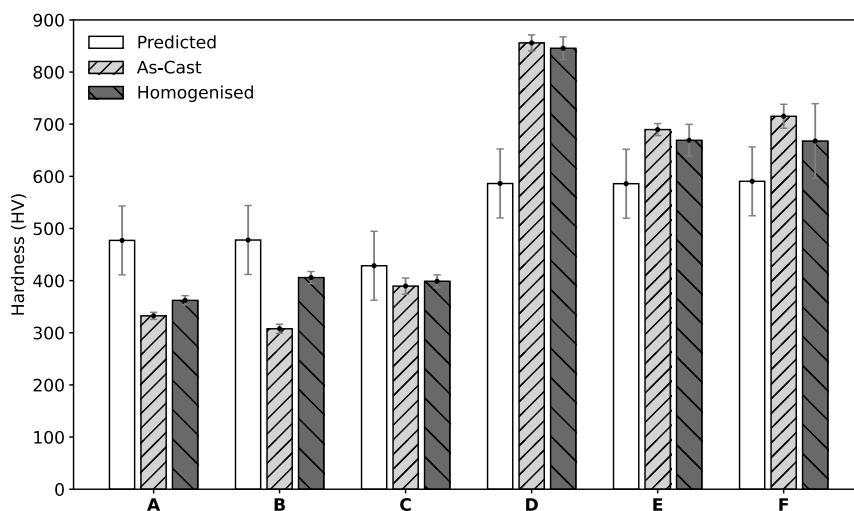
Figure 8. A bar chart showing the comparison between the ML-predicted, as-cast, and homogenized microindentation hardness in HV of the six alloys fabricated in this study. Error bars on the predicted hardness values represent the MAE score of the ML regression model. The error bars on the experimentally measured as-cast and homogenized samples represent the standard deviation. The results of the hardness analysis are included in Table S2, Supporting Information.

observed to display multiple secondary phases, the ML-predicted hardness is significantly higher than that of both the as-cast and homogenised conditions. However some agreement within the range of uncertainty of the experimentally determined values was observed. Furthermore, a marginal increase in hardness was

exhibited from the as-cast to the homogenized state for all three alloys, which was found to be consistent with the observed changes in microstructure. In all cases for alloys D–F, predicted to form single-phase FCC solid solutions by the ML but not CALPHAD, the ML-predicted hardness is observed to be

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
ENGINEERING
MATERIALS**

www.aem-journal.com

significantly lower than the measured hardness in both the as-cast and homogenized form. In addition, a drop in hardness from the as-cast to the homogenized condition is observed in all cases for these alloys. This apparent reduction could be due to a decrease in the intermetallic phase fraction, or morphological changes as the alloys move from a skeletal structure to discrete particles. However, given that the changes are within the measurement error, this was not investigated further.

7. Discussion

7.1. Phase Formation Results Discussion

It is clear from the experimental results, shown in the BSE micrographs of Figures 6 and 7, that all alloys fabricated did not exhibit a single-phase FCC solid solution, with significant intermetallic formation observed. In fact, only four out of the six compositions exhibit an FCC phase, with the other two showing BCC or B2 matrix phases, clearly demonstrating the inaccuracies in the phase prediction of the ML classification models. Alloy B is observed to form the lowest fraction of intermetallic phase and this is also reflected in the confidence measure of the ML phase predictions, scoring one of the highest FCC phase prediction confidences for the downselected alloys. In general, for the alloy systems downselected to be experimentally analyzed, the confidence measure of the ML phase predictions is low. Pushing toward alloy systems with higher hardness was prioritized over alloy systems likely to produce a single-phase FCC solid-solution, but at a significantly lower hardness. If only compositions predicted to form an FCC phase with confidence>90% in both models X and Z are considered, then of the 24 613 predicted to form an FCC phase by the ML, only 443 would satisfy these criteria and all are predicted by CALPHAD to form a single FCC phase. Furthermore, all 443 of these alloys are a compositional variation of CrMnFeCoNi, as expected, as this system dominates the literature with respect to FCC HEAs. In addition, these compositions have a significantly lower hardness predicted by the ML than those downselected for experimental testing, with an average predicted hardness of only 105 HV. Hence, high-hardness-predicted compositions were prioritized at the expense of high confidence in prediction of a single-phase FCC solid-solution to enable further exploration of HEA compositional space. To assist in alloy downselection, equilibrium phase fraction calculations at 1000 °C were performed using CALPHAD and the TCNi8 database on the compositions predicted to form single-phase FCC by both models X and Z. To draw effective comparisons between the experimentally fabricated alloys and CALPHAD, a long duration exposure at 1000 °C and subsequent quench would be required. However, CALPHAD predictions of the same alloy systems as a function of temperature still do not correctly identify the microstructure of the alloy compositions, as described in Table 6, and in most cases predict more complex intermetallic phases that are not observed in the microstructure.

