



Machine learning prediction of the mechanical properties of injection-molded polypropylene through X-ray diffraction analysis

Ryo Tamura, Kenji Nagata, Keitaro Sodeyama, Kensaku Nakamura, Toshiki Tokuhira, Satoshi Shibata, Kazuki Hammura, Hiroki Sugisawa, Masaya Kawamura, Teruki Tsurimoto, Masanobu Naito, Masahiko Demura & Takashi Nakanishi

To cite this article: Ryo Tamura, Kenji Nagata, Keitaro Sodeyama, Kensaku Nakamura, Toshiki Tokuhira, Satoshi Shibata, Kazuki Hammura, Hiroki Sugisawa, Masaya Kawamura, Teruki Tsurimoto, Masanobu Naito, Masahiko Demura & Takashi Nakanishi (2024) Machine learning prediction of the mechanical properties of injection-molded polypropylene through X-ray diffraction analysis, *Science and Technology of Advanced Materials*, 25:1, 2388016, DOI: [10.1080/14686996.2024.2388016](https://doi.org/10.1080/14686996.2024.2388016)

To link to this article: <https://doi.org/10.1080/14686996.2024.2388016>



© 2024 The Author(s). Published by National Institute for Materials Science in partnership with Taylor & Francis Group.



[View supplementary material](#)



Published online: 15 Aug 2024.



[Submit your article to this journal](#)



Article views: 347



[View related articles](#)



[View Crossmark data](#)

Machine learning prediction of the mechanical properties of injection-molded polypropylene through X-ray diffraction analysis

Ryo Tamura ^{a,b,*}, Kenji Nagata ^{a,b,*}, Keitaro Sodeyama ^{a,b}, Kensaku Nakamura ^{a,c}, Toshiki Tokuhira ^{a,c}, Satoshi Shibata ^{a,d}, Kazuki Hammura ^{a,e}, Hiroki Sugisawa ^{a,f}, Masaya Kawamura ^{a,f}, Teruki Tsurimoto ^{a,f}, Masanobu Naito ^{a,g}, Masahiko Demura ^{a,b,h} and Takashi Nakanishi ^{a,i}

^aMaterials Open Platform for Chemistry, National Institute for Materials Science, Ibaraki, Japan;

^bCenter for Basic Research on Materials, National Institute for Materials Science, Ibaraki, Japan;

^cProcess Technology Laboratory, R&D Center, Mitsui Chemicals, Inc., Chiba, Japan;

^dEssential Chemicals Research Laboratory, Sumitomo Chemical Co., Chiba, Japan;

^ePlatform Laboratory for Science & Technology, Asahi KASEI Corporation, Shizuoka, Japan;

^fScience & Innovation Center, Mitsubishi Chemical Corporation, Kanagawa, Japan;

^gResearch Center for Macromolecules and Biomaterials, National Institute for Materials Science, Ibaraki, Japan;

^hResearch Network and Facility Services Division, National Institute for Materials Science, Ibaraki, Japan;

ⁱResearch Center for Materials Nanoarchitectonics (MANA), National Institute for Materials Science, Ibaraki, Japan

ABSTRACT

Predicting the mechanical properties of polymer materials using machine learning is essential for the design of next-generation of polymers. However, the strong relationship between the higher-order structure of polymers and their mechanical properties hinders the mechanical property predictions based on their primary structures. To incorporate information on higher-order structures into the prediction model, X-ray diffraction (XRD) can be used. This study proposes a strategy to generate appropriