Polymer informatics beyond homopolymers

Shivank S. Shukla, Christopher Kuenneth, and Rampi Ramprasad*

Polymers are diverse and versatile materials that have met a wide range of material application demands. They come in several flavors and architectures (e.g., homopolymers, copolymers, polymer blends, and polymers with additives). Searching this enormous space for suitable materials with a specific set of property/performance targets is thus nontrivial, painstaking, and expensive. Such a search process can be made effective by the creation of rapid and accurate property predictors. In this article, we present a machine learning framework to predict the thermal properties of homopolymers, copolymers, and polymer blends. A universal fingerprinting scheme capable of handling this entire polymer chemical class has been developed and a multitask deep learning algorithm is trained simultaneously on a large data set of glass-transition, melting, and degradation temperatures. The trained models demonstrate precision and scalability to other properties when relevant data becomes accessible.

Impact statement

The chemical and structural variations that can be achieved with the polymers is staggering. Such extraordinary and diverse possibilities translate to attractive combinations of physical properties impacting several application spaces, making the polymeric class of materials ubiquitous in our modern society. This chemistry–structure–property diversity is accompanied by a major challenge. Searching the chemo-structural space to identify suitable application-relevant candidates with the right set of target properties is nontrivial, requiring advanced rapid property prediction and search schemes. In this article, we present a data-driven machine learning framework to instantaneously predict the thermal properties (an important property class) of a dizzyingly large class of polymer archetypes, namely, homopolymers, copolymers, and polymer blends. A state-of-the-art machine learning algorithm has been developed and trained on a large data set of glass-transition, melting, and degradation temperatures, to make instantaneous predictions of these properties for any new-to-the-world polymer that falls in this large important polymer chemical class. This prediction scheme paves the way for discovering polymers with unprecedented thermal stability by allowing searches of enormous chemical spaces at scale.

Introduction

Polymeric materials come in a variety of flavors and architectures, such as homopolymers, copolymers, polymer blends, and polymers with additives such as dopants, plasticizers, and organic/inorganic fillers.[1–4] The extraordinary chemical and structural diversity offered by such materials lead to wide-ranging and attractive combinations of physical properties impacting several application spaces, ranging from structural, electrical, packaging, chemical separation, health care, energy, and sustainable technologies.[5–17]

In an effort to simultaneously optimize multiple (correlated or uncorrelated) properties, the community has explored and developed polymer varieties beyond neat homopolymers, namely, copolymers, polymer blends, and polymers with additives. Finding optimal candidates possessing a predefined set of property attributes has largely been guided by experience, intuition, and trial-and-error approaches. An exhaustive search of the relevant chemical spaces is nontrivial given the vast expanse of the spaces. Over the last decade or so, polymer informatics approaches have attempted to aid this search process by offering data-driven machine learning (ML) models to rapidly predict the properties of new polymer formulations and to recommend candidate materials that may meet multi-property target requirements.[18–27] Nevertheless, these polymer informatics efforts have thus far largely focused on neat homopolymers,[28] although notable exceptions exist within recent attempts to address copolymer chemistries,[29–34] polymers with dopants, and polymer composites.[35]
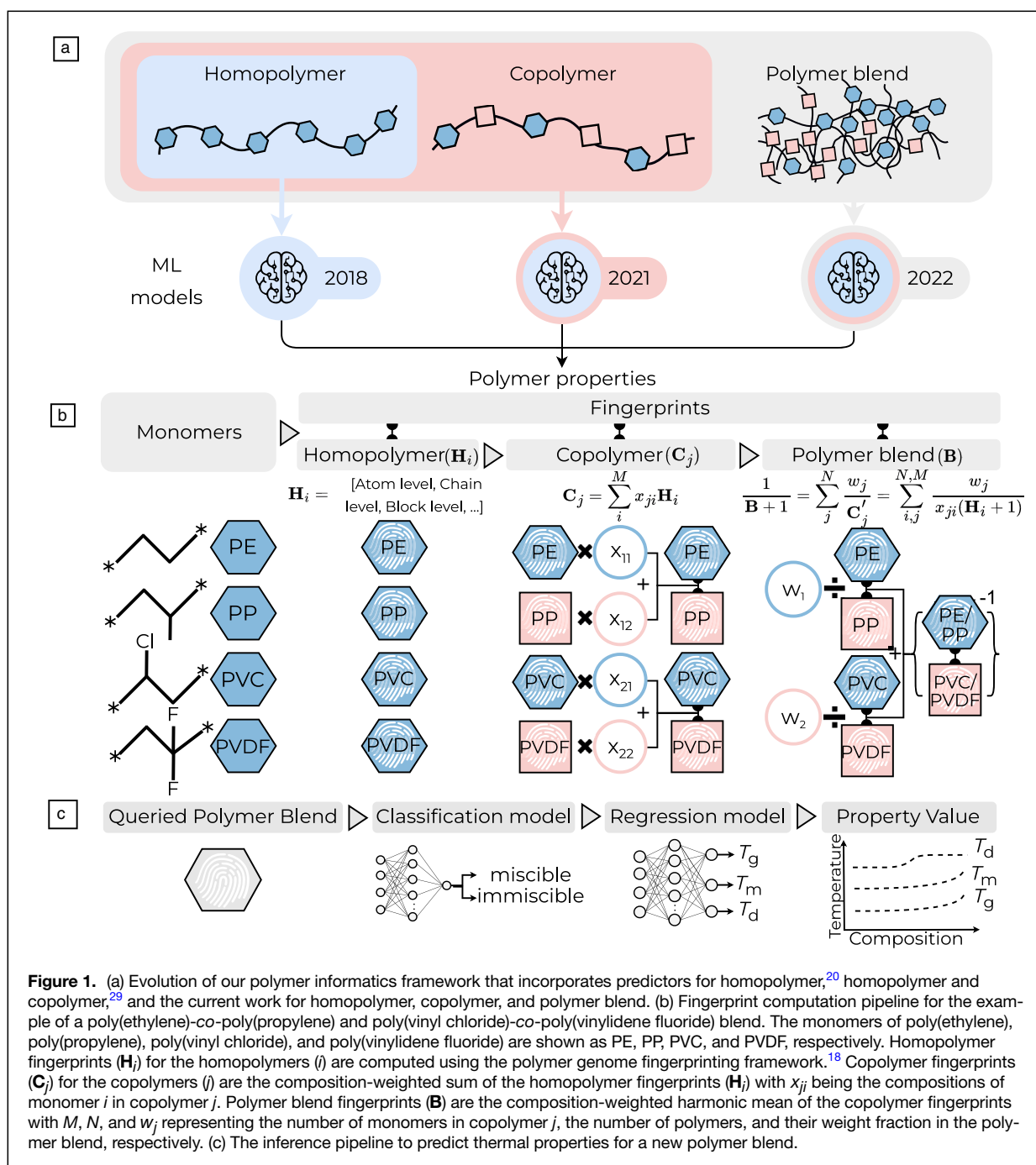
In this contribution, we lay the groundwork to handle neat homopolymers, copolymers, and polymer blends within one unified multitask neural network polymer informatics framework. For definiteness, we focus on thermal properties, namely,

