General Paper

Materials Informatics Approach to Predictive Models for Elastic Modulus of Polypropylene Composites Reinforced by Fillers and Additives

Yuko IKEDA,^a Michihiro OKUYAMA,^a Yukihiro NAKAZAWA,^a Tomohiro OSHIYAMA,^{*a}
Kimito FUNATSU^b

^a KONICA MINOLTA, Inc., 2970 Ishikawa-machi, Hachioji-shi, Tokyo 192-8505, Japan

^b Department of Chemical System Engineering, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656, Japan

e-mail: tomohiro.oshiyama@konicaminolta.com

(Received: July 31, 2020; Accepted for publication: January 5, 2021; Online publication: June 11, 2021)

Advanced processes are useful when developing polymer composites because there are an enormous number of possible combinations of fillers and additives to realize polymers with desired properties. Materials informatics is a data-driven approach to find novel materials or a suitable combination of materials from material data sheets. Here, we used materials informatics to construct a predictive model for the elastic modulus of polypropylene composites. To apply materials informatics to existing experimental data, we described explanatory variables by a combination of 0 and 1 representing polypropylene, or by the content ratio of filler and additive, without using materials property data. We constructed a predictive model for the elastic modulus of polypropylene composites using a partial least square regression model with dummy variables. To validate the predictive model, comparisons were made between measured and predicted elastic moduli for eight new polypropylene composites. The residual was less than 300 MPa for the range 1,000–3,000 MPa. We improved the accuracy of the prediction for composites with high filler content ratio by applying a nonlinear support vector regression model. The predictive model is therefore useful for identifying suitable combinations of polypropylene, filler and additive to achieve a desired elastic modulus.

Keywords: Materials informatics, Elastic modulus, Polypropylene composite, PLS, SVR, Dummy variables

1 INTRODUCTION

High performance polymer composites reinforced by fillers and additives are light weight and high strength and are therefore attractive in manufacturing. Carbon fiber reinforced polymer composites are increasingly used in place of traditional materials such as steel and aluminum alloys in aircraft and automobile industries. To realize the desirable mechanical properties of polymer composites, the content ratio of materials and process parameters should be controlled. However, experiments with all possible combinations of polymers, fillers and additives are impractical. Designing a materials composition based on a researchers' experience and intuition is costly and time consuming. Materials informatics (MI) has received considerable recent attention as a data-driven approach for developing materials. MI has been applied to develop inorganic solid electrolyte materials and

a spin-driven thermoelectric material with a high thermopower [1–4]. MI has also been used to develop organic materials such as efficient light emitting materials [5], and to predict the solubility of organic solvents [6]. MI has been far less applied to polymers and polymer composites. This is because of the difficulty in collecting data sets of polymer structures and properties based on similar experimental conditions. Another reason is that there are many commercial brands based on the same monomers but with different average molecular weights. Processing parameters such as temperature, time, pressure and mixing conditions also have a large influence on the experimental results.

There have been pioneering studies based on polymer MI. Polymers with high glass transition temperatures (T_g) were proposed through machine learning using experimental T_g data sets from the polymer database “PoLyInfo” provided by the National Institute for Materials Science [7]. In the analysis, monomer structures in

Open Access This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License. <http://creativecommons.org/licenses/by-nc/4.0/>

polymers were converted into fingerprint descriptors, i.e., digitized information for molecular structures [8]. Elastic moduli of polymer composites comprising polypropylene and talc were also analyzed using a classification approach. The authors used experimental data sets from the literature and specified properties that contributed toward a high elastic modulus [9].

High-performance polymer structures have been proposed based on data-driven approaches using the quantitative structure-activity relationship method [10, 11], Bayesian optimization [12] and transfer learning [13]. In these cases, MI was applied according to the characteristics of the polymer and polymer composites and the quality of the data sets, and high quality results were obtained.

In the current study, MI was applied to develop polypropylene composites and to propose predictive models for elastic moduli of polypropylene composites reinforced by fillers and additives [14]. Two procedures were performed. First, a predictive model for elastic moduli was constructed using supervised experimental data sets with a regression model. Second, the model was used to list all possible material compositions and to propose compositions with desirable elastic moduli. These compositions were then verified experimentally.

