Scripta Materialia 183 (2020) 1–5

Contents lists available at ScienceDirect

Scripta Materialia

journal homepage: www.elsevier.com/locate/scriptamat

Machine learning-based prediction models for formation energies of interstitial atoms in HCP crystals

Daegun You[a], Shraddha Ganorkar[a], Sooran Kim[b], Keonwook Kang[c], Won-Yong Shin[d,*], Dongwoo Lee[a,*]

[a] School of Mechanical Engineering, Sungkyunkwan University, Suwon, Gyeonggi-do, South Korea
[b] Department of Physics Education, Kyungpook National University, Daegu, South Korea
[c] Department of Mechanical Engineering, Yonsei University, Seoul, South Korea
[d] Department of Computational Science and Engineering, Yonsei University, Seoul, South Korea

ARTICLE INFO

Article history:
Received 8 November 2019
Revised 7 February 2020
Accepted 25 February 2020
Available online 13 March 2020

Keywords:
Formation energy
Interstitial atom
HCP crystal
Machine learning
First-principles calculation

ABSTRACT

Prediction models of the formation energies of H, B, C, N, and O atoms in various interstitial sites of *hcp*-Ti, Zr, and Hf crystals are developed based on machine learning. Parametric models such as linear regression and brute force search (BFS) as well as nonparametric algorithms including the support vector regression (SVR) and the Gaussian process regression (GPR) are employed. Readily accessible chemical and geometrical descriptors allow straightforward implementation of the prediction models without any expensive computational modeling. The models based on BFS, SVR, and GPR show the excellent performance with $R^2 > 96\%$.

Interstitial atoms play significant roles in determining physical and chemical properties of solids. Fundamental understanding of the stabilities of interstitial atoms at various sites of host crystalline is crucial to design proper processing conditions of compounds and to ensure their service lives in important technological applications. Formation energy E_f of an interstitial atom is a measure of the atom's stability [1–7]. Experimental techniques, such as nano-calorimetry [6–9] and nuclear reaction analysis [10], can be employed to determine E_f. These experiments, however, often suffer from low resolutions to determine E_f at low concentration of the interstitials. First-principles calculations based on density functional theory (DFT) offer an alternative method to determine E_f with high accuracy, although it is time-consuming to calculate E_f for a broad range of materials systems. For a broad range of crystal compounds without interstitial atoms inside, the values of E_f can readily be accessed from DFT-based materials libraries, such as Materials Project and AFLOW [11,12]. Additionally, prediction models based on machine learning (ML) are useful tools for predicting physical and chemical properties, including formation energy, band gap, stability of crystals, glass-forming ability, and configurational energy of various materials [13–24]. However, both the materials database and the prediction models for E_f of interstitial atoms are still limited and insufficiently explored.

In this work, we present a strategy to develop prediction models for E_f of various interstitial atoms of H, B, C, N, and O at the manifold interstitial sites of Bond-centered (BC), Crowdion (CD), Hexahedral (Hexa), Plane-centered (PC), Octahedral (Octa), and Tetrahedral (Tetra) in *hcp* Ti, Zr, and Hf hosts. Recently, linear regression (LR) models were proposed in [5] to predict E_f of interstitial atoms in Zr. Although the suggested models use the easily accessible chemical and geometrical parameters and showed 97% accuracy, the analyses were limited to a small set of data. Here, we use ML-based models for an expanded set of interstitial–host systems. The values of E_f for various interstitial-host systems were acquired by first-principles calculations. Then, LR, support vector regression (SVR) [25–28], Gaussian process regression (GPR) [29–31], and LR with non-linear heuristic forms [17–19] by brute force searching (BFS) were implemented to predict E_f with easily accessible input parameters. The workflow of this study is summarized in Fig. 1.

The structure of a typical *hcp* crystal (space group: $P6_3/mmc$) with various interstitial sites is shown in Fig. 1 and the corresponding positions are provided in Table 1. Each interstitial-host system was created by inserting an interstitial atom (H, B, C, N, O) at an interstitial site (CD, Hexa, BC, PC, Octa, Tetra) of a host (Ti, Zr,

* Corresponding author.
E-mail addresses: wy.shin@yonsei.ac.kr (W.-Y. Shin), dongwoolee@skku.edu (D. Lee).