7.2. Hardness Results Discussion

In general, the ML model incorrectly predicts the observed hardness of the experimentally fabricated alloys, although there is some agreement within errors, as seen in Figure 8. Further scrutiny of Figure 8 reveals distinct patterns and systematic errors in the ML hardness predictions. Alloys A–C are found to be systematically overpredicted. In contrast, for alloys D–F, there is significant underprediction of the hardness compared to the experimentally measured values in both the as-cast and homogenized state. The relative overprediction for alloys A–C compared to the underprediction for alloys D–F is less pronounced. In most cases, taking into consideration the measurement errors, the as-cast and homogenized values are comparable, with the notable exception of alloy B, where the precipitation of Ni_3Ti was found to increase the hardness following homogenization. It is encouraging that the hardness regression model Y captures a difference for the dicrete alloy classes from the input features, even though this difference is not captured by the phase model. Nevertheless, the ML did accurately capture the correct trends, with the alloys experimentally determined to display an FCC matrix phase and a secondary intermetallic phase predicted to be a lower hardness than those observed to form BCC and B2 intermetallic-based microstructures.

The disparity between the ML hardness and measured hardness for alloys D–F occurs because these alloys have been downselected as they were predicted to be the highest hardness alloys to form a single-phase FCC solid solution. As highlighted by the SHAP summary plots in Figure 5 and the permutation feature importance plots in Figure 4, all models place a very high importance on *VEC* and models Y and Z place very high importance on $\Delta\varepsilon$. It is expected that $\Delta\varepsilon$ would be related to alloy hardness and indeed the SHAP summary plots show that larger values of $\Delta\varepsilon$ lead to the model predicting higher values of hardness, Figure 5c. Similarly, in model Z, smaller values of $\Delta\varepsilon$ are less likely to predict FCC phase formation, Figure 5b. Furthermore, larger values of VEC are shown to be more likely to predict FCC phase formation and a lower alloy hardness, Figure 5. Hence, although the classification and regression models are not coupled, the model interpretation metrics discussed in Section 4, reveal that similar emphasis is placed on input features in all models. Therefore, it would be expected that if the classification of ML model predicts a single-phase FCC solid solution, then the parallel regression ML model would also predict a lower hardness. Consequently, as the alloys are not in fact single-phase FCC solid solutions and are in some cases not based on an FCC matrix at all, it would be expected that they would display a significantly higher hardness than that predicted by the ML model. Similarly, the maximum single-phase FCC hardness recorded in the training database was 537 HV. Thus, the ML model would not be expected to extrapolate much beyond this maximum value and certainly would not predict hardness values in excess of 650 HV, such as those exhibited experimentally for alloys D–F. The overprediction observed for alloys A–C is in contrast to the observed microstructure of the alloy. It would be expected that as the model predicts single-phase FCC formation, the presence of secondary intermetallic, seen in the BSE micrographs in Figures 6 and 7, would lead to an increase in alloy hardness over the predictions. Instead, the opposite trend is observed.

The combined under- and overprediction of the hardness values from the ML regression model Y, compared to the measured values, are believed to be in part due to the experimental HEA mechanical properties database used in this study to train the

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
ENGINEERING
MATERIALS**

www.aem-journal.com

model. The database contains only 366 values, including values calculated through an empirical relationship with yield strength and imputed values. Of these entries in the database, there are significantly more BCC-forming compositions than FCC, constituting 30.3% and 18.0% of the database respectively. Furthermore, additional analysis of the 66 FCC-labeled compositions revealed that the distribution of hardness values had a mean of 237 HV but a median of 189 HV, a lower quartile of 132 HV, and an upper quartile of 295 HV. In fact, of the 66 FCC-labeled compositions, only 15 were found to be quoted as having a hardness of >400 HV. Additionally, there is an absence of data in the region $320 < HV < 420$. This could potentially indicate that the compositions with higher hardness were mislabeled as single-phase FCC in their respective publications and may have instead exhibited the formation of intermetallic phases.