Shivank S. Shukla, School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, USA
Christopher Kuenneth, School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, USA; Faculty of Engineering Science, Universität Bayreuth, Bayreuth, Germany
Rampi Ramprasad, School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, USA; rampi.ramprasad@mse.gatech.edu
*Corresponding author
doi:10.1557/s43577-023-00561-0

the glass-transition temperature (T_g), the melting temperature (T_m), and the degradation temperature (T_d). Our method combines a novel fingerprinting scheme for polymer blends with a unified fingerprinting approach for all three types of polymers. We fine-tuned a ML architecture to accurately predict thermal properties, including the glass-transition temperature (T_g), the melting temperature (T_m), and the degradation temperature (T_d) of all three types of polymers, including miscible and immiscible polymer blends. Furthermore, we developed a classification model to predict the miscibility of polymer blends. Past efforts have mainly focused on homopolymers and copolymers, but our

framework can handle all three types of polymers, and it can predict these properties for any new homopolymer, copolymer, or polymer blend.

As portrayed in **Figure 1**a, homopolymers are a subset of copolymers, and copolymers are a subset of polymer blends. Homopolymers are defined by one monomer repeat unit and copolymers by multiple monomer units. Polymer blends are a physical mixture of two or more homopolymer(s) and/or copolymer(s). In this work, we assume that our copolymers are random (i.e., the multiple repeat units are distributed randomly along the polymer backbone); we make this assumption

Figure 1. (a) Evolution of our polymer informatics framework that incorporates predictors for homopolymer,[20] homopolymer and copolymer,[29] and the current work for homopolymer, copolymer, and polymer blend. (b) Fingerprint computation pipeline for the example of a poly(ethylene)-*co*-poly(propylene) and poly(vinyl chloride)-*co*-poly(vinylidene fluoride) blend. The monomers of poly(ethylene), poly(propylene), poly(vinyl chloride), and poly(vinylidene fluoride) are shown as PE, PP, PVC, and PVDF, respectively. Homopolymer fingerprints ($\mathbf{H}_i$) for the homopolymers (*i*) are computed using the polymer genome fingerprinting framework.[18] Copolymer fingerprints ($\mathbf{C}_j$) for the copolymers (*j*) are the composition-weighted sum of the homopolymer fingerprints ($\mathbf{H}_i$) with x_{ji} being the compositions of monomer *i* in copolymer *j*. Polymer blend fingerprints ($\mathbf{B}$) are the composition-weighted harmonic mean of the copolymer fingerprints with *M*, *N*, and w_j representing the number of monomers in copolymer *j*, the number of polymers, and their weight fraction in the polymer blend, respectively. (c) The inference pipeline to predict thermal properties for a new polymer blend.

because we do not have data that specify the particular archetype of the copolymer. However, if the copolymer type is known, an ML architecture trained by Tao et al. can be used to predict thermal properties of block and gradient copolymers.[36,37] Their framework considers the sequence of repeating units and relative composition of the monomers, which can be crucial for predicting thermal properties, whereas our approach only considers the relative composition of monomers. Polymer blends could be miscible or immiscible; our framework first predicts which category the polymer blend belongs to, and then, subsequently, predicts the appropriate number of critical temperatures (miscible polymer blends are characterized by one T_g, one T_m, and one T_d, whereas immiscible two-phase polymer blends could display two T_g, two T_m, and one T_d).

Our machine learning procedure starts by first converting the chemical structure of homopolymers, copolymers, and polymer blends into numerical vectors called fingerprints; the chemical structure itself is specified using SMILES strings[38] of the repeat units, the composition of the copolymers (if relevant), and the weight fraction of the polymer blend components (if relevant). These aspects are captured in Figure 1b. The fingerprints of the chemical structures, along with T_g, T_m, and T_d data are fed into our multitask neural network architecture,[29] with the ultimate output being a trained model that can predict whether a polymer blend (if the queried case is a polymer blend) is miscible or not, followed by the thermal properties of the queried case. This new polymer informatics capability is able to predict T_g, T_m, and T_d with a RMSE of 15 K, 17 K, and 23 K, respectively. Needless to say, this framework can be extended to handle any other property class, as long as the requisite data for these properties are available.

