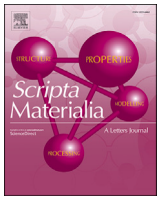
Scripta Materialia 228 (2023) 115335

Contents lists available at ScienceDirect

Scripta Materialia

journal homepage: www.journals.elsevier.com/scripta-materialia



Accelerated alloy discovery using synthetic data generation and data mining ☆

Rangasayee Kannan [a,*], Peeyush Nandwana [b]

[a] Manufacturing Science Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[b] Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

ARTICLE INFO

Keywords:
Alloy design
Data Mining
Synthetic Data
Machine Learning
CALPHAD

ABSTRACT

The search for new alloys with improved properties is never ending with infinite combinations and amounts of alloying elements in the alloy. Advancements in machine learning have made navigating this enormous search space feasible. However, training the machine learning models and tuning their hyper-parameters to make accurate predictions can be time-consuming and often require high-performance computing resources. Furthermore, the quality of the predictions depend on the availability of sufficient training data. Here, we present a generic approach to accelerate alloy discovery by coupling high throughput CALPHAD calculations, synthetic data generation, and data mining. As a demonstration of the approach, we design super bainitic steels that form bainite at 200°C in lower transformation times.

Bronze age marked the beginning of alloy design for humans when copper and tin were melted together to fabricate weapons and agricultural tools [1]. Traditionally, alloy design relied on human intuition, involving learning patterns on the role of each alloying element. With the knowledge and the development of thermodynamics, alloy design transitioned to a science based approach with the development of multiple alloys for engineering applications. Most of the alloys have a single principle element with additions of other elements to achieve specific properties. However, recently, complex concentrated alloys/high entropy alloys (HEA) with multiple elements in near equiatomic compositions are being used with superior properties and performance. Optimizing the composition of each element is challenging since there are infinite number of alloying element combinations, and fabricating and testing each of them is a cumbersome task. The complexity of the problem is further exacerbated with the increase in number of alloying elements in HEA's which have a wider search space, with at least 10^{50} alloy variants [1]. While the recent developments in computational thermodynamic tools utilizing CALPHAD have helped to a certain extent [2], considering the vast composition space, conducting calculations can be computationally costly and time consuming, often requiring super-computing resources.

Advancements in machine learning have enabled rapid development of new alloys for targeted applications. Both, a forward and an inverse design approach have been used to design alloys using machine learning [1,3–9]. The forward approach involves training machine learning models to predict properties using composition as input [7–9] and the inverse design approach involves training the machine learning model, usually a deep neural network (DNN) to predict the composition with the required properties as input [1,5,6], also known as Generative Deep Learning. Both these approaches have several hyper-parameters, which need to be tuned for optimal convergence of the predictions. The computational time required to tune these hyper-parameters increases with the complexity of the data. A major drawback of such approaches is the availability of sufficient training data to train the machine learning models, that affects the accuracy of the predictions. This can be challenging for alloy design, especially with the limited data available in the open literature. To overcome the challenge of limited available data, either high-throughput computational thermodynamics can be used to generate training data, and/or experiments

* Corresponding author.
E-mail addresses: kannanr1@ornl.gov (R. Kannan), nandwanap@ornl.gov (P. Nandwana).

https://doi.org/10.1016/j.scriptamat.2023.115335
Received 15 November 2022; Received in revised form 5 January 2023; Accepted 28 January 2023
Available online 6 February 2023

can be performed to validate the predicted data, which can in turn provide feedback to the training data like the active learning approach used for materials design proposed recently [1]. However, both these approaches can be time-consuming, which reduces the discovery rate of new alloys.

Synthetic data generated by algorithms which learn the statistical trends in the real data have been used in various fields like traffic monitoring, healthcare, fraud detection, and to detect defects in steel castings to validate mathematical models and to act as training data for machine learning models [10–15]. Apart from being used as training data for machine learning models, the amount of synthetic data generated is at the discretion of the user, is faster, and allows for exploring data spaces which can be missed by real data. All the above mentioned features of synthetic data are ideal to explore novel alloy compositions in an accelerated manner. To the best of the authors' knowledge, synthetic data has not been used as a tool for alloy design.