https://doi.org/10.1016/j.scriptamat.2020.02.042
1359-6462/

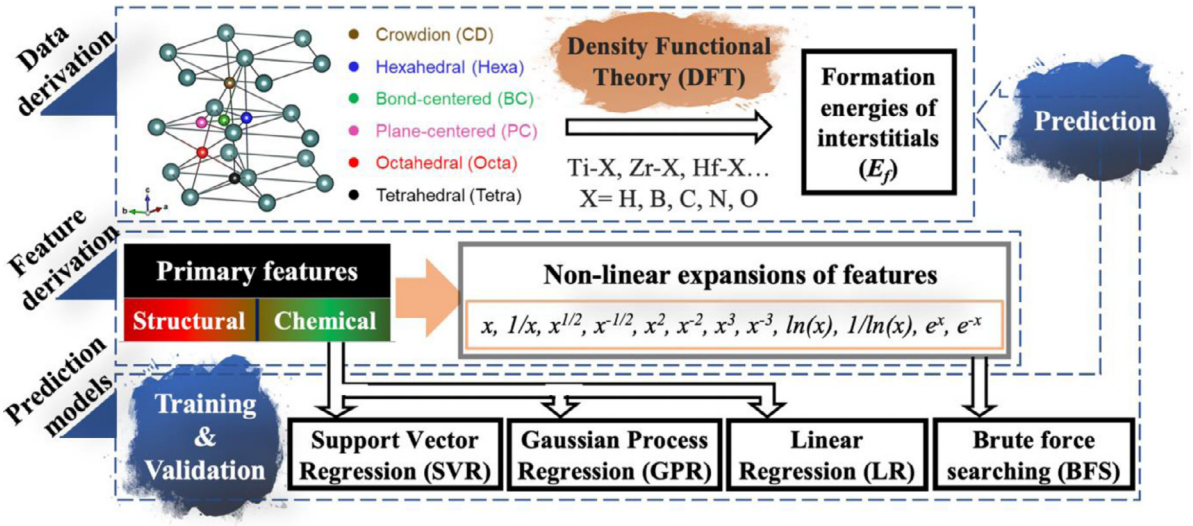

Fig. 1. Workflow of the ML-based prediction models of E_f; Data derivation – determination of E_f of the interstitial atoms in the hcp crystals by DFT simulations, Feature derivation – selection and expansion of the primary features, Prediction models – implementation of parametric (LR and BFS) and nonparametric (SVR and GPR) models.

Table 1
Positions of various interstitial sites in the hcp crystal.

Interstitial site	Position
Crowdion (CD)	(0.33 0.17 0.75)
Octahedral (Octa)	(0.33 0.67 0.25)
Bond-centered (BC)	(0.17 0.33 0.50)
Hexahedral (Hexa)	(0.00 0.00 0.50)
Plane-centered (PC)	(0.33 0.67 0.50)
Tetrahedral (Tetra)	(0.67 0.33 0.12)

Hf). A total of $5 \times 6 \times 3 = 90$ structures were then relaxed by using DFT simulations (Section 1 in the Supplementary Material). For a relaxed structure, E_f of an interstitial atom can be determined as [5–7]

$$E_f = E_{\text{total}} - N_{\text{host}} \times E_{\text{host}} - 1 \times E_{\text{interstitial}}, \quad (1)$$

where E_{total} is the total energy of the relaxed supercell including the interstitial and host atoms, E_{host} and $E_{\text{interstitial}}$ are the energy/atom in the bulk forms of the host and interstitial elements, respectively, and N_{host} is the number of atoms in the host crystal. The lattice parameters and E_{host} were determined as Ti: ($a = 2.93$ Å, $c = 4.64$ Å, $E_{Ti} = -7.94$ eV); Zr: ($a = 3.24$ Å, $c = 5.17$ Å, $E_{Zr} = -8.55$ eV); and Hf: ($a = 3.20$ Å, $c = 5.06$ Å, $E_{Hf} = -9.55$ eV), which agree well with the data in Materials Project [11].

Fig. 2(a)–(e) show E_f calculated by Eq. (1) for the host-X (host = Ti, Zr, Hf; $X = H$, B, C, N, and O) systems at various interstitial sites. Table SI of the Supplementary Material shows the exact values of E_f. Some of the host-X pairs were found to be unstable as the interstitial atom slides to the nearby interstitial site upon the relaxation. E_f for the unstable systems cannot be determined, making the total number, n, of the E_f values for the analyses as 73. The values of E_f for some of the systems were acquired from literature and compared in Fig. 2 [6,7,32–34], showing reasonable agreements with the results from the current study.