Results and discussion

Data

The data set used in this study for homopolymer, copolymer, and polymer blends T_g, T_m, or T_d were collected from multiple sources cited elsewhere,[18,29,39–41] including from the poLy-Info database.[41] (The copyrights of this database are owned by the National Institute for Materials Science [NIMS].) In order to maintain consistency, we only use T_g and T_m data points measured using differential scanning calorimetry (DSC), and T_d values measured via thermogravimetric analysis (TGA).[29] Each copolymer data point has two comonomers, and each polymer blend data point consists of two constituents (two homopolymers, a homopolymer–copolymer mixture, or two copolymers). We infer the polymer blend miscibility for each data point from the presence of one or two T_g values in the data set. This is also known as technological miscibility.[42] If the polymer blend data point has only one T_g value, we consider it miscible; otherwise, it is classified as immiscible. A three-component selector vector is used to indicate the property (T_g, T_m, or T_d) and miscibility (miscible or immiscible) of the data point. The property (for homopolymers, copolymers,
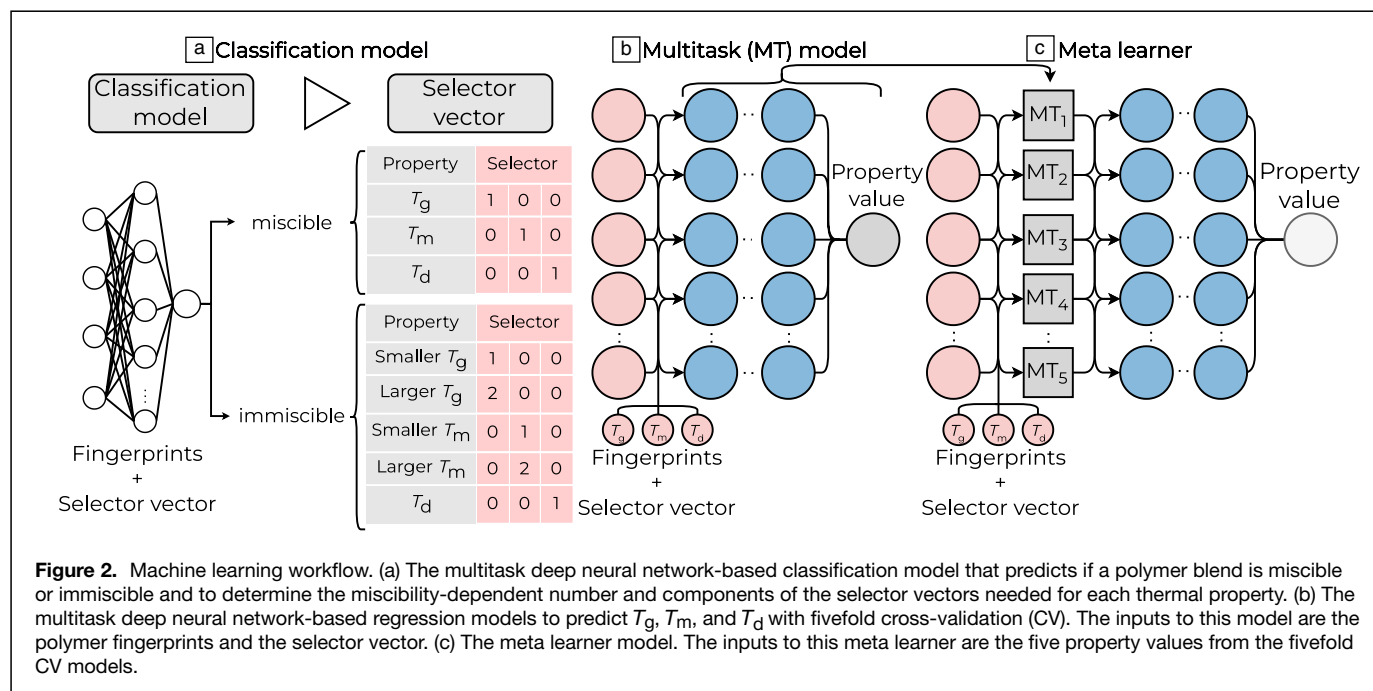
and polymer blends) is encoded in the selector vector at the position of the nonzero component as shown in **Figure 2**a. For polymer blends, the miscibility information is encoded at the appropriate component taking on values of 1 or 2 as shown in Figure 2a.

Fingerprinting

The repeat units of the monomers of homopolymers, copolymers, and polymer blends in the data set are represented using simplified molecular input line system (SMILES) strings.[38] We use stars [*] to denote the end points of the repeat unit. SMILES strings cannot directly be ingested by conventional ML models and require conversion to numerical vectors. This conversion is performed using a previously pioneered handcrafted fingerprinting scheme[20] (see the "Methods" section) that has shown great performance for predicting properties of polymers in many previous works.[18–20,23,28] For copolymers,[29] we compute fingerprints as the composition-weighted sum of the homopolymer fingerprint vectors ($\mathbf{C}_j = \sum_i^N x_{ji}\mathbf{H}_i$), as shown in Figure 1b. $\mathbf{H}_i$, x_{ji}, and N denote the fingerprint vector of a homopolymer (i), relative compositions of homopolymers in a copolymer (j), and the total number of comonomer components (in this work, $N = 1, 2$), respectively. For polymer blends, we use the composition-weighted harmonic mean of fingerprint vectors of the constituents in the polymer blend. To compute the polymer blend fingerprints, we use $1/\mathbf{B} + 1 = \sum_j^M w_j/\mathbf{C}'_j = \sum_{i,j}^{N,M} w_j/x_{ji}\mathbf{H}_i + 1$. Here, $\mathbf{B}$, w_j, $\mathbf{C}'_j$, and M represent the polymer blend fingerprint, the relative composition of the jth constituent, fingerprint vector of the jth polymer blend constituent, and the number of constituents in the polymer blend, respectively. This equation resembles the mathematical form of the Fox equation.[43] A scalar factor of one was added to the fingerprint vector components (and later subtracted) to avoid singularities caused by fingerprint components with the value of zero. The complete fingerprinting pipeline for homopolymers, copolymers, and polymer blends is shown in Figure 1b.