7.3. Future Research Directions

As demonstrated and discussed above, the accuracy and confidence of predictions of FCC phase formation by the ML models constructed in this study are lacking. In addition, the hardness model overpredicts in the case of alloys A–C and underpredicts for alloys D–F. Therefore, there is significant potential for improvement of the prediction capability. The simplest and most obvious is the availability of additional high quality and quantity data on the microstructural and mechanical properties of HEAs away from the more commonly studied CrMnFeCoNi compositional space. Following this, if the available data are increased, new complex models such as neural networks can be considered for application to this case study and beyond. In this work, more complex models were not considered due to the small amount of available training data being unsuitable for their construction. Additionally, a smaller compositional granularity can be considered for the virtual candidate search space to minimize the space between compositions and more effectively identify phase transitions from the ML outputs. However, this reduced compositional granularity comes at the cost of more time-intensive computation. Finally, in terms of improving the ML model, the feature space can be refined to include new features that better describe both mechanical properties and phase formation of HEAs.

Additionally, for the purpose of alloy discovery, to meet the proposed design case in this study, more alloy systems predicted by this model can be selected for experimental analysis. This will determine their suitability to meet the application design criteria and understand the combined ML and CALPHAD predictions. Despite the erroneous ML predictions and the large uncertainties obtained throughout this study, the alloys that have been experimentally evaluated have resulted in compositions that offer promise for the intended application. Further work on these alloys will seek to evaluate carbon reinforcement to both suppress detrimental brittle intermetallic phase formation and improve mechanical properties, enhancing their applicability to the design case.

8. Conclusions

The study aimed to apply a ML methodology to the design of hardmetal matrix compositions for metal-forming tooling applications. This was simplified to two key design criteria, that of

high hardness and a single-phase FCC structure as a proxy for ductility and toughness. Hence, a series of RF ML models have been constructed and trained from experimentally determined public databases on the phase and hardness of HEAs to enable predictions of these properties in unexplored compositional spaces. In contrast to the majority of literature, CALPHAD was not used to supplement data or as an input to the models. Instead, CALPHAD was utilized to provide further microstructural investigation following the ML and aid in downselection of suitable alloy compositions for experimental analysis. Six compositions were chosen to be fabricated and their microstructural and mechanical properties were investigated to enable assessment of their ability to meet the design criteria and assess the performance of the ML predictions. These six compositions consisted of the three unique alloy systems predicted to be the hardest single-phase FCC by the ML and the three hardest compositions predicted by both the ML and CALPHAD to form a single-phase FCC microstructure.

Interrogation of the ML models constructed for phase and hardness prediction revealed a strong dependency on features such as VEC and $\Delta\varepsilon$ influencing the prediction outcomes, despite the models not being coupled. This was an encouraging result, indicating that the hardness regression model could correctly capture microstructural parameters in subsequent predictions. This was further reinforced by the experimental results obtained. Although none of the alloys that were experimentally evaluated comprised a single-phase FCC structure, three alloys demonstrated an FCC matrix with the remaining alloys relying primarily on a BCC or B2 phase acting as the matrix. This apparent microstructural change was captured by the hardness model that was found to systematically overpredict the hardness of the FCC-matrix alloys and underpredict the BCC/B2-matrix alloys.

In contrast to the satisfactory outcomes of the hardness regression model, the phase prediction classification models were found to be inaccurate compared to the experimental microstructural assessments. The discrepancy between experiment and prediction was believed to be primarily due to the databases used in the ML training. The imbalanced data contained within the databases, coupled with the bias of FCC compositions toward the CrMnFeCoNi system, resulted in predictions herein having reduced confidence indicators when downselected with increased hardness being prioritized.