As illustrated in Fig. 2, E_f depends more strongly on the interstitial atom and the site rather than the host atom. This is probably due to the chemical similarity (i.e., EN and VE) of the host atoms - Ti, Zr, and Hf. E_f is found to be the minimum at Octa and the maximum at PC for most of the systems considered in this work. Note that in the host-H systems (Fig. 2(a)), however, Tetra or Octa are the most stable ones, which is consistent with the previous reports [32–34,37], while CD or BC are the least stable ones. Tetra is unstable for many of the systems but is stable only in some of the host-H (Fig. 2(a)) or host-O (Fig. 2(e)) systems [38–40]. As seen in Fig. 2(f), these results may be attributed to the combined effects of the large differences in the atomic radii and EN between the constituent atoms, i.e., H and O have the largest differences with the host atoms in atomic radii and EN, respectively. Among the systems considered, the host-O pairs (Fig. 2(e)) are the most stable with the lowest E_f, ranging from −5.56 to −2.23 eV. The host-N pairs (Fig. 2(d)) are the second most stable ones, with E_f ranging from −3.91 to 0.01 eV. The higher stability of O and N interstitial atoms can largely be attributed to their high electronegativity (Fig. 2(f)). The larger difference between electronegativities of constituent atoms results in charge transfer from the host to the interstitial atom, thereby forming the mixed ionic-covalent bond nature [4].

To develop prediction models for E_f, we carefully selected the chemical and geometrical input parameters (i.e., features or descriptors) based on their accessibility and their influence on E_f. Miedema [41–43] and Hume-Rothery rules [44–47] are the semi-empirical models for the stability of crystals, which corroborate the role of various chemical and geometrical features such as electronegativity, atomic radii, valance electron, coordination number, crystal structure, electron density at the boundary of Wigner-Seitz cell, etc. Based on these theories and the observations from Fig. 2, we chose the chemical features as the electronegativity difference (ΔEN) between host-interstitial atoms, and the number of valence electrons VE of the interstitial atoms [36]. We also considered the average electronegativity EN_{avg} of the constituent elements, as both ΔEN and EN_{avg} are known to determine a bonding type according to the van Arkel diagram [48–51]. For the geometrical features, we considered the average distance between the surfaces d_{surf} of an interstitial atom and its host, the bonding similarity of interstitial site I_{sim}, and the shape factor η, each of which can be determined as:

$$d_{\text{surf}} = \left| d_{\text{avg}} - \left(R_{\text{host}} + R_{\text{imp}} \right) \right| \quad (2)$$