Performance

ML model development involved fivefold cross-validation (CV) and a meta learner, as described in the "Methods" section. The averaged performance scores of the five CV and the meta learner classification models are illustrated in **Figure 3**a. By comparing the precision (P), recall (R), accuracy (A), and F1 score, we find that P is generally higher than R. The reason for this is the imbalance of miscible and immiscible polymer blends (ratio $\approx 5/1$) in our data set (see **Table I**). The performance metrics of the classification meta learner improve from the fivefold cross-validation classification model for T_g as illustrated in Figure 3a. Figure 3b and Supplementary information Table S1 show the root-mean-square error (RMSE) values of the fivefold CV models of the property-predictive regression models. The low RMSE values for T_g, T_m, and T_d of homopolymers, copolymers, and polymer

Figure 2. Machine learning workflow. (a) The multitask deep neural network-based classification model that predicts if a polymer blend is miscible or immiscible and to determine the miscibility-dependent number and components of the selector vectors needed for each thermal property. (b) The multitask deep neural network-based regression models to predict T_g, T_m, and T_d with fivefold cross-validation (CV). The inputs to this model are the polymer fingerprints and the selector vector. (c) The meta learner model. The inputs to this meta learner are the five property values from the fivefold CV models.

blends provide confidence in the novel fingerprinting scheme for polymer blends and the usage of multitask models for this problem. All RMSE values for thermal properties are also of the same order of magnitude as experimental measurement errors. These RMSE values are also slightly better than the performance of homopolymers and copolymers as reported in our past publications.[21,28,29] For all types of polymers, RMSEs of fivefold CV are the lowest for T_g, followed by T_m, and then T_d.

The parity plots of meta learner predictions for all the thermal properties on different polymer data sets are shown in **Figure 4**. These meta learner predictions are based on the 80% data set used to train the cross-validation models. The low overall RMSE values (including all the types of polymers) of 15 K, 17 K, and 23 K, and high R^2 of values 0.98, 0.97, and 0.96 for T_g, T_m, and T_d, respectively, signify high

performance across the three distinct classes of polymers. Incorporating a hyperparameter-tuned meta learner on top of the cross-validation model further improves the performance of the multitask model as shown in Table S1. We use Monte Carlo dropout to estimate the uncertainty in our predictions and report them within a 95% confidence interval.[44] Our use of Monte Carlo dropout allows us to quantify the uncertainties in our predictions.

Sample prediction
Model predictions and experimentally measured values across the whole composition range for two selected miscible (**Figure 5**a–b) and two immiscible polymer blends (Figure 5c–d) are illustrated in Figure 5. The smoothness of the predictions across the composition range indicates that the ML model learned a smooth mapping between

Table I. The number of homopolymers, copolymers, and polymer blend data points for the glass-transition (T_g), melting (T_m), and degradation (T_d) temperatures.