Here, we propose an approach to use high throughput CALPHAD data to generate synthetic data, and use data mining tools on the synthetic data to explore a wide range of compositions to design novel alloys. For example, we use the proposed approach to design nanostructured bainitic steels. Nanostructured bainitic steels typically have strength in the range 1250 MPa – 2500 MPa, elongation in the range of 5% – 30%, fracture toughness in the range of 30 – 50 $MPam^{1/2}$, endurance limit of 850 MPa – 1050 MPa, good resistance to tempering/softening up to a temperature of 350°C – 500°C, and superior impulse loading properties [16]. The properties of the nanostructured bainitic steels are comparable to maraging steels and tool steels [17,18], however, the alloy cost (USD/ton) of nano-bainitic steels is 90 times cheaper than their equivalent maraging steels [17], making them ideal candidates for high-volume applications like tooling and cutting blades, which require high hardness, wear resistance, and tempering/softening resistance at service temperatures. Though such properties are attractive, bulk nanostructured steels have not yet been fully commercialized owing to the long isothermal holding time at temperature above martensite start temperature required to produce the nanostructure, which is of the order of days (2-60 days) and has its associated costs [19]. This remains a major drawback for the implementation of nanostructured bainitic steels in a wider range of markets. Though the transformation time can be reduced by the addition of Al and/or Co (up to 3-4 wt.% combined) [20], the minimum reported transformation time is around 10 hours, which is still long compared to the manufacturing time of industrial components, which is around 4-6 hours [21]. Designing new alloys with reduced transformation time can open the application space for nanostructured steels, and we demonstrate the proposed accelerated alloy design approach to increase the discovery rate of nanostructured bainitic steels.

Fig. 1 shows the schematic of the proposed workflow. Firstly, input data either generated using high throughput CALPHAD, experimental measurements or from open literature is used to generate synthetic data. Once synthetic data has been generated, and the required compositions were identified by using data mining, the synthetic data was validated. The synthetic data can be validated using one of the three available sources with increasing validation accuracy and decreasing quantity of data to validate against. Firstly, select compositions or properties from the generated synthetic data can be validated by running CALPHAD on the synthetically generated compositions. This method can validate many synthetically generated compositions, providing better validation statistics. Secondly, the synthetically generated compositions can be validated by conducting experiments which can be resource intensive and therefore, encompass a significantly narrow range. Finally, the synthetically generated compositions can be validated by data published in the open literature which can be limited in compositions as well as pedigree. Therefore, validating the synthetic data against experimental measurements and data published in open literature can produce a highly accurate validation score but can only validate a much narrower subset of the compositions. Here we demonstrate the use of CALPHAD computations (current case study) and experimental/literature-reported measurements (refer to Supplementary) to validate the synthetic data. Finally, once a validated synthetic data generator was developed, alloys meeting the required properties were mined and visualized using a data mining approach to generate an ensemble of new alloys.

In the current case study to design nanostructured bainitic steels, input data was generated using high throughput CALPHAD technique. Firstly, the composition range of each element from the list of all the reported nanostructured bainitic steels [16] was tabulated as shown in Table 1.

Based on the composition range in Table 1, 5000 hypothetical compositions within the range were generated using the property model module of ThermoCalc. The amount (in Wt. %) of each element was varied to follow a Gaussian distribution such that the mean was the average amount of each element and the tails of the Gaussian distribution were the maximum and minimum amount of each element. The possible combinations of variations in the alloy were controlled depending on the number of sample sets needed. Detailed procedure can be found in our article in Ref. [22]. With the generated compositions, martensite start temperature (M_s) and Vickers hardness (HV) were calculated using the TC Python module of ThermoCalc [2,23].[1] M_s was calculated to down select compositions with M_s below 200 °C so that martensite does not form during bainitic transformation at 200 °C. To determine the kinetics of bainite formation, the time to start bainite transformation at 200 °C (t_{200}) was calculated using MUCG83[2] code originally developed in FORTRAN [25,26] and later re-written in Python [27] for all the 5000 compositions. Fig. 2 summarizes the input data on 5,000 compositions generated from high throughput CALPHAD. It can be seen from the parallel coordinate plot visualization in Fig. 2a that majority of the compositions have a bainite start time less than 200,000 seconds. Some of the compositions have a bainite start time greater than 500,000 seconds. It can be seen that the martensite start temperature of all the compositions falls in the range −31°C - 198°C indicating that bainite can form at 200°C without the formation of martensite. The hardness for these compositions have a wide range from 550 - 720 HV. Correlation matrix of the input 5000 data in Fig. 2b shows that the time to form bainite has a strong negative correlation with Al and Co content in the alloy which agrees with the data in the literature [20]. The purpose of the correlation matrix is to validate if the synthetic data generator has learned the correlations between different elements and the bainite start at 200°C, martensite start temperature (M_s), and hardness (HV) in the real data while generating the synthetic datasets.

Following the generation of the real data input using high throughput CALPHAD, the FAST_ML preset [28] of synthetic data vault package [29,30] which optimizes for fast modeling and sampling time while also taking advantage of basic, machine learning techniques to learn from the real data was used to generate synthetic data. The FAST_ML preset

[1] ThermoCalc calculates the M_s temperature based on the calculation of thermodynamic the barrier required for diffusionless transformation of austenite to martensite and compares that with the available driving force for austenite transformation, and M_s is defined as the temperature at which the available driving force is larger than the thermodynamic barrier for the diffusionless transformation. For hardness measurements, ThermoCalc considers contributions from intrinsic strength of the pure elements, grain boundary strengthening, solid solution strengthening, and precipitation strengthening. For this particular case study precipitation strengthening was not considered, and the obtained yield strength was reported as the Vickers hardness (HV). More details can be found in Ref. [24].