$$I_{\text{sim}} = (1 - d_{\text{std}}) \times CN \quad (3)$$

$$\eta = \frac{V^{2/3}}{A}, \quad (4)$$

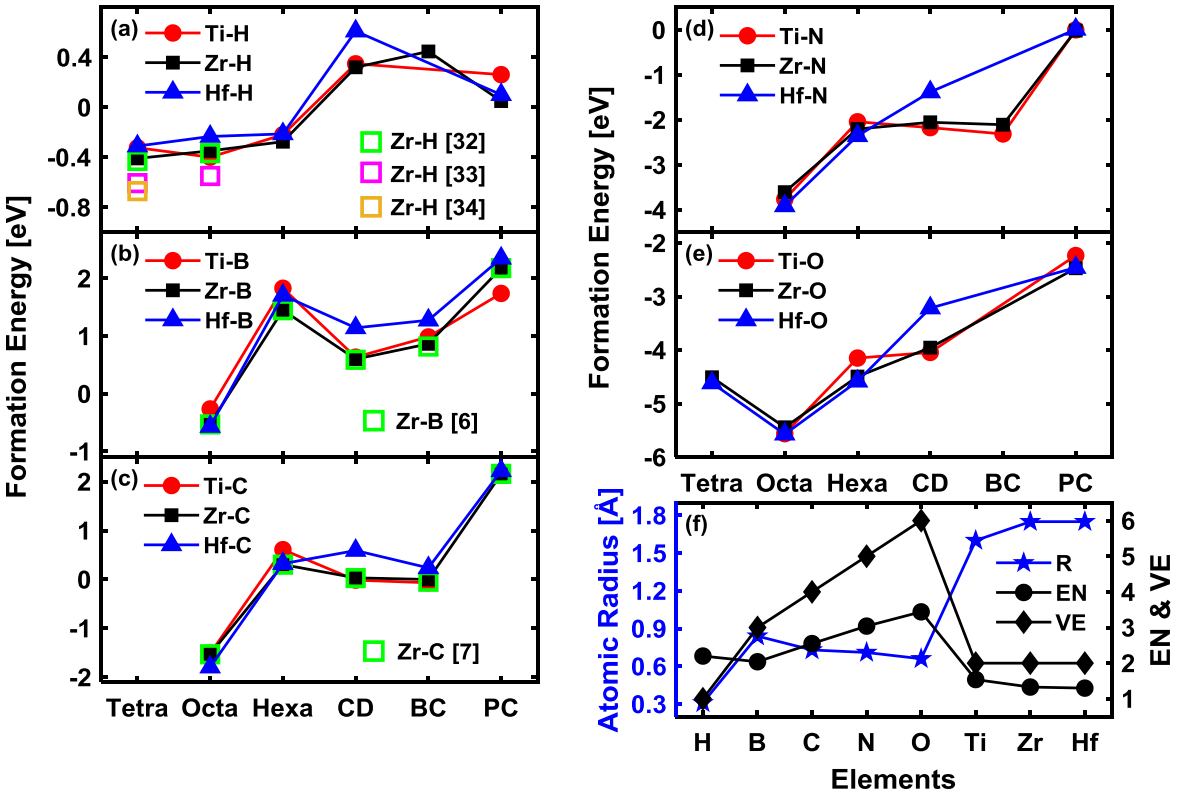

Fig. 2. (a)–(e) The formation energies E_f of interstitial atoms at the various sites in the Ti, Zr, and Hf host crystals with the comparison of the reference data [6, 7, 32–34], (f) covalent atomic radii (R) [35], the number of valence electrons (VE) [36], and Pauling's electronegativity (EN) of the host and interstitial atoms. For the transition metals, we use the lowest possible valence electrons.

respectively, where d_{avg} and d_{std} are the average and standard deviation of the distances between the center of an interstitial atom and its neighboring host atoms; R_{host} and R_{imp} are the atomic radii of the host and interstitial atoms; CN is the coordination number of the interstitial atom; and V and A are the volume and surface area of the interstitial site, respectively. In ref [5], d_{surf} was shown to be the strongest geometrical feature for predicting E_f in *hcp*-Zr. Bonding similarity I_{sim} correlates negatively with the deviation of the bond lengths from equality and is proportional to CN. For example, Octa has a large value for I_{sim} because the interstitial site has six bonds ($CN = 6$) with the same length. The parameter η is a dimensionless volume to the surface ratio of the interstitial site. The values of all the primary features (i.e., EN_{avg}, ΔEN, VE, d_{surf}, I_{sim}, and η) for each system are summarized in Tables SII and SIII in the Supplementary Material. It is worth noting that all the primary features can readily be determined without carrying out any DFT simulations.

The correlations among the seven variables (the six primary features and the output parameter E_f) were analyzed using Spearman's rank correlation coefficient, which is a nonparametric measure of statistical dependence between the rankings of two variables [52–57]. The Spearman correlation coefficient $\rho_{i\text{-}ii}$ between the i th and ii-th variable is given by

$$\rho_{i-ii} = \frac{\sum_{j=1}^{n}\left\{\left(r_{i,\ j}-\bar{r}_i\right)\cdot\left(r_{ii,\ j}-\bar{r}_{ii}\right)\right\}}{\sqrt{\left\{\sum_{j=1}^{n}\left(r_{i,\ j}-\bar{r}_i\right)^2\right\}\cdot\left\{\sum_{j=1}^{n}\left(r_{ii,\ j}-\bar{r}_{ii}\right)^2\right\}}}, \quad (5)$$

where $r_{i,j}$ and $r_{ii,j}$ are the ranks of each system j among the total dataset $n = 73$ in the i th and ii-th variables, respectively. The Spearman correlations for all the variables are depicted in Fig. 3. The chemical features (e.g., EN_{avg}, ΔEN, and VE) tend to show higher correlations with E_f compared to the geometrical ones (e.g., I_{sim}, d_{surf}, and η). However, the chemical features are correlated with each other significantly; thus, making use of all the chemical features would not noticeably improve the prediction performance.

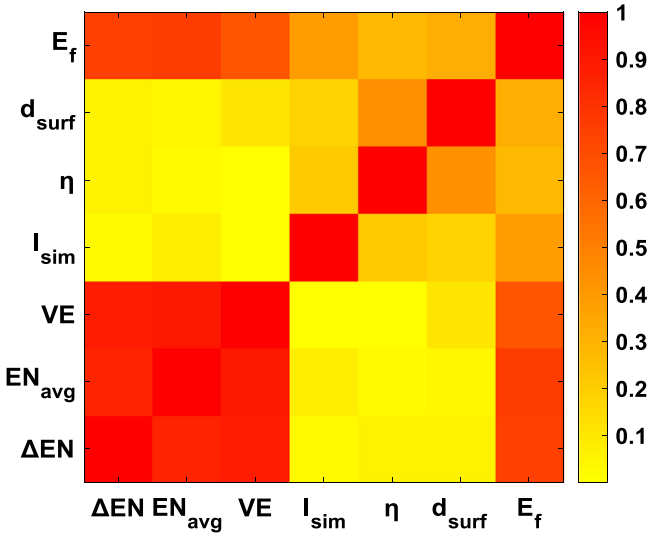

Fig. 3. The correlation among the features and E_f determined by the absolute values of the Spearman's rank coefficient.