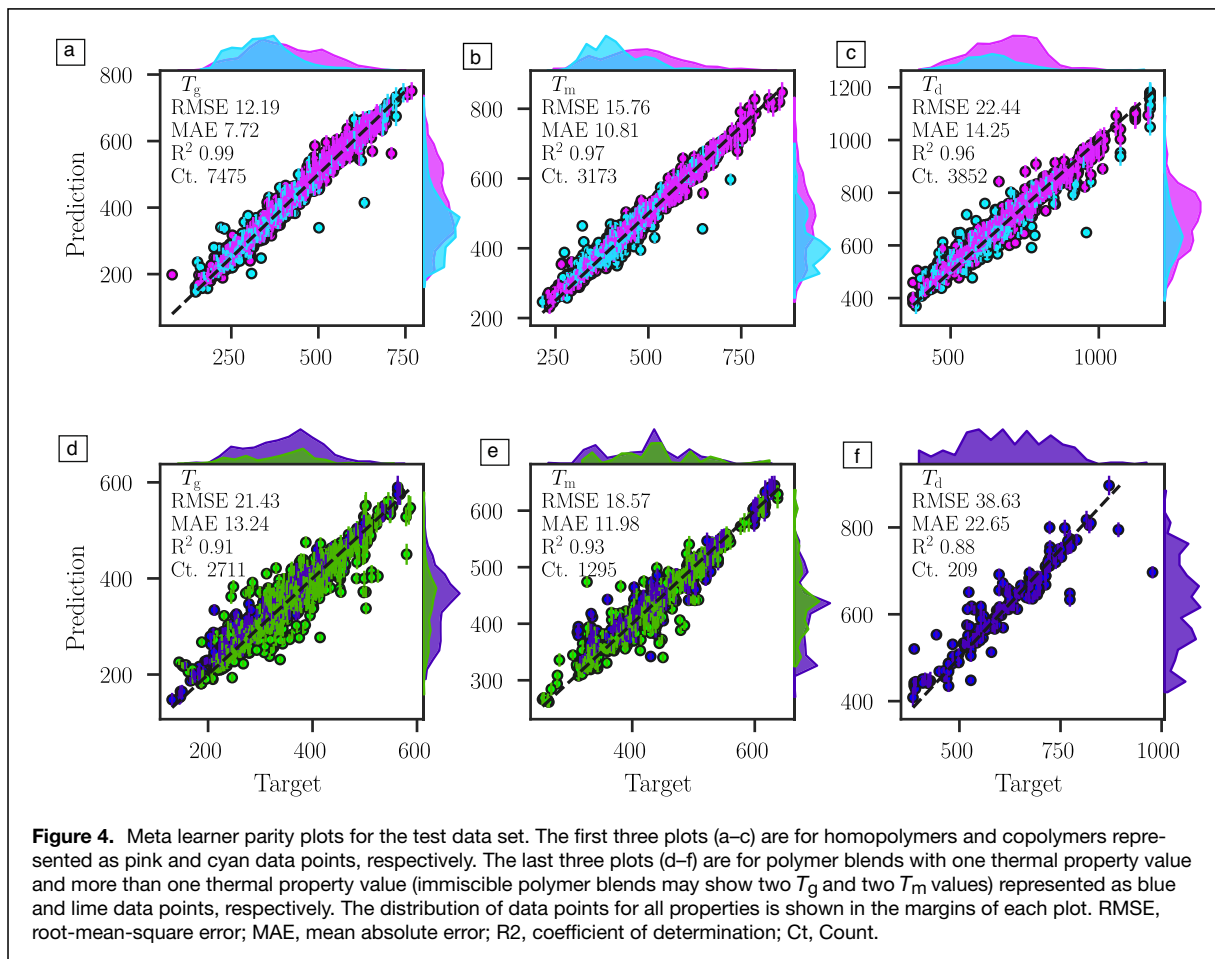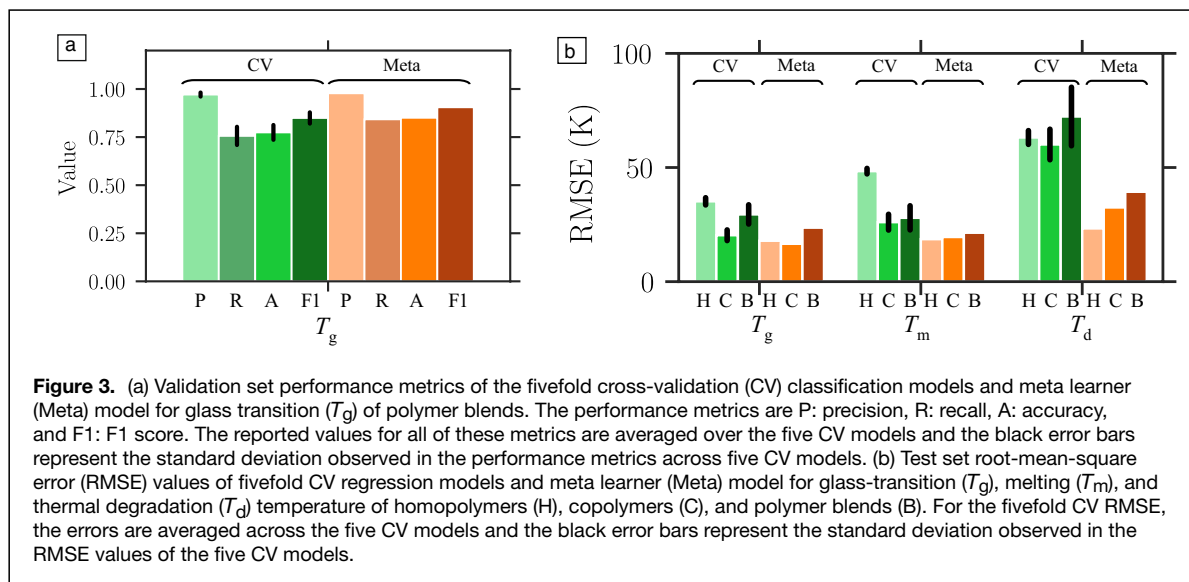| Property	Homopolymer	Copolymer	Polymer Blend		Range	Method	Total
			M	IM			
T_g	5072	4426	2541	440	[131, 587]	DSC	12,039
T_m	2079	1988	1001	330	[253, 639]	DSC	5068
T_d	3520	1360	261	–	[383, 977]	TGA	5141
Total	10,671	7774	3803	770			23,018

The 7774 copolymer data points consists of 1569 distinct copolymer chemistries. The 4573 polymer blend data points consist of 626 distinct combinations of two constituents (two homopolymers, two copolymers, or a homopolymer and a copolymer). Polymer blends can be miscible (M) or immiscible (IM). The reported measurement method, differential scanning calorimetry (DSC), or thermogravimetric analysis (TGA) is also indicated.
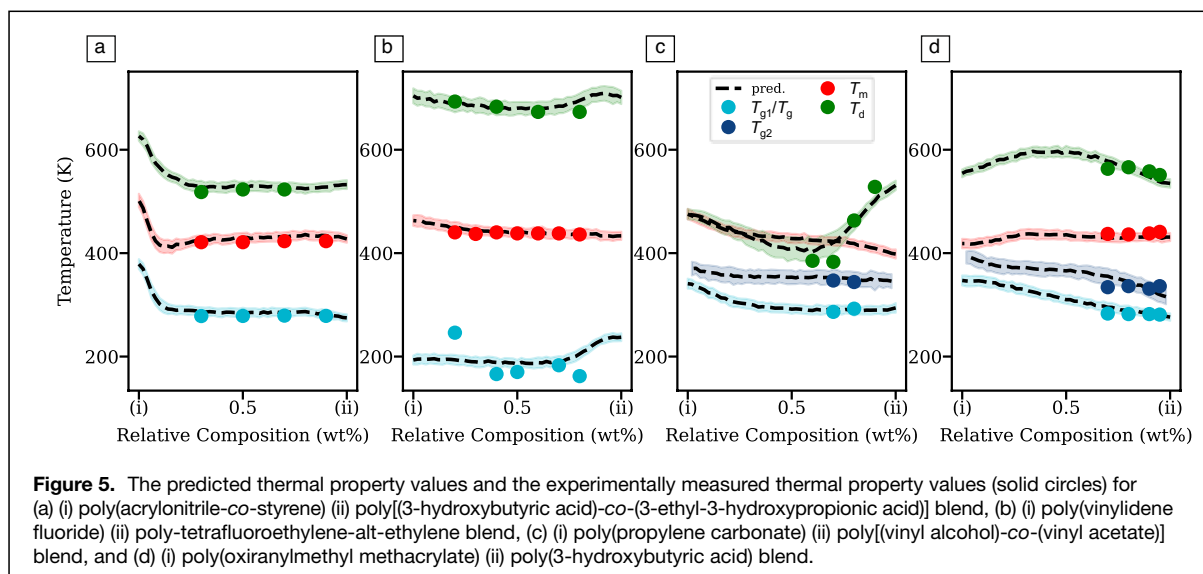
the fingerprint space and polymer properties. The experimental data points are in close agreement with the predictions, except for a few points in Figure 5b that fall outside the shaded bands, which indicate the uncertainty of the predictions.