[2] MUCG83 is a thermodynamic tool for modeling the thermodynamics and kinetics of solid-state phase transformations in steels. The tool considers the majority of the important alloying elements in bainitic steels like C, Si, Mn, Ni, Mo, Cr, V, Co, Cu, Al and W. The tool calculates the driving force for allotriomorphic ferrite and diffusionless ferrite formation from austenite by calculating the T_o and T_o + 400 J/mol temperatures enabling the construction of TTT curve. Detailed explanation and applicability of the code can be found in Refs. [25,26].

Table 1
Composition range of elements in nanostructured bainitic steels reported in literature.

Element	C	Si	Mn	Mo	Cr	V	Al	Co	Ni
Min. (Wt. %)	0.7	1.5	0.1	0.04	0.26	0.1	0	0.14	0.16
Max. (Wt. %)	1.15	2.9	3.76	1.45	2.3	1.58	3	3.87	4.12

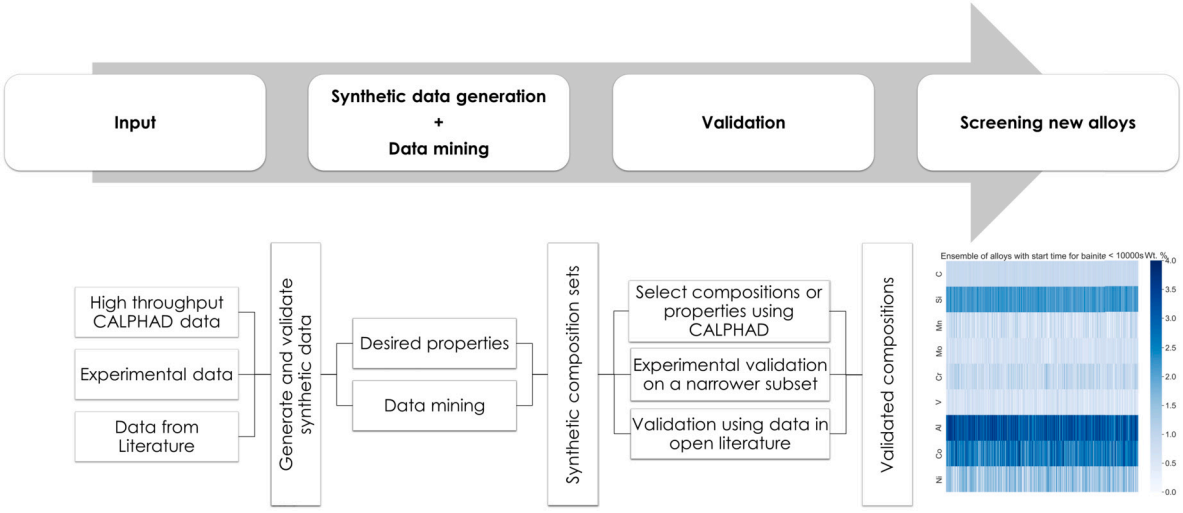


Fig. 1. Schematic showing the workflow for accelerated alloy discovery.

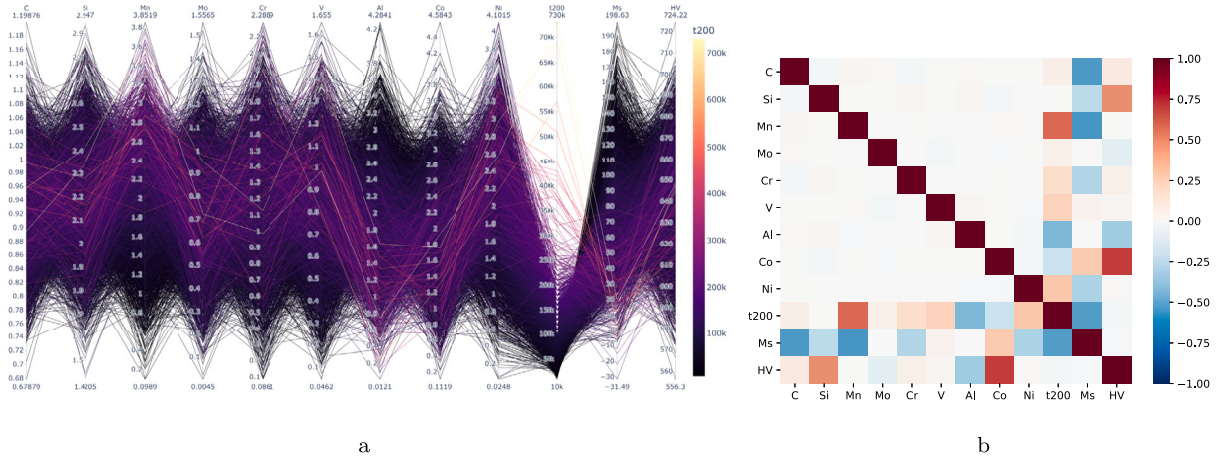


Fig. 2. Input data on 5000 compositions generated using high throughput CALPHAD visualized using parallel coordinate plot. The correlations between individual compositions and the time to bainite start (t200), martensite start temperature (M_s), and hardness (HV) are shown using the Pearson correlation matrix. Interactive version of the parallel coordinate plot is available in the Supplementary material.