The first procedure was complicated by commercial polypropylene homo-polymers having various average molecular weights, monomer compositions in copolymer and tacticities. The average molecular weight and also the molecular weight distribution both effect the physical and mechanical properties of polymers. Experimentally collecting these parameters is hugely time consuming. To overcome this and exploit entire experimental data sets, we used descriptors comprising brand names and content ratios of each material as explanatory variables. The second section describes the setting of descriptors based on the combination and content ratio of polypropylenes, fillers and additives. The procedure and results for constructing a predictive model for elastic moduli using a partial least square (PLS) regression model [15] are also described. The third section discusses the applicability of the prediction model. Improving the accuracy of the predictive model for composites with high filler content ratio using a nonlinear support vector regression (SVR) model is described.

2 CONSTRUCTION OF PREDICTIVE MODELS FOR ELASTIC MODULUS

2.1 Description of explanatory variables using dummy variables derived from brand names of materials and content ratio

In this study, 180 experimental data sets of elastic moduli comprising 11 polypropylenes, 18 fillers and 20 additives were used. These materials are all commercially available. In many cases, thermophysical parameters of polymers are partially missing and detailed polymer structures are unknown. Therefore, we described the explanatory variables by a combination of 0 and 1 representing polypropylene, filler and additive or by the content ratio of the filler and additive. These dummy variables enabled us to use entire experimental data sets.

x_{p_i} was defined as the descriptor of polypropylenes which was

described in equation (1) and p_i was defined as the polypropylene brand name, where $i = 1$ to 11.

$$x_{p_i} = \begin{cases} 1 (i = \alpha) \\ 0 (i \neq \alpha) \end{cases} \quad (1)$$

x_{f_i} and x_{a_i} were defined as the descriptor of fillers and additives which were described in equations (2) and (3), respectively. f_i was defined as the filler brand name, where $i = 1$ to 18. a_i was defined as the additive brand name, where $i = 1$ to 20. c_{f_i} and c_{a_i} represent the filler and additive content ratio, respectively.

$$x_{f_i} = \begin{cases} c_{f_i} (i = \alpha) \\ 0 (i \neq \alpha) \end{cases} \quad (2)$$

$$x_{a_i} = \begin{cases} c_{a_i} (i = \alpha) \\ 0 (i \neq \alpha) \end{cases} \quad (3)$$

Using equations (1), (2) and (3), explanatory variables were set by the vector representation which showed the brand names and content ratios in the polymer composites. For example, in the vector representation of equation (4), the brand names of the polypropylene, filler and additive are p_1 , f_1 and a_{20} , respectively, and the c_{f_1} and $c_{a_{20}}$ values are 10% and 5%, respectively.

$$\mathbf{x} = \begin{bmatrix} x_{p_1} \\ \vdots \\ x_{p_{11}} \\ x_{f_1} \\ \vdots \\ x_{f_{18}} \\ x_{a_1} \\ \vdots \\ x_{a_{20}} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 0 \\ 10 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 5 \end{bmatrix} \quad (4)$$

2.2 Procedure of constructing predictive model for elastic moduli

The predictive model for the elastic moduli of polypropylene composites was constructed using PLS regression based on the 49 explanatory variables in equations (1), (2) and (3). All 180 data sets were used to construct the PLS model. A machine learning library of scikit-learn in Python3 was used in the PLS regression. To avoid multicollinearity problems among the explanatory variables, an orthogonal set of latent variables was used.

The accuracy of the predictive model was evaluated using data sets obtained in our experiments. Leave-one-out cross-validation (LOOCV) was performed to validate the PLS model. In LOOCV, the predicted model was constructed using 179 data sets as training data sets, and the performance was tested on data set number 180.

Figure 1. Relationship between predicted t and number of latent variables.

Figure 2. Relationship between measured elastic modulus and modulus predicted by LOOCV using four latent variables.

Tests of all 180 data sets by repeating this procedure 180 times confirmed the generalization of the PLS model. In LOOCV, the number of an orthogonal set of latent variables was determined by the minimum value of the root mean square error (RMSE, t) represented by equation (5).

$$t = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (5)$$

In equation (5), y_i is an observed value in the i -th test data and $\hat{y}_i$ is a predicted value in the i -th test data for the prediction model using 179 training data sets except for the i -th data set. **Figure 1** shows the relationship between the predicted t and number of latent variables. The minimum t was 309 MPa when the number of latent variables was four and the R-squared value was 0.73.