Model validation

As a direct and true test of the generalizability of our trained meta learner model, we performed the following. The entire data set was divided into two parts. We used 90% of the data to train the fivefold cross-validation and the meta learner model,

Figure 3. (a) Validation set performance metrics of the fivefold cross-validation (CV) classification models and meta learner (Meta) model for glass transition (T_g) of polymer blends. The performance metrics are P: precision, R: recall, A: accuracy, and F1: F1 score. The reported values for all of these metrics are averaged over the five CV models and the black error bars represent the standard deviation observed in the performance metrics across five CV models. (b) Test set root-mean-square error (RMSE) values of fivefold CV regression models and meta learner (Meta) model for glass-transition (T_g), melting (T_m), and thermal degradation (T_d) temperature of homopolymers (H), copolymers (C), and polymer blends (B). For the fivefold CV RMSE, the errors are averaged across the five CV models and the black error bars represent the standard deviation observed in the RMSE values of the five CV models.

Figure 4. Meta learner parity plots for the test data set. The first three plots (a–c) are for homopolymers and copolymers represented as pink and cyan data points, respectively. The last three plots (d–f) are for polymer blends with one thermal property value and more than one thermal property value (immiscible polymer blends may show two T_g and two T_m values) represented as blue and lime data points, respectively. The distribution of data points for all properties is shown in the margins of each plot. RMSE, root-mean-square error; MAE, mean absolute error; R2, coefficient of determination; Ct, Count.

Figure 5. The predicted thermal property values and the experimentally measured thermal property values (solid circles) for (a) (i) poly(acrylonitrile-*co*-styrene) (ii) poly[(3-hydroxybutyric acid)-*co*-(3-ethyl-3-hydroxypropionic acid)] blend, (b) (i) poly(vinylidene fluoride) (ii) poly-tetrafluoroethylene-alt-ethylene blend, (c) (i) poly(propylene carbonate) (ii) poly[(vinyl alcohol)-*co*-(vinyl acetate)] blend, and (d) (i) poly(oxiranylmethyl methacrylate) (ii) poly(3-hydroxybutyric acid) blend.

Table II. The performance of the trained multitask (a) classification and (b) prediction meta learner model for the thermal properties on the testing data set.

a			b				
Property	T_g	T_m	Property		T_g	T_m	T_d
Accuracy	0.86	0.82	Homopolymer	R^2	0.90	0.84	0.67
				RMSE	33.45	45.42	63.13
Precision	0.94	0.95	Copolymer	R^2	0.95	0.84	0.73
				RMSE	19.03	30.21	50.16
Recall	0.94	0.80	Blend (Miscible)	R^2	0.79	0.85	0.51
				RMSE	33.71	26.79	66.71
F1 score	0.94	0.87	Blend (Immiscible)	R^2	0.73	0.88	
				RMSE	37.97	23.91	

while the remaining 10% was reserved for testing. This 10% subset of data was never seen by our model and was used to test both the classification and prediction models. The performance of the meta learner classification model on the testing data set is summarized in **Table II**a, demonstrating promising performance for both thermal properties. Similarly, Table IIb presents a summary of the meta learner prediction model's performance on the 10% testing data set. Additionally, the parity plots for all predictions can be found in Figure S2. Notably, our observations indicate that the prediction performance for homopolymers and copolymers surpassed that of polymer blends on the testing data set, primarily due to the larger availability of data points, consistent with the findings from our fivefold cross-validation model. Furthermore, we noticed variations in performance based on thermal properties, with the model demonstrating enhanced accuracy for T_g and T_m compared to T_d. Overall, the results suggest reasonable agreement between the predicted values and the measured values for all polymer classes and different thermal properties, although additional data could further improve the model's

performance. To expand the data set for polymer blends, simulated data for T_g can be employed. Previous studies have employed molecular dynamics simulations to generate T_g data for miscible blends.[45,46] Furthermore, features encompassing processing parameters, morphological details, and experimental conditions can be utilized for all thermal properties to train more accurate models.

Conclusion

In this study, we developed an ML framework capable of predicting properties of homopolymers, copolymers, and polymer blends simultaneously. Our framework demonstrates notable performance in terms of prediction accuracy and expands the chemical domain for this property class, showing improvements over previous works in terms of prediction accuracy and coverage.[29] Powered by multitask predictors and a large data set of 23,018 thermal data points, this framework enables the prediction of polymer properties that fall in a broad technologically relevant class. For the success of the model building, we

designed a polymer blend fingerprinting scheme that extends and is based on homopolymer and copolymer fingerprints that we have used in the past.

There are several ways in which this work can be utilized to further expand the capabilities of polymer informatics in terms of accuracy and scope. Incorporating the impact of processing parameters and morphology in the case of polymer blends can lead to improvements in prediction accuracy and versatility. The conceptual ideas used in this work to set up the ML pipeline for thermal properties prediction for different polymer flavors can be extended to other properties (e.g., electronic and mechanical) when suitable data are available. Our approach can be adapted to different polymer classes, such as polymers with additives or polymer composites, by encoding features corresponding to the additive and fillers in the fingerprinting step.