of the synthetic data vault package utilized Gaussian Copula machine learning model to learn from the input data while adhering to the basic statistical properties like the min/max values, average, and standard deviation of the real input data. The FAST_ML synthetic data generator was used to generate 1,000,000 synthetic composition combinations. Fig. 3a shows the visualization of the generated 1,000,000 synthetic data using parallel coordinate plots. It can be seen that upon visualizing the synthetic data, the generated synthetic data covers composition spaces not represented by the real data input. Since the parallel coordinate plots are color coded with respect to the time to start bainite transformation, a clear correlation between the time to start bainitic transformation and Mn, Co, and Al content is observed unlike the real data which only has 5000 composition sets. It should also be noted that the generated synthetic data has the maximum and minimum bounds of composition similar to the real data input. To test if the generated synthetic data captures the inter-variable correlations, correlation matrices were plotted for the synthetic data in Fig. 3b. It can be seen that the synthetic data generator has learned the inter-variable correlations, with the correlation matrix of the generated synthetic data similar to the real data input (Refer Fig. 2b). Further metrics regarding the quality of the generated synthetic data can be found in the Supplementary material. To validate the transformation time for bainite start at 200°C, 1,000,000 compositions generated by synthetic data were fed into the MUCG83 code to calculate the bainite start temperature at 200°C. The results from the synthetic data generator are compared to the MUCG83 predictions in Fig. 3c, and it can be seen that the synthetic data generator predicts the transformation time fairly well with an R^2 of 0.9671, indicating that the synthetic data generator the synthetic data generator accurately generates the compositions close to the real data without the need for thermodynamic calculations. To understand the quality of the generated synthetic data at the extremes of transformation time, probability plot for transformation time for bainite at 200°C was plot-