Figure 2 shows the relationship between the measured elastic modulus and the modulus predicted by LOOCV using four latent variables. Although the PLS model constructed by LOOCV showed a relatively large residual at high modulus, the 49 dummy variables yielded a moderately accurate model. It was necessary to clarify the applicable domain of the model because the data density at high elastic modulus was low.

Figure 3. Frequency distribution of absolute residuals.

3 CONSIDERATION OF PLS PREDICTIVE MODEL FOR ELASTIC MODULI OF POLYPROPYLENE COMPOSITES

The accuracy of the predictive model was validated by visualizing the PLS residuals. The reason for the large residual at low predictive accuracy required investigating to improve the model.

3.1 Visualization of residual of predictive model

To confirm the applicability of the predictive PLS model, the difference between observed and predicted values were obtained. Absolute values of the residuals were obtained as shown in **Figure 3**.

Figure 1 shows that the t was 309 MPa, so the percentage of residuals less than 300 MPa was analyzed. The percentage of data having residuals of 300 MPa or less was 90% at < 2,500 MPa and 15% at ≥ 2,500 MPa. The predictive accuracy was low at high modulus because of the low data density. It was therefore necessary to increase the number of experimental data sets to improve the predictive accuracy.

3.2 Generation of virtual material compositions

We then verified the accuracy of the PLS regression by comparing the experimental and predicted elastic moduli for polypropylene composites. Using 11 polypropylenes, 18 fillers and 20 additives (content ratios are shown in **Table 1**) resulted in 575,484 virtual compositions. The predicted elastic moduli for the possible polypropylene composites was therefore obtained using the PLS regres-

Table 1. Number of virtual material compositions.

Material	Polypropylene	Filler	Additive
Number of Materials ^{a)}	11 + 1	18 + 1	20 + 1
Candidates ^{b)}	0,1	0,1,2,2.5,3,5,6,10,20,30,40,50,60	0,0.1,1,2,3,4,5,6,10,20,30,40
Number of candidates ^{c)}	${}_{12}C_1$	${}_{18}C_1 \times 12 + 1$	${}_{20}C_1 \times 11 + 1$

^{a)} One was added to the number of materials as the case where no polypropylene, filler and additive were included.

^{b)} For the polypropylene, the brand name was represented by a combination of 0 or 1. For the filler and additive, the content ratios (wt.%) were described.

^{c)} Number of virtual material compositions was calculated by: ${}_{12}C_1 \times ({}_{18}C_1 \times 12 + 1) \times ({}_{20}C_1 \times 11 + 1) = 575,484$

Figure 4. Data for 575,484 virtual compositions (green dots) and 180 experimental supervised data sets. The color bar indicates the predictive elastic modulus in MPa. **Figure 6.** Euclid distances between experimental candidates and most neighboring data set. **Figure 5.** Zoomed in area around the supervised data sets in **Figure 4** (blue dots), and experimental candidates A to H (orange dots).

sion. Based on the predicted values, eight experimental samples were selected to satisfy a variety of data and material types.

The following procedure was used to select experimental candidates. The data structures of the 575,484 virtual compositions and 180 experimental supervised data sets were visualized as shown in **Figure 4**. We used three latent variables in the visualization because the t was 335 MPa, which was close to the minimum t . The visualized results are shown in **Figure 4**. The virtual compositions were more widely distributed than the supervised data sets. Experimental candidates were selected in the vicinity of the supervised data sets because of the high predictive accuracy. **Figure 5** shows the zoomed in area around the supervised data sets in **Figure 4**.

Using three latent variables, the Euclid distances between the experimental candidates and most neighboring supervised data set were quantified as shown in **Figure 6**. The predicted values of the experimental candidates A to H are shown in **Table 2**. Experimental candidates A to G were selected where high accuracy was expected in the prediction. Experimental candidate H was selected where a high elastic modulus was expected, despite the data point deviating from the supervised data sets.