However, despite the experimental and predicted discrepancies, the methodology employed identified compositions that are suitable for further experimental evaluation toward the intended use case. Furthermore, the construction of the models and their rigorous analysis have revealed several areas for improvement for both the ML architectures as well as highlighting the need for reliable, extensive, and expansive databases.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data and code that support the findings of this study are openly available at https://doi.org/10.15131/shef.data.25233514, https://doi.org/10.15131/shef.data.25250770.v1, and respectively.

Keywords

[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, *Adv. Eng. Mater.* **2004**, *6*, 299.

[2] B. Cantor, I. T. H. Chang, P. Knight, A. J. B. Vincent, *Mater. Sci. Eng. A, Struct. Mater.* **2004**, *375–377*, 213.

[3] E. P. George, D. Raabe, R. O. Ritchie, *Nat. Rev. Mater.* **2019**, *4*, 515.

[4] S. Akrami, P. Edalati, M. Fuji, K. Edalati, *Mater. Sci. Eng.: R: Rep.* **2021**, *146*, 100644.

[5] F. He, Z. Wang, Q. Wu, S. Niu, J. Li, J. Wang, C. T. Liu, *Scr. Mater.* **2017**, *131*, 42.

[6] D. B. Miracle, *Mater. Sci. Technol.* **2015**, *31*, 1142.

[7] E. Pickering, N. Jones, *Int. Mater. Rev.* **2016**, *61*, 183.

[8] D. B. Miracle, *JOM* **2017**, *69*, 2130.

[9] D. B. Miracle, O. N. Senkov, *Acta Mater.* **2017**, *122*, 448.

[10] K. Biswas, J.-W. Yeh, P. P. Bhattacharjee, J. T. M. DeHosson, *Scr. Mater.* **2020**, *188*, 54.

[11] M. C. Troparevsky, J. R. Morris, P. R. C. Kent, A. R. Lupini, G. M. Stocks, *Phys. Rev. X* **2015**, *5*, 011041.

[12] B. Gwalani, D. Choudhuri, K. Liu, J. T. Lloyd, R. S. Mishra, R. Banerjee, *Mater. Sci. Eng. A, Struct. Mater.* **2020**, *771*, 138620.

[13] M. Laurent-Brocq, L. Perrière, R. Pirès, Y. Champion, *Mater. Des.* **2016**, *103*, 84.

[14] W. Ren, Y. F. Zhang, W. L. Wang, S. J. Ding, N. Li, *Mater. Des.* **2016**, *235*, 112454.

[15] A. M. Manzoni, U. Glatzel, *Mater. Charact.* **2019**, *144*, 512.

[16] J. Joseph, M. Senadeera, Q. Chao, K. F. Shamlaye, S. Rana, S. Gupta, S. Venkatesh, P. Hodgson, M. Barnett, D. Fabijanic, *J. Alloys Compd.* **2021**, *888*, 161496.

[17] Y.-T. Chen, Y. J. Chang, H. Murakami, S. Gorsse, A. C. Yeh, *Scr. Mater.* **2020**, *187*, 177.

[18] A. Shafiei, *Metall. Mater. Trans. A: Phys.* **2022**, *53*, 4349.

[19] Y. Lu, Y. Dong, H. Jiang, Z. Wang, Z. Cao, S. Guo, T. Wang, T. Li, P. K. Liaw, *Scr. Mater.* **2020**, *187*, 202.

[20] G. L. W. Hart, T. Mueller, C. Toher, S. Curtarolo, *Nat. Rev. Mater.* **2021**, *6*, 730.

[21] W. Huang, P. Martin, H. L. Zhuang, *Acta Mater.* **2019**, *169*, 225.

[22] Y. Zeng, M. Man, K. Bai, Y.-W. Zhang, *Mater. Des.* **2021**, *202*, 109532.

[23] Y. Yan, D. Lu, K. Wang, *Comput. Mater. Sci.* **2021**, *199*, 110723.

[24] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* **2022**.

[25] M. Calvo-Dahlborg, S. Mehraban, N. P. Lavery, S. G. R. Brown, J. Cornide, J. Cullen, J. Cieslak, Z. Leong, R. Goodall, U. Dahlborg, *J. Alloys Compd.* **2021**, *865*, 158799.

[26] C. Wen, Y. Zhang, C. Wang, D. Xue, Y. Bai, S. Antonov, L. Dai, T. Lookman, Y. Su, *Acta Mater.* **2019**, *170*, 109.