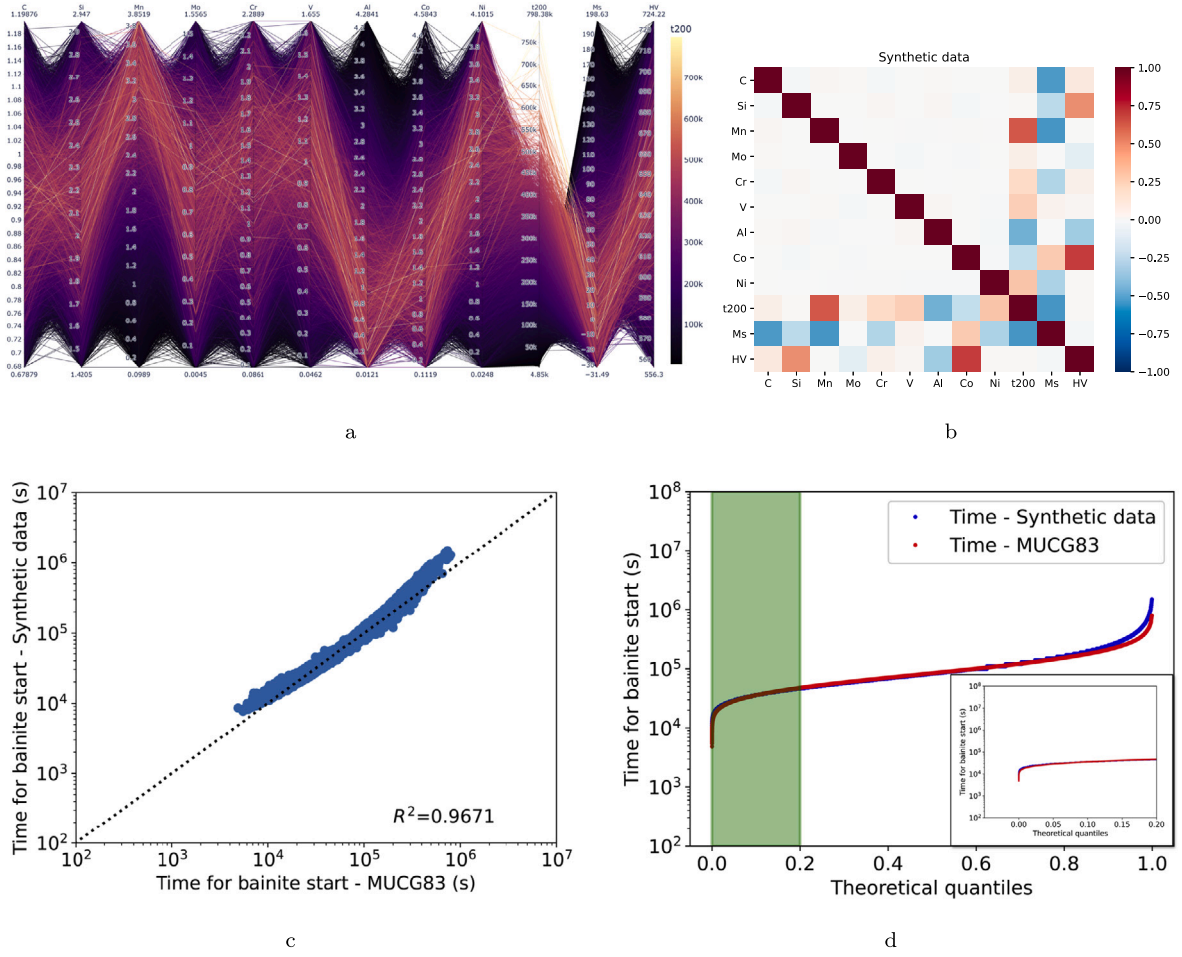


Fig. 3. (a) Synthetic data containing 1,000,000 compositions generated by the synthetic data generator visualized using parallel coordinate plot. (b) The correlations between individual compositions and the time to bainite start (t200), martensite start temperature (M_s), and hardness (HV) for the synthetic data are compared with the real data generated in Fig. 2 using the Pearson correlation matrix. (c) Parity plot of time to start bainite at 200°C for synthetic data and the real data calculated using MUCG83, (d) Probability plot for time to start bainite at 200°C for synthetic data and real data. Interactive version of the parallel coordinate plot is available in the Supplementary material.

ted for the real data and the generated synthetic data. It can be seen that the predictions from the synthetic data agree well with the real data from MUCG83 even in the extremes of transformation time, especially for lower bainite transformation time, as shown in the inset for lower quantiles (Refer Fig. 3d).

Finally the validated synthetic data was mined to identify compositions which generate bainite at 200°C with time to transformation less than 10,000 s. 10,000 s was used as a baseline since the reported nanostructured bainitic alloys have time to bainite start greater than 10,000 s at 200°C [16]. Fig. 4 shows the results after data mining of the synthetic data where the alloys are color coded with respect to the time to start bainite transformation using a diverging color scheme centered at 10,000 s. It can be seen that the alloys reported in literate [16] have a start time for bainite transformation > 10,000 s at 200°C . Upon mining the synthetically generated compositions, 535 nanostructured bainitic compositions with a transformation time <10,000 s at 200°C can be designed by using the current approach (Refer to the inset in Fig. 4 a). The composition of the generated alloys are represented by a heat map in Fig. 4b, where each column represents an alloy and the intensity of the color represents the wt. % of the specific element in the alloy. The alloys in Fig. 4b have been ordered based on the time for bainite transformation at 200°C, which is shown by a separate color bar, where it can be seen that the designed alloys form bainite at a time lower than existing alloys with some of them forming bainite in as low as 5000 s. This indicates that the alloy design space is wide open to design and fabricate new nanostructured bainitic steel alloys which have a lower isothermal hold time to generate the nanostructure. The proposed approach can also design alloys meeting more than one required property simultaneously. For example based on the correlation maps presented in Fig. 2, it can be seen that a lower time to start bainite transformation (t_{200}) results in a higher M_s temperature. Therefore, it is crucial to design alloys which have a lower t_{200} and at the same time have a low enough M_s temperature so as to avoid the embrittling martensite formation. From Fig. 3, upon generating the synthetic data, it can be seen from the color coded parallel coordinate plots that alloys with a lower t_{200} generally tend to have a lower overall hardness. Thus, an ideal nanostructured bainitic alloy would be the one with low t_{200}, low M_s, and a high hardness. By providing these conditions to the data mining step, alloys can be designed meeting three conditions concurrently. Fig. 4c shows one such example of alloys having a transformation time to form bainite <10,000 s, M_s < 150°C, and hardness >650 HV. The proposed approach is generic and can be applied to other alloy systems, which have a wider search space, for example, multi-component high entropy alloys (HEA's) which have at least 10^{50} alloy variants. Applicability of this approach to HEA's is shown in the Supplementary material, where we used the current approach to design alloy compositions with low thermal expansion coefficients and validated the data with predicted and experimental results by Rao et al. [1].