The selected polypropylene, filler, additive and content ratios are shown in **Table 2**. The polypropylenes, fillers and additives were commercially available products with unknown molecular structures. We set the representation of explanatory variables by brand name and content ratio and calculated the predictive accuracy using the dummy variables.

3.3 Experimental procedure

The 180 supervised experimental data sets and experimental data sets for the eight new samples were obtained as follows. The mixtures of polypropylenes, fillers and additives were kneaded by a biaxial kneader (MC15, Xplore Instruments) to obtain polypropylene composites. The kneading temperature was 200 °C and the rotational speeds were 80 rpm and 130 rpm before and after the intake of the material, respectively. Kneading was continued for 5 min. Molded samples were produced by an injection molding machine (IM12, Xplore Instruments). The molded form was under the conditions of ISO527-2-1BA. During molding, the cylinder and mold temperatures were 200 °C and 60 °C, respectively. The injection pressure was 10–15 bar and the injection time was 18 s. The elastic moduli of the obtained polypropylene composites were

Table 2. Selected experimental candidates and content ratios.

Exp.	Polypropylene ^{a)}	Filler ^{a)}	Content ratio/wt.%	Additive ^{a)}	Content ratio/wt.%	Predicted value/ MPa
A	p ₃	f ₇	2.5	a ₁₄	1	1002
B	p ₄	f ₈	2	a ₁₆	1	1620
C	p ₄	f ₁₆	10	a ₇	6	1905
D	p ₅	f ₁₇	10	a ₈	0.1	2446
E	p ₄	f ₄	20	a ₁₄	40	2128
F	p ₁₀	f ₁₅	20	a ₃	0.1	3057
G	p ₅	f ₁₈	20	a ₇	6	3052
H	p ₅	f ₁₅	60	a ₁₆	10	3955

^{a)} Four polypropylenes (p₃, p₄, p₅, p₁₀), seven fillers (f₄, f₇, f₈, f₁₅, f₁₆, f₁₇, f₁₈) and five additives (a₃, a₇, a₈, a₁₄, a₁₆) were selected as experimental candidates.

Figure 7. Comparison between observed and elastic modulus.

evaluated by a universal testing machine (TENSILON RTC1250A, A&D Co. Ltd.) under a crosshead speed of 1 mm/min at an initial load of 0.3 N.

3.4 Specification for applicability by additional experiments

We verified the accuracy of the model using the eight experimentally prepared polypropylene composites. **Figure 7** compares the measured and predicted elastic moduli for these eight composites. Experimental samples A to G had predicted values from 1,000 MPa to 3,148 MPa, and the residuals were within 300 MPa, which were close to the value of t (309 MPa) obtained by LOOCV. This indicated that the predictive model was accurate in the range of 1,000 MPa to 3,000 MPa.

The predictive accuracy was low at high elastic modulus, as evidenced by the residual for experimental sample H of 2,176 MPa. The reason for this low accuracy at high elastic modulus is discussed in the following section.

3.5 Investigation of low predictive accuracy at high elastic modulus

To investigate the reason for the low predictive accuracy at high elastic modulus, we focused on the single data point with the largest residual of 1,619 MPa among the 180 data sets in **Figure 2**. This data are hereafter referred to as δ . Elucidating the reason for

Figure 8. Relationship between measured and PLS-predicted elastic modulus and content ratio of f₅.

the decrease in predictive accuracy of δ was anticipated to lead to improvements in the predictive model.

The filler content of f₅ was 40%, which was a high content ratio compared with those of the other polypropylene composites. Increasing the filler content ratio generally leads to an increase in elastic modulus. However, this increase in elastic modulus has been reported to plateau and the decrease for high filler content ratios [16].

We extracted sets from the 180 experimental data sets whose filler contents of f₅ were 40% (δ), 30%, 20%, 10%, and 1%. The median elastic modulus was used in the case of multiple data sets using different polypropylenes in the same content ratio. **Figure 8** shows the measured and PLS-predicted values of elastic modulus, and the relationship between the median value and content of f₅.

Figure 8 shows that the predicted elastic modulus had a linear relationship with the content ratio of filler f₅. The measured elastic modulus deviated from this linear relationship. **Figure 8** shows that the observed value was extremely low compared with the experimental value when the f₅ content ratio was 40%. There was an optimal filler concentration and the optimal value differed between fillers because of the increasing viscosity, component incompatibility and other factors. To predict the optimal filler content, it was necessary to improve the predictive accuracy at high elastic modulus.