Methods

Polymer genome fingerprints

Polymer SMILES strings (e.g., `[*]CC[*]` for polyethylene) are converted to numerical fingerprint vectors using a handcrafted fingerprinting scheme. For homopolymers/monomers, previous works have shown that effective fingerprinting involves using three hierarchical levels of descriptors that span different length scales.[47,48] Starting from the atomic level, we determine the presence of predetermined atomic fragments or motifs, which are sequences of contiguous atoms.[21] Moving up the hierarchy, we employ larger length-scale descriptors that utilize quantitative structure–property relationships (QSPRs), such as van der Waals surface area, topological polar surface area (TPSA), fraction of atoms in rings, and fraction of rotatable bonds.[21] Finally, we use chain-level descriptors that characterize the physical structure of the polymer, including the shortest topological distance between rings, fraction of atoms in side chains, and length of the longest side chain.[21] These three levels of descriptors enable us to capture diverse and critical features of the polymers that are important for their properties. The benefits of these descriptors are that they have sufficient chemico-structural information to describe a wide range of physical and chemical attributes that control various polymer properties, they can distinguish between two different monomers, and they are invariant to different specifications of the polymer SMILES strings of the same polymer.

Multitask models and meta learner

The polymer fingerprints along with the selector vector and thermal property values for homopolymers, copolymers, and polymer blends are used to train our predictive ML models. Before training, the thermal property values are curated using min-max normalization to scale them to [0,1]. For the classification models, polymer fingerprints, selector vectors, and the number of thermal property values for T_g and T_m of polymer blends are used for training. The ML architecture is shown in Figure 2a, where the multitask model is a concatenation-based conditioned multitask deep neural network. This multitask model is trained on 80% of the data set through fivefold cross-validation (CV) and the remaining 20% is utilized to train the meta learner. We utilized TensorFlow[49] to implement all of our models. Adam optimization along with stochastic weight averaging is used for updating the weights of the network. The learning rate is initially set at 10^{-3} and is changed in the later phases of training through the learning rate scheduler along with early stopping to prevent overfitting. Hyperparameters of our ML model that include the number of layers, number of neurons in each layer, initial learning rate, dropout rates, and the layer where the selector vector is concatenated were tuned using the HyperBand algorithm implemented in the KerasTuner.[50] All values of the hyperparameters are present in Table S2. The ML architecture and the hyperparameter tuning steps are the same for the prediction and classification models.

The predicted thermal property values from the fivefold CV models are used as inputs to the meta learner as shown in Figure 2c. The predictive meta learner is an ensemble model that predicts the final thermal property value from the five thermal property values of the five CV models.[29] The meta learner for classifying miscibility is an ensemble model that predicts the final probabilities for miscibility of polymer blends. Both meta learners are trained on the remaining 20% of the data set that the cross-validation models have never seen. Similar to the multitask models, the hyperparameter optimization of the meta learner is achieved through the HyperBand algorithm implemented in the KerasTuner.[50]

Author contributions

S.S.S curated the data set and trained the ML models. C.K. provided assistance for training and evaluating the ML models. R.R. conceptualized the work and provided his guidance. All authors discussed the results and commented on the manuscript.

Data availability

The data sets used in this study are from PoLyInfo database https://polymer.nims.go.jp/en/. (The copyrights of this database are owned by the National Institute for Materials Science [NIMS].)

Code availability

The code is available at https://github.com/Ramprasad-Group/Polymer_informatics_beyond_homopolymers.

Competing interests

The authors declare no competing financial interest.

Supplementary information

The online version of this article contains supplementary material available at https://doi.org/10.1557/s43577-023-00561-0.