The LR models for predicting E_f according to different numbers of features were developed. For the models with less than six primary features, we selected the features by sequentially removing the highly correlated input features (in order of *VE and* ΔEN) first, and removing the less correlated input features to the output parameter (in order of η, d_{surf}, and I_{sim}). For the LR models, the form $E_f = \theta_0 + \sum_{i=1}^{m}\theta_i x_i$ was used, where θ_i represents the regression

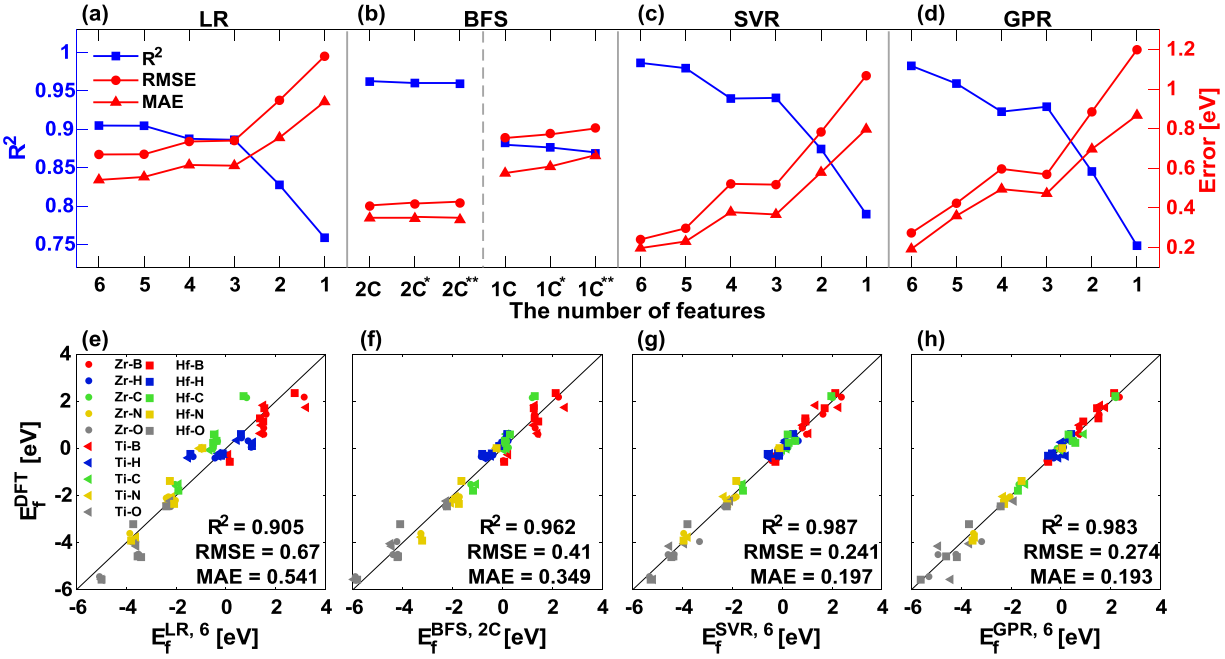

Fig. 4. Performances of the prediction models. (a)–(d) Plots of the coefficient of determination (RR^2), RMSE, and MAE according to the number of features in descending order. (e)–(h) Comparison of E_f from the DFT simulations and the prediction models for various host-interstitial systems. The models from left to right are LR, BFS, SVR, and GPR for both upper and bottom panels.

coefficient, m the total number of considered features, and x_i the value of the i th feature. The coefficients can be determined by minimizing the error $\delta = \|E_f^{DFT} - (\theta_0 + \sum_{i=1}^{m} \theta_i x_i)\|_2^2$. The detailed procedure for the error minimization and the exact form of each equation can be found in Section 2 of the Supplementary Material. Fig. 4(a) shows the performance of the LR model versus the number of features, m. The LR model with all 6 features results in R R^2 = 0.905 with a root mean square error (RMSE) of 0.67 eV and a mean absolute error (MAE) of 0.541 eV. There is only a slight reduction in R^2 as the number of feature decreases until m = 3 (R R^2 = 0.886), but the accuracy starts to degrade abruptly as the number of features is further reduced.