Table 3. Candidates and optimized values of hyper parameters. ϵ was fixed to 0.1 in the linear-SVR model.

Kernel	Linear		RBF		
Hyper parameter	C	ϵ	C	γ	ϵ
Candidates	$2^{-10}, 2^{-9}, \dots, 2^9, 2^{10}$	-	$2^{-5}, 2^{-4}, \dots, 2^9, 2^{10}$	$2^{-20}, 2^{-19}, \dots, 2^9, 2^{10}$	$2^{-10}, 2^{-9}, \dots, 2^{-1}, 2^0$
Number of candidates	21	-	16	31	31
Optimized value	0.125	0.1	32	0.125	0.000977

Figure 9. Relationship between measured elastic modulus and elastic modulus predicted by the linear kernel function (left) and RBF kernel function (right) in SVR. **Figure 10.** Results from the linear SVR function (left) and RBF SVR function (right) analyzed by filler content. Color bars indicate filler content (%).

3.6 Improved predictive accuracy at high elastic modulus by nonlinear regression model

A nonlinear SVR model was used to improve the predictive accuracy at high elastic modulus [17]. An RBF kernel was used for nonlinear analysis and a linear kernel was used for comparison.

A machine learning library of scikit-learn in Python3 was used in the SVR. The 180 data sets were divided into training and test data sets (75% and 25%, respectively) similarly to the case for PLS regression. The SVR model was validated by 5-fold cross-validation. The fast algorithm was used to search for the three hyper parameters (C , γ , ϵ) [18]. Candidates and optimized values

are shown in Table 3.

In the linear SVR model, the predictive value for the 40% filler content of f_5 increased linearly with the elastic modulus, which was similar to the result for the PLS regression. The nonlinear RBF SVR model improved the accuracy at high elastic modulus as shown in Figure 9. The SVR results were analyzed to confirm the influence of filler content as shown in Figure 10. In the linear kernel function, the data sets for high filler content deviated from the linear relationship. In the nonlinear RBF SVR, the correlation between the observed and predicted elastic modulus was improved at high elastic modulus. Table 4 also shows R-squared and RMSE results

Table 4. R-squared and RMSE results for the SVR models.

Method	R-squared (MPa)	RMSE (MPa)
SVR/Linear	0.64	309
SVR/RBF	0.71	278

Figure 11. Relationship between measured and SVR-predicted elastic modulus and content ratio of f_5 for the linear SVR function (left) and nonlinear SVR function (right).

for the nonlinear RBF model. **Figure 11** shows that the decrease in elastic modulus at 40% filler content of f_5 was reproduced. The nonlinear model therefore improved the predictive accuracy at high elastic modulus.

4 CONCLUSION

MI was applied to predictive models for the elastic moduli of polypropylene composites reinforced by fillers and additives. In analyzing 180 experimental data sets, the explanatory variables were described by a combination of 0 and 1 representing polypropylene, or by the content ratio of fillers and additives. A predictive model for the elastic moduli of polypropylene composites was constructed using PLS regression with dummy variables and validated by LOOCV. Additional experiments were conducted to confirm the accuracy of the predictive model. The residual was less than 300 MPa in a range of 1,000 MPa to 3,000 MPa. The residual was close to t (309 MPa) in LOOCV, which indicates that the prediction model was accurate. A nonlinear SVR model was applied for high filler content ratio to improve the accuracy.

The predictive model based on the brand names and content ratio of experimental data sets as explanatory variables was therefore effective. It is often difficult to describe explanatory variables of polymer structures, and the MI approach used here overcomes this problem.

The effectiveness of dummy variables was shown in the data analysis as is the case with continuous variables [19]. Fingerprint descriptors were also used as variables for MI [20]. Our study shows that the approach is practical in designing the polymer composites to achieve a desired elastic modulus. The Bayesian optimization is another approach for an efficient search of desired property of polymers. We would like to compare the results with other meth-

ods in the next step. The MI-based predictive model can rapidly select a suitable combination of polypropylene, filler and additive to achieve a desired elastic modulus. Our further investigations involve increasing the number of data points, especially at high elastic modulus, to establish a more accurate predictive model.