In summary, an accelerated alloy design approach is presented to discover new alloys meeting multiple properties at the same time. Com-

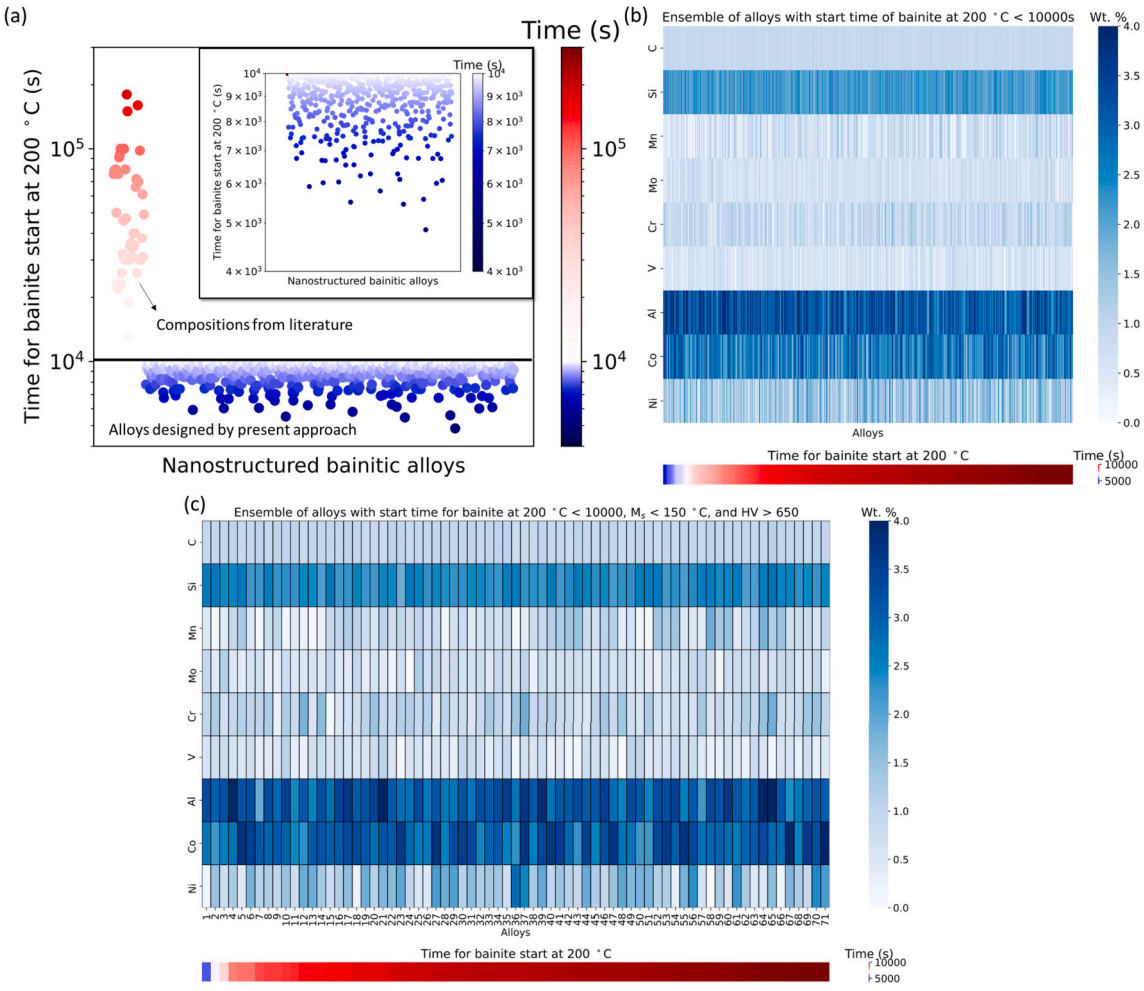


Fig. 4. (a) Transformation time of nanostructured bainitic alloys reported in literature and alloys designed by current approach, (b) the ensemble of alloy compositions designed by current approach with transformation time to bainite start less than 10,000s, and (c) Ensemble of alloys meeting multiple conditions of t200<10,000 s, M_s < 150°C, and hardness >650 HV. In (b) and (c) each column represents an alloy, and each cell in the column represents the composition of the element in Wt. %. The alloys are sorted based on the transformation time for bainite formation at 200°C (shown in separate color bar). Interactive versions of the composition map for (b) and (c) are available in the Supplementary material. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