The BFS model is another parametric model that expands and convolves the primary features to the compound features in non-linear forms, and uses these new features for modeling E_f. With the 6 primary features (i.e., EN_{avg}, ΔEN, VE, d_{surf}, I_{sim}, and η), we utilized 12 functions of x, $\frac{1}{x}$, $\sqrt{x}$, $\frac{1}{\sqrt{x}}$, x^2, $\frac{1}{x^2}$, x^3, $\frac{1}{x^3}$, $\ln x$, $\frac{1}{\ln x}$, e^x, and $\frac{1}{e^x}$, to produce a total of $6 \times 12 = 72$ prototypical features [17–19]. Multiplication of two or three of the prototypical features yields 54,025 distinct compound features (i.e., $\frac{EN_{avg}}{\ln \eta}$, $\frac{\Delta EN \times d_{surf}^2}{\sqrt{I_{sim}}}$, etc.), which constitute the prediction models in the form of $E_f = \theta_0 + \theta_1 f_{c1}$ (1C BFS model) or $E_f = \theta_0 + \theta_1 f_{c1} + \theta_2 f_{c2}$ (2C BFS model). Here, f_{c1} and f_{c2} are the compound or prototypical features. As in the LR models, the coefficients were determined by minimizing the error. Both the 1C and 2C BFS models were subjected to searching the best LR fit by training only, which gives 6,459,645 LR trained models. The 10 best-fit BFS models for each 1C and 2C case were subjected to 10-fold cross-validation (Section 2 in the Supplementary Material). Fig. 4(b) and (f) represent the results of the 2C (RR^2 = 0.962, RMSE = 0.41 eV, MAE = 0.349 eV) and 1C (RR^2 = 0.882, RMSE = 0.755 eV, MAE = 0.575 eV) BFS models with the highest accuracy. In Fig. 4(b), the second and third best BFS models using 2C or 1C features are also shown and labeled as 2C*/2C** or 1C*/1C**, respectively. The best models with the same number of the compound features show similar performances.

ML-based nonparametric models have also been developed. These models utilize 1 to 6 primary features based on the Spearman's rank coefficients, as in the LR models. Two algorithms including SVR and GPR were employed. SVR is a supervised learning algorithm that uses a kernel function, which enables the non-linear mapping of data into a high-dimensional space [25–28]. GPR is based on the Bayesian method, also known as the stochastic procedure learning [29–31], which utilizes the probability density and noise in observation values. We used radial-basis functions (RBF) for the kernel function for both SVR and GPR and assumed noise-freeness. Each model of SVR and GPR was subjected to 10-fold cross-validation (Section 2 of the Supplementary Material). Fig. 4(c), (d), (g), and (h) display the results from SVR and GPR. From these figures, we find that R^2 = {0.987, 0.983}, RMSE = {0.241 eV, 0.274 eV}, and MAE = {0.197 eV, 0.193 eV}, for SVR and GPR, respectively when m = 6. As in the LR models, both SVR and GPR show relatively high accuracies (R R^2 > 0.9) for m decreases from 6 to 3, but the performance decreases abruptly when m decreases from 3 to 1.

As seen in Fig. 4, the accuracy performance of SVR, GPR, and BFS with the 6 primary features are outstanding by exhibiting R R^2 > 0.96, while the performance of the LR model with the same number of features is inferior. This difference is largely attributed to the non-linear relationships not only between the input and output parameters but also among the input parameters. The chemical and geometrical features highly interact with each other and have non-linear relationships to determine the formation energy. Note that while LR and BFS models allow direct access to analytical expressions for E_f, SVR and GPR models are inadequate to do so due to the implicit regression processes. Therefore, the BFS models are beneficial as they have explicit forms as well as high accuracy. Nonparametric models of SVR and GPR are beneficial because of their low computational cost and satisfactory accuracy: the use of 3 features with these models lead to RR^2 > 0.92. The models based on the different algorithms indicate that the features related to electronegativity are the most critical factor for predicting E_f, which coincides with both the Spearman's coefficient (see Fig. 3) and the trend identified from the analyses in Fig. 2.

The transferability of the models developed in this work was tested for other *hcp* host-interstitial systems of Co-H and Re-H. As described in Section 3 in the Supplementary Material, the models predict the test systems reasonably well. The application of the developed model for predicting E_f to the material systems with high interstitial concentration would be less accurate because our models use the geometrical features based on the undeformed host crystals, while the host-interstitial system would be more deformed at higher concentrations. However, similar ML approaches can be adopted for high accuracy by using sufficiently large training data for higher concentration cases.