SUPPLEMENTARY MATERIALS

The 180 supervised experimental data sets used in this study were shown in Table S1.

ACKNOWLEDGMENTS

We thank Aidan G. Young, Ph.D., from Edanz Group (<https://en-author-services.edanzgroup.com/ac>) for editing a draft of this manuscript.

REFERENCES

- [1] R. Jarem, K. Kanamori, I. Takeuchi, M. Nakayama, H. Yamasaki, T. Saito, *Sci. Rep.*, **8**, 5845 (2018). DOI:10.1038/s41598-018-23852-y, PMID:29643423
- [2] F. Ren, L. Ward, T. Williams, K. J. Laws, C. Wolverton, J. Hattrick-Simpers and A. Mehta, *Sci. Adv.*, **4**, eaaq1566 (2018).
- [3] R. Yuan, Z. Liu, P. V. Balachandran, D. Xue, Y. Zhou, X. Ding, *Adv. Mater.*, **30**, 1702884 (2018). DOI:10.1002/adma.201702884
- [4] Y. Iwasaki, R. Sawada, V. Stanev, M. Ishida, A. Kirihaara, Y. Omori, H. Someya, I. Takeuchi, E. Saitoh and S. Yoroazu, *npj Comput Mater* **5**, 103 (2019).

- [5] R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, D. Duvenaud, D. Maclaurin, M. A. Blood-Forsythe, H. S. Chae, M. Einzinger, D. G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. Huang, S. I. Hong, M. Baldo, R. P. Adams, A. Aspuru-Guzik, *Nat. Mater.*, **15**, 1120 (2016). DOI:10.1038/nmat4717 , PMID:27500805
- [6] A. Suzuki, Y. Kikura, K. Tanaka, K. Funatsu, *J. Comput. Chem. Jpn.*, **19**, 1 (2018).
- [7] PoLyInfo, <http://polymer.nims.go.jp>
- [8] H. Yamada, W. Stephen, C. Liu, R. Yoshida, Japanese Joint Statistical Meeting, 19, 1 (2018).
- [9] M. McBride, N. Persson, E. Reichmanis, M. Grover, *Processes (Basel)*, **6**, 79 (2018). DOI:10.3390/pr6070079
- [10] S. Goto, M. Arakawa, K. Funatsu, *J. Comput. Aided Chem.*, **10**, 30 (2009). DOI:10.2751/jcac.10.30
- [11] S. Takano, H. Kaneko, *J. Comput. Chem. Jpn.*, **18**, 115 (2019). DOI:10.2477/jccj.2019-0004
- [12] T. Minami, M. Kawata, T. Fujita, K. Murofushi, H. Uchida, K. Omori, Y. Okuno, *MRS Adv.*, **4**, 1125 (2019). DOI:10.1557/adv.2019.57
- [13] H. Yamada, C. Liu, S. Wu, Y. Koyama, S. Ju, J. Shiomi, J. Morikawa, R. Yoshida, *ACS Cent. Sci.*, **5**, 1717 (2019). DOI:10.1021/acscentsci.9b00804 , PMID:31660440
- [14] Y. Ikeda, M. Okuyama, Y. Nakazawa, T. Oshiyama, *KONICA MINOLTA TECHNOLOGY REPORT*, **16**, 136 (2019).
- [15] S. Wold, M. Sjöström, L. Eriksson, *Chemom. Intell. Lab. Syst.*, **58**, 109 (2001). DOI:10.1016/S0169-7439(01)00155-1
- [16] R. Nasrin, A. H. Bhuiyan, M. A. Gafur, *Int., J. Compos. Mater.*, **5**, 155 (2015).
- [17] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York (2006).
- [18] H. Kaneko, K. Funatsu, *Chemom. Intell. Lab. Syst.*, **142**, 64 (2015). DOI:10.1016/j.chemolab.2015.01.001
- [19] S. Garavaglia, A. Sharma, "A Smart Guide to Dummy Variables: Four Applications and a Macro" (1988).
- [20] R. Ramprasad, R. Batra, A. Mannodi-Kanakkithodi, C. Kim, *npj Comput Mater* **3**, 54 (2017).