[27] A. Debnath, A. M. Krajewski, H. Sun, S. Lin, M. Ahn, W. Li, S. Priya, J. Singh, S. Shang, A. M. Beese, Z.-K. Liu, W. F. Reinhart, *J. Mater. Inf.* **2021**, *1*, 3.

[28] R. Machaka, *Comput. Mater. Sci.* **2021**, *188*, 110244.

[29] D. Q. Zhao, S. P. Pan, Y. Zhang, P. K. Liaw, J. W. Qiao, *Appl. Phys. Lett.* **2021**, *118*, 231904.

[30] Y. V. Krishna, U. K. Jaiswal, M. R. Rahul, *Scr. Mater.* **2021**, *197*, 113804.

[31] K. Kaufmann, K. S. Vecchio, *Acta Mater.* **2020**, *198*, 178.

[32] U. Bhandari, C. Zhang, C. Zeng, S. Guo, A. Adhikari, S. Yang, *Crystals* **2021**, *11*, 1.

[33] C. Wen, C. Wang, Y. Zhang, S. Antonov, D. Xue, T. Lookman, Y. Su, *Acta Mater.* **2021**, *212*, 116917.

[34] F. Liu, X. Xiao, L. Huang, L. Tan, Y. Liu, *Mater. Today Commun.* **2022**, *30*, 103172.

[35] Y. Sun, Z. Lu, X. Liu, Q. Du, H. Xie, J. Lv, R. Song, Y. Wu, H. Wang, S. Jiang, Z. Lu, *Appl. Phys. Lett.* **2021**, *119*, 201905.

[36] C. Zhang, Y. Yang, *MRS Bull.* **2022**, *47*, 158.

[37] E. Prieto, A. Vaz-Romero, J. Gonzalez-Julian, S. Guo, P. Alvaredo, *Int. J. Refract. Met. Hard Mater.* **2021**, *99*, 105592.

[38] E. P. George, W. A. Curtin, C. C. Tasan, *Acta Mater.* **2020**, *188*, 435.

[39] X. Liu, J. Zhang, Z. Pei, *Prog. Mater. Sci.* **2023**, *131*, 101018.

[40] X. Liu, P. Xu, J. Zhao, W. Lu, M. Li, G. Wang, *J. Alloys Compd.* **2022**, *921*, 165984.

[41] R. Arróyave, *J. Phase Equilib. Diffus.* **2022**, *43*, 606.

[42] C. Yang, C. Ren, Y. Jia, G. Wang, M. Li, W. Lu, *Acta Mater.* **2022**, *222*, 117431.

[43] D. Beniwal, P. Singh, S. Gupta, M. J. Kramer, D. D. Johnson, P. K. Ray, *npj Comput. Mater.* **2022**, *8*.

[44] S. Gorsse, M. H. Nguyen, O. N. Senkov, D. B. Miracle, *Data Brief* **2018**, *21*, 2664.

[45] X. Huang, C. Jin, C. Zhang, H. Zhang, H. Fu, *Mater. Des.* **2021**, *211*, 110177.

[46] R. Jain, U. Lee, S. Samal, N. Park, *J. Alloys Compd.* **2023**, *956*, 170193.

[47] L. Shen, L. Chen, J. Huang, J. He, Z. Li, J. Pan, F. Chang, P. Dai, Q. Tang, *Intermetallics* **2023**, *162*, 108030.

[48] S. Gorsse, M. H. Nguyen, O. N. Senkov, D. B. Miracle, *Data Brief* **2020**, *32*, 106216.

[49] R. Machaka, G. T. Motsi, L. M. Raganya, P. M. Radingoana, S. Chikosha, *Data Brief* **2021**, *38*, 107346.

[50] I. Toda-Caraballo, P. E. J. Rivera-Díaz-del-Castillo, *Acta Mater.* **2015**, *85*, 14.

[51] S. Guo, C. Ng, J. Lu, C. Liu, *J. Appl. Phys.* **2011**, *109*.

[52] A. Y.-T. Wang, R. J. Murdock, S. K. Kauwe, A. O. Oliynyk, A. Gurlo, J. Brgoch, K. A. Persson, T. D. Sparks, *Chem. Mater.* **2020**, *32*, 4954.