In conclusion, ML-based prediction models and descriptors for the formation energies E_f of the various interstitial atoms H, B, C, N, and O at multiple interstitial sites of *hcp* Ti, Zr, and Hf crystals have been investigated. The formation energies calculated from DFT were used as the output parameter for the models. The six primary chemical and geometrical input parameters were selected based on their impacts on E_f and also their accessibilities. The LR model exhibited relatively low accuracy, while the non-linear heuristic forms via the BFS model and the nonparametric models built upon SVR and GPR showed high accuracies. The procedures for feature selection and for prediction model derivation discussed in this paper can be used to study a wider range of host-interstitial systems to achieve fundamental understanding of the stabilities of interstitial atoms.

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

This work was supported by the National Research Foundation of Korea (NRF) under Grant No. NRF-2017R1E1A1A01078324 and Samsung Research Funding and Incubation Center for Future Technology under Grant No. SRFC-MA1802-06.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.scriptamat.2020.02.042.

References

[1] Z. Lei, X. Liu, Y. Wu, H. Wang, S. Jiang, S. Wang, X. Hui, Y. Wu, B. Gault, P. Kontis, D. Raabe, L. Gu, Q. Zhang, H. Chen, H. Wang, J. Liu, K. An, Q. Zeng, T.-G. Nieh, Z. Lu, Nature 563 (7732) (2018) 546–550.
[2] Z. Li, C.C. Tasan, H. Springer, B. Gault, D. Raabe, Sci. Rep. 7 (2017) 40704.
[3] Z. Wang, I. Baker, Z. Cai, S. Chen, J.D. Poplawsky, W. Guo, Acta Mater. 120 (2016) 228–239.
[4] D. Lee, J.J. Vlassak, K. Zhao, ACS Appl. Mater. Inter. 8 (17) (2016) 10995–11000.
[5] D. You, S. Ganorkar, M. Joo, D. Park, S. Kim, K. Kang, D. Lee, J. Alloys Compd. 787 (2019) 631–637.
[6] D. Lee, J.J. Vlassak, K. Zhao, Nano Lett. 15 (10) (2015) 6553–6558.
[7] D. Lee, G.D. Sim, K. Zhao, J.J. Vlassak, Nano Lett. 15 (12) (2015) 8266–8270.
[8] D. Lee, B. Zhao, E. Perim, H. Zhang, P. Gong, Y. Gao, Y. Liu, C. Toher, S. Curtarolo, J. Schroers, J.J. Vlassak, Acta Mater. 121 (2016) 68–77.
[9] D. Lee, J.J. Vlassak, Scr. Mater. 165 (2019) 73–77.
[10] C.S. Zhang, B. Li, P.R. Norton, J. Alloys Compd 231 (1) (1995) 354–363.
[11] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, APL Mater 1 (1) (2013) 011002.
[12] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R.H. Taylor, L.J. Nelson, G.L.W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, O. Levy, Comput. Mater. Sci. 58 (2012) 227–235.
[13] A. Seko, T. Maekawa, K. Tsuda, I. Tanaka, Phys. Rev. B 89 (5) (2014) 054303.
[14] Y. Saad, D. Gao, T. Ngo, S. Bobbitt, J.R. Chelikowsky, W. Andreoni, Phys. Rev. B 85 (10) (2012) 104104.
[15] G. Pilania, J.E. Gubernatis, T. Lookman, Phys. Rev. B 91 (21) (2015) 214302.
[16] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, R. Ramprasad, Sci. Rep. 3 (2013) 2810.
[17] G. Pilania, A. Mannodi-Kanakkithodi, B.P. Uberuaga, R. Ramprasad, J.E. Gubernatis, T. Lookman, Sci. Rep. 6 (2016) 19375.
[18] C. Kim, G. Pilania, R. Ramprasad, Chem. Mater 28 (5) (2016) 1304–1311.
[19] L.M. Ghiringhelli, J. Vybiral, S.V. Levchenko, C. Draxl, M. Scheffler, Phys. Rev. Lett. 114 (10) (2015) 105503.