References

1. C. Hagiopol, *Copolymerization: Toward a Systematic Approach* (Springer, New York, 1999)
2. D.R. Paul, *Polymer Blends, Volume 1* (Elsevier, Amsterdam, 2012)
3. V. Ambrogi, C. Carfagna, P. Cerruti, V. Marturano, "Additives in Polymers," in *Modification of Polymer Properties,* ed. by C.F. Jasso-Gastinel, J.M. Kenny (Elsevier, Amsterdam, 2017), chap. 4, pp. 87–108. https://doi.org/10.1016/B978-0-323-44353-1.00004-X
4. D.R. Paul, J.W. Barlow, *J. Macromol. Sci. Part C* **18**, 109 (1980). https://doi.org/10.1080/00222358008080917
5. A.N. Singh, R.D. Thakre, J.C. More, P.K. Sharma, Y.K. Agrawal, *Polymer* **54**, 1077 (2015)
6. P.N. Bartlett, P.R. Birkin, *Synth. Met.* **61**, 15 (1993)
7. S.P. Nunes, *Macromolecules* **49**, 2905 (2016)
8. Y. Sun, G. Shi, *J. Polym. Sci. B Polym. Phys.* **51**, 231 (2013)
9. L.O. Ekebafe, D.E. Ogbeifun, F.E. Okieimen, *Biokemistri* **23**(2), 81 (2011)
10. F. Puoci, F. Iemma, U.G. Spizzirri, G. Cirillo, M. Curcio, N. Picci, *Am. J. Agric. Biol. Sci.* **3**, 299 (2008)
11. M. Rasek, J. Drozd, *Pol. J. Appl. Chem.* **53**(3), 267 (2009)
12. L.M. Robeson, *Polym. Eng. Sci.* **24**, 587 (1984)
13. O. Okay, *Prog. Polym. Sci.* **25**, 711 (2000)
14. E. Elhefian, *J. Chem. Soc. Pak.* **36**, 11 (2014)
15. L. Yu, K. Dean, L. Li, *Prog. Polym. Sci.* **31**, 576 (2006)
16. P. Pötschke, D.R. Paul, *J. Macromol. Sci. Part C* **43**, 87 (2003)
17. K. Hamad, M. Kaseem, Y.G. Ko, F. Deri, *Polym. Sci. Ser. A* **56**, 812 (2014)
18. C. Kim, A. Chandrasekaran, T.D. Huan, D. Das, R. Ramprasad, *J. Phys. Chem. C* **122**, 17575 (2018)
19. A. Chandrasekaran, C. Kim, R. Ramprasad, "Polymer Genome: A Polymer Informatics Platform to Accelerate Polymer Discovery," in *Machine Learning Meets Quantum Physics*, ed. by K.T. Schütt, S. Chmiela, O.A. von Lilienfeld, A. Tkatchenko, K. Tsuda, K.-R. Müller, Lecture Notes in Physics Book Series, vol. 968 (Springer, Cham, 2020), pp. 397–412. https://doi.org/10.1007/978-3-030-40245-7_18
20. H. Doan Tran, C. Kim, L. Chen, A. Chandrasekaran, R. Batra, S. Venkatram, D. Kamal, J.P. Lightstone, R. Gurnani, P. Shetty, M. Ramprasad, J. Laws, M. Shelton, R. Ramprasad, *J. Appl. Phys.* **128**, 171104 (2020)
21. C. Kuenneth, A.C. Rajan, H. Tran, L. Chen, C. Kim, R. Ramprasad, *Patterns* **2**, 100238 (2021)
22. G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, R. Ramprasad, *Sci. Rep.* **3**, 2810 (2013)
23. R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, *NPJ Comput. Mater.* **3**, 54 (2017)
24. S. Wu, Y. Kondo, M.-A. Kakimoto, B. Yang, H. Yamada, I. Kuwajima, G. Lambard, K. Hongo, Y. Xu, J. Shiomi, C. Schick, J. Morikawa, R. Yoshida, *NPJ Comput. Mater.* **5**, 66 (2019)
25. C. Yan, X. Feng, C. Wick, A. Peters, G. Li, *Polymer* **214**, 123351 (2021)
26. J.W. Barnett, C.R. Bilchak, Y. Wang, B.C. Benicewicz, L.A. Murdock, T. Bereau, S.K. Kumar, *Sci. Adv.* **6**, eaaz4301 (2020)
27. M.-X. Zhu, H.-G. Song, Q.-C. Yu, J.-M. Chen, H.-Y. Zhang, *Int. J. Heat Mass Transf.* **162**, 120381 (2020)
28. L. Chen, G. Pilania, R. Batra, T.D. Huan, C. Kim, C. Kuenneth, R. Ramprasad, *Mater. Sci. Eng.* **144**, 100595 (2021)
29. C. Kuenneth, W. Schertzer, R. Ramprasad, *Macromolecules* **54**, 5957 (2021)
30. Y. Zhang, X. Xu, *J. Mol. Graph. Model.* **103**, 107796 (2021)
31. G. Pilania, C.N. Iverson, T. Lookman, B.L. Marrone, *J. Chem. Inf. Model.* **59**, 5013 (2019)
32. Z. Jiang, J. Hu, B.L. Marrone, G. Pilania, X.B. Yu, *Materials* (Basel) **13**(24), 5701 (2020). https://doi.org/10.3390/ma13245701
33. S. Goswami, R. Ghosh, A. Neog, B. Das, *Mater. Today* **46**, 5838 (2021)
34. Y. Zhang, X. Xu, *Mol. Cryst. Liq. Cryst.* **730**, 9 (2021)
35. M.-X. Zhu, Q.-C. Yu, H.-G. Song, T.-X. Chen, J.-M. Chen, *ACS Appl. Energy Mater.* **4**, 1449 (2021). https://doi.org/10.1021/acsaem.0c02647
36. L. Tao, J. Byrnes, V. Varshney, Y. Li, *iScience* **25**, 104585 (2022)
37. L. Tao, T. Arbaugh, J. Byrnes, V. Varshney, Y. Li, *STAR Protoc.* **3**, 101875 (2022)
38. D. Weininger, *J. Chem. Inf. Comput. Sci.* **28**, 31 (1988)
39. C. Kim, A. Chandrasekaran, A. Jha, R. Ramprasad, *MRS Commun.* **9**, 860 (2019)
40. A. Jha, A. Chandrasekaran, C. Kim, R. Ramprasad, *Model. Simul. Mater. Sci. Eng.* **27**, 24002 (2019)
41. PoLyInfo. https://polymer.nims.go.jp/en/
42. E. Manias, L.A. Utracki, "Thermodynamics of Polymer Blends," in *Polymer Blends Handbook*, ed. by L.A. Utracki, C.A. Wilkie (Springer, Dordrecht, 2014), pp. 171–289. https://doi.org/10.1007/978-94-007-6064-6_4
43. W. Brostow, R. Chiu, I.M. Kalogeras, A. Vassilikou-Dova, *Mater. Lett.* **62**, 3152 (2008)
44. Y. Gal, Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," *33rd International Conference on Machine Learning, ICML 2016* (New York, June 19–24, 2016), pp. 1651–1660
45. A.D. Glova, S.G. Falkovich, D.I. Dmitrienko, A.V. Lyulin, S.V. Larin, V.M. Nazarychev, M. Karttunen, S.V. Lyulin, *Macromolecules* **51**, 552 (2018)
46. J. Li, S. Jin, G. Lan, S. Chen, L. Li, *J. Mol. Graph. Model.* **84**, 182 (2018)
47. H. Doan Tran, C. Kim, L. Chen, A. Chandrasekaran, R. Batra, S. Venkatram, D. Kamal, J.P. Lightstone, R. Gurnani, P. Shetty, M. Ramprasad, J. Laws, M. Shelton, R. Ramprasad, *J. Appl. Phys.* **128**(17), 171104 (2020). https://doi.org/10.1063/5.0023759
48. A. Mannodi-Kanakkithodi, G. Pilania, T.D. Huan, T. Lookman, R. Ramprasad, *Sci. Rep.* **6**, 20952 (2016)
49. A. Martin, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefow, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems (2015). https://www.tensorflow.org/
50. T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, L. Invernizzi, KerasTuner (2019). https://github.com/keras-team/keras-tuner ☐