pared to the existing alloy design strategies using CALPHAD or machine learning methods involving generative adversarial networks (GAN), the proposed approach is accelerated. Firstly, conducting high throughput CALPHAD over a wide search space like 1,000,000 compositions can be computationally expensive, whereas the proposed method which generates synthetic data based on the initial input from CALPHAD accelerates the alloy search process. Secondly, inverse design using GAN can be difficult to train since the competing neural networks in the generator and discriminator may fail to converge if proper hyperparameters are not chosen for both the generator and discriminator [31,32]. Compared to training GAN, generating synthetic data is faster and the entire process of setting up the synthetic data generator and mining the compositions can be performed in a few hours using a commercial consumer-use computer without custom hardware upgrades. The proposed approach, therefore, can be used in industrial and academic settings to guide the design of new alloy compositions with improved properties. As a case study, we used the proposed approach to design new nanostructured bainitic steels with reduced transformation time for bainite transformation and also validated it for concentrated alloys like high entropy alloys. This demonstrates that the approach can be used for alloy design of dilute as well as concentrated alloys.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Research was performed at the U.S. Department of Energy's Manufacturing Demonstration Facility, located at Oak Ridge National Laboratory. This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. Research was co-sponsored by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy, Advanced Manufacturing Office.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.scriptamat.2023.115335.

References

[1] Z. Rao, P. Tung, R. Xie, Y. Wei, H. Zhang, A. Ferrari, et al., Machine learning-enabled high-entropy alloy discovery, Science (2022), https://doi.org/10.1126/SCIENCE.ABO4940/SUPPL_FILE/SCIENCE.ABO4940_SM.PDF, arXiv:2202.13753.

[2] J.O. Andersson, T. Helander, L. Höglund, P. Shi, B. Sundman, Thermo-Calc & DICTRA, computational tools for materials science, Calphad 26 (2) (2002) 273–312, https://doi.org/10.1016/S0364-5916(02)00037-8.

[3] Q. Wu, Y. Jia, Z. Wang, F. He, Y. Wei, J. Li, et al., Rapid alloy design from superior eutectic high-entropy alloys, Scr. Mater. 219 (2022) 114875, https://doi.org/10.1016/J.SCRIPTAMAT.2022.114875.

[4] Y.V. Krishna, U.K. Jaiswal, R.M. R, Machine learning approach to predict new multiphase high entropy alloys, Scr. Mater. 197 (2021) 113804, https://doi.org/10.1016/J.SCRIPTAMAT.2021.113804.

[5] A. Debnath, A.M. Krajewski, H. Sun, S. Lin, M. Ahn, W. Li, et al., Generative deep learning as a tool for inverse design of high entropy refractory alloys, J. Mater. Inform. 1 (1) (2021) 3, https://doi.org/10.20517/JMI.2021.05, arXiv:2108.12019, https://jmijournal.com/article/view/4294.

[6] Q.M. Hu, R. Yang, The endless search for better alloys, Science (New York, NY) 378 (6615) (2022) 26–27, https://doi.org/10.1126/SCIENCE.ADE5503/ASSET/7AEE98D0-A56D-4E82-8F07-DCC0EABD0A24/ASSETS/GRAPHIC/SCIENCE.ADE5503-F1.SVG, https://www.science.org/doi/10.1126/science.ade5503.

[7] V.S. Goud, R.M. R, G. Phanikumar, Prediction of growth velocity of undercooled multicomponent metallic alloys using a machine learning approach, Scr. Mater. 207 (2022) 114309, https://doi.org/10.1016/J.SCRIPTAMAT.2021.114309.

[8] W. Guan, Y. Zhao, Y. Liu, S. Kang, D. Wang, L. Cui, Force data-driven machine learning for defects in friction stir welding, Scr. Mater. 217 (2022) 114765, https://doi.org/10.1016/J.SCRIPTAMAT.2022.114765.

[9] X. Jiang, B. Jia, G. Zhang, C. Zhang, X. Wang, R. Zhang, et al., A strategy combining machine learning and multiscale calculation to predict tensile strength for pearlitic steel wires with industrial data, Scr. Mater. 186 (2020) 272–277, https://doi.org/10.1016/J.SCRIPTAMAT.2020.03.064.

[10] G. Soltana, M. Sabetzadeh, L.C. Briand, Synthetic data generation for statistical testing, in: ASE 2017 - Proceedings of the 32nd IEEE/ACM International Conference on Automated Software Engineering, 2017, pp. 872–882.

[11] J.W. Anderson, K.E. Kennedy, L.B. Ngo, A. Luckow, A.W. Apon, Synthetic data generation for the internet of things, in: Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014, 2014, pp. 171–176.

[12] L. Zhang, A. Gonzalez-Garcia, J. Van De Weijer, M. Danelljan, F.S. Khan, Synthetic data generation for end-to-end thermal infrared tracking, IEEE Trans. Image Process. 28 (4) (2019) 1837–1850, https://doi.org/10.1109/TIP.2018.2879249, arXiv:1806.01013.

[13] A. Boikov, V. Payor, R. Savelev, A. Kolesnikov, Synthetic data generation for steel defect detection and classification using deep learning, Symmetry 13 (7) (2021) 1176, https://doi.org/10.3390/SYM13071176, https://www.mdpi.com/2073-8994/13/7/1176/htm.