[20] W. Ye, C. Chen, Z. Wang, I.-H. Chu, S.P. Ong, Nat. Commun. 9 (1) (2018) 3800.
[21] F. Ren, L. Ward, T. Williams, K.J. Laws, C. Wolverton, J. Hattrick-Simpers, A. Mehta, Sci. Adv. 4 (4) (2018) eaaq1566.
[22] Y.T. Sun, H.Y. Bai, M.Z. Li, W.H. Wang, J. Phys. Chem. Lett. 8 (14) (2017) 3434–3439.
[23] L. Ward, S.C. O'Keeffe, J. Stevick, G.R. Jelbert, M. Aykol, C. Wolverton, Acta Mater. 159 (2018) 102–111.
[24] A.R. Natarajan, A. Van der Ven, npj Comput. Mater 4 (1) (2018) 56.
[25] C.-C. Chang, C.-J. Lin, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) 1–27.
[26] S.R. Gunn, ISIS technical report 14 (1) (1998) 5–16.
[27] A.J. Smola, B. Schölkopf, Stat Comput 14 (3) (2004) 199–222.
[28] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, 1995.
[29] J. Quiñonero-Candela, C.E. Rasmussen, J. Mach, Learn. Res. 6 (2005) 1939–1959.
[30] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, University Press Group Limited, 2006.
[31] J. Snoek, H. Larochelle, R.P. Adams, Adv. Neural Inf. Process. Syst. 25 (2012) 2951–2959.
[32] Y. Zhang, C. Jiang, X. Bai, Sci. Rep. 7 (2017) 41033.
[33] C. Domain, R. Besson, A. Legris, Acta Mater. 50 (13) (2002) 3513–3526.
[34] Y. Fukai, The Metal-Hydrogen System, Springer, 1993.
[35] B. Cordero, V. Gómez, A.E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán, S. Alvarez, Dalton Trans. 21 (2008) 2832–2838.
[36] J. Rumble, CRC Handbook of Chemistry and Physics, CRC Press LLC, 2018.
[37] S.K. Nayak, C.J. Hung, V. Sharma, S.P. Alpay, A.M. Dongare, W.J. Brindley, R.J. Hebert, npj Comput. Mater 4 (1) (2018) 11.
[38] R.G. Hennig, D.R. Trinkle, J. Bouchet, S.G. Srinivasan, R.C. Albers, J.W. Wilkins, Nat. Mater. 4 (2005) 129.
[39] H.H. Wu, D.R. Trinkle, Phys. Rev. Lett. 107 (4) (2011) 045504.
[40] H.H. Wu, P. Wisesa, D.R. Trinkle, Phys. Rev. B 94 (1) (2016) 014307.
[41] A.R. Miedema, Philips Tech. Rev. 33 (6) (1973) 149–160.
[42] A.R. Miedema, F.R.d. Boer, P.F.d. Chatel, J. Phys. F: Met. Phys. 3 (8) (1973) 1558.
[43] A.R. Miedema, R. Boom, F.R. De Boer, J. Less-Common Met. 41 (2) (1975) 283–298.
[44] L.S. Darken, R.W. Gurry, Physical Chemistry of Metals (Metallurgy and Metallurgical Engineering Series), McGraw-Hill, New York, 1953.
[45] T. Massalski, Physical metallurgy, North-Holland, 1996, pp. 135–204.
[46] U. Mizutani, MRS Bull. 37 (2) (2012) 169.
[47] W.H. Rothery, J. Inst. Met. 35 (1926) 295–354.
[48] L.C. Allen, J. Am. Chem. Soc. 111 (25) (1989) 9003–9014.
[49] L.C. Allen, J.F. Capitani, G.A. Kolks, G.D. Sproul, J. Mol. Struct. 300 (1993) 647–655.
[50] A.E.V. Arkel, Molecules and Crystals in Inorganic Chemistry, Interscience Publishers, New York, NY, USA, 1956.
[51] J.A.A. Ketelaar, Chemical Constitution: An Introduction to the Theory of the Chemical Bond, 2 ed., Elsevier, New York, NY, USA, 1958.
[52] C. Spearman, Am. J. Appl. Psychol. 15 (1) (1904) 72–101.
[53] T. Cleff, Exploratory Data Analysis in Business and Economics, Springer, Switzerland, 2014.
[54] M.G. Kendall, Biometrika 33 (1945) 239–251.
[55] Y. Iwasaki, A.G. Kusne, I. Takeuchi, npj Comput. Mater. 3 (1) (2017) 4.
[56] B. Medasani, A. Gamst, H. Ding, W. Chen, K.A. Persson, M. Asta, A. Canning, M. Haranczyk, npj Comput. Mater 2 (1) (2016) 1.
[57] Y. Zhang, C. Ling, npj Comput. Mater 4 (1) (2018) 25.