[53] A. S. Bundela, M. R. Rahul, *J. Alloys Compd.* **2022**, *908*.

[54] Y. Zhang, C. Wen, C. Wang, S. Antonov, D. Xue, Y. Bai, Y. Su, *Acta Mater.* **2020**, *185*, 528.

[55] G. Pilania, *Comput. Mater. Sci.* **2021**, *193*, 110360.

[56] A. Choudhury, T. Konnur, P. P. Chattopadhyay, S. Pal, *Eng. Comput.* **2020**, *37*, 1003.

[57] R. Caruana, A. Niculescu-Mizil, in *Proc. of the 23rd Int. Conf, on Machine Learning - ICML '06 2006* **2006**, pp. 161–168.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
ENGINEERING
MATERIALS**

www.aem-journal.com

[58] L. Breiman, *Mach. Learn.* **2001**, *45*, 5.

[59] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825.

[60] N. Islam, W. Huang, H. L. Zhuang, *Comput. Mater. Sci.* **2018**, *150*, 230.

[61] H. Dalianis, *Clinical Text Mining: Secondary use of Electronic Patient Records* (Ed: H. Dalianis), Springer International Publishing, Cham **2018**, pp. 45–53.

[62] B. Akhil, A. Bajpai, N. P. Gurao, K. Biswas, *Modell. Simul. Mater. Sci. Eng.* **2021**, *29*, 85005.

[63] M. J. Campbell, *Statistics at Square One*, Wiley Blackwell, Hoboken, NJ **2021**.

[64] C. Chen, L. Ma, Y. Zhang, P. K. Liaw, J. Ren, *Intermetallics* **2023**, *154*, 107819.

[65] S. Risal, W. Zhu, P. Guillen, L. Sun, *Comput. Mater. Sci.* **2021**, *192*.

[66] L. Zhang, H. Chen, X. Tao, H. Cai, J. Liu, Y. Ouyang, Q. Peng, Y. Du, *Mater. Des.* **2020**, *193*.

[67] J. J. Gilman, *Electronic Basis of The Strength of Materials*, Cambridge University Press **2003**.

[68] W. Hume-Rothery, *Elements of Structural Metallurgy*, Institute of Metals, London **1961**.

[69] W. D. Callister, D. G. Rethwisch, *Materials Science and Engineering*, Wiley, Hoboken, NJ **2015**.

[70] S. M. Lundberg, S.-I. Lee, in *Proc. of the 31st Int. Conf. on Neural Information Processing Systems*, Curran Associates Inc., Long Beach, CA **2017**, pp. 4768–4777.

[71] B. Cao, S. Yang, A. Sun, Z. Dong, T.-Y. Zhang, *J. Mater. Inf.* **2022**, *2*, 4.

[72] S. A. Kube, S. Sohn, D. Uhl, A. Datye, A. Mehta, J. Schroers, *Acta Mater.* **2019**, *166*, 677.

[73] G. Pawley, *J. Appl. Crystallogr.* **1981**, *14*, 357.

[74] ASTM **1984**.

[75] Z. Wang, Y. Huang, Y. Yang, J. Wang, C. T. Liu, *Scr. Mater.* **2015**, *94*, 28.

[76] F. Tian, L. K. Varga, N. Chen, J. Shen, L. Vitos, *Intermetallics* **2015**, *58*, 1.

[77] S. Fang, X. Xiao, L. Xia, W. Li, Y. Dong, *J. Non-Cryst. Solids* **2003**, *321*, 120.

[78] U. Mizutani, H. Sato, *Crystals* **2017**, *7*, 9.

[79] A. Takeuchi, A. Inoue, *Mater. Trans.* **2005**, *46*, 2817.

[80] S. Guo, C. T. Liu, *Prog. Nat. Sci.: Mater. Int.* **2011**, *21*, 433.

[81] Y. Zhang, Y. J. Zhou, J. P. Lin, G. L. Chen, P. K. Liaw, *Adv. Eng. Mater* **2008**, *10*, 534.

[82] X. Yang, Y. Zhang, *Mater. Chem. Phys.* **2012**, *132*, 233.

[83] J. M. Rickman, G. Balasubramanian, C. J. Marvel, H. M. Chan, M. T. Burton, *J. Appl. Phys.* **2020**, *128*.