[14] H.P. Das, R. Tran, J. Singh, X. Yue, G. Tison, A. Sangiovanni-Vincentelli, et al., Conditional synthetic data generation for robust machine learning applications with limited pandemic data, Proc. AAAI Conf. Artif. Intell. 36 (11) (2022) 11792–11800, https://doi.org/10.1609/AAAI, arXiv:2109.06486, https://ojs.aaai.org/index.php/AAAI/article/view/21435.

[15] J. Dahmen, D. Cook, SynSys: a synthetic data generation system for healthcare applications, Sensors 19 (5) (2019) 1181, https://doi.org/10.3390/S19051181, https://www.mdpi.com/1424-8220/19/5/1181/htm.

[16] H. Bhadeshia, Bainite in steels: theory and practice, 3 ed., https://books.google.com/books?hl=en%26lr=%26id=SwqWDwAAQBAJ%26oi=fnd%26pg=PP1%26dq=bainite+in+steels%26ots=nHUE1Ljram%26sig=m-ue8mJyJhRO8sJo2GtU9XdxzSU%23v=onepage%26q=bainiteinsteels%26f=false, 2019.

[17] F.G. Caballero, H.K. Bhadeshia, Very strong bainite, Curr. Opin. Solid State Mater. Sci. 8 (3–4) (2004) 251–257, https://doi.org/10.1016/J.COSSMS.2004.09.005.

[18] A. Kwiatkowski da Silva, I.R. Souza Filho, W. Lu, K.D. Zilnyk, M.F. Hupalo, L.M. Alves, et al., A sustainable ultra-high strength Fe18Mn3Ti maraging steel through controlled solute segregation and α-Mn nanoprecipitation, Nat. Commun. 13 (1) (2022) 1–8, https://doi.org/10.1038/s41467-022-30019-x, https://www.nature.com/articles/s41467-022-30019-x.

[19] H. Huang, M.Y. Sherif, P.E. Rivera-Díaz-Del-Castillo, Combinatorial optimization of carbide-free bainitic nanostructures, Acta Mater. 61 (5) (2013) 1639–1647, https://doi.org/10.1016/J.ACTAMAT.2012.11.040.

[20] C. Garcia-Mateo, F.G. Caballero, H.K. Bhadeshia, Acceleration of low-temperature bainite, ISIJ Int. 43 (11) (2003) 1821–1825, https://doi.org/10.2355/ISIJINTERNATIONAL.43.1821.

[21] F.G. Caballero, C. Garcia-Mateo, Super-bainite, in: Encyclopedia of Materials: Metals and Alloys, 2021, pp. 73–83, https://digital.csic.es/handle/10261/226455.

[22] R. Kannan, G.L. Knapp, P. Nandwana, R. Dehoff, A. Plotkowski, B. Stump, et al., Data mining and visualization of high-dimensional ICME data for additive manufacturing, Integr. Mater. Manuf. Inn. 11 (1) (2021) 57–70, https://doi.org/10.1007/S40192-021-00243-2, https://link.springer.com/article/10.1007/s40192-021-00243-2.

[23] TC-Python API reference documentation — TC-Python 2022b documentation, https://download.thermocalc.com/docs/tc-python/2022b/html/.

[24] A. Stormvinter, A. Borgenstam, J. Ågren, Thermodynamically based prediction of the martensite start temperature for commercial steels, Metall. Trans. A, Phys. Metall. Mater. Sci. 43 (10) (2012) 3870–3879, https://doi.org/10.1007/S11661-012-1171-Z/FIGURES/15, https://link.springer.com/article/10.1007/s11661-012-1171-z.

[25] H.K. Bhadeshia, Thermodynamic analysis of isothermal transformation diagrams, Met. Sci. 16 (3) (1982) 159–165, https://doi.org/10.1179/030634582790427217, https://www.tandfonline.com/doi/abs/10.1179/030634582790427217.

[26] MAP program MUCG83, https://www.phase-trans.msm.cam.ac.uk/map/steel/programs/mucg83.html.

[27] GitHub MUCG83, https://github.com/arthursn/mucg83.

[28] Getting started - fast ML preset, https://github.com/sdv-dev/SDV/discussions/786.

[29] Synthetic Data Generation for tabular, relational and time series data, https://github.com/sdv-dev/SDV.

[30] N. Patki, R. Wedge, K. Veeramachaneni, The synthetic data vault, in: Proceedings - 3rd IEEE International Conference on Data Science and Advanced Analytics, DSAA 2016, 2016, pp. 399–410, https://github.com/sdv-dev/SDV.

[31] I. Goodfellow, NIPS 2016 tutorial: generative adversarial networks, preprint, arXiv:1701.00160, 2016.

[32] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, et al., Improved techniques for training GANs, in: Advances in Neural Information Processing Systems, 2016, p. 29, https://github.com/openai/improved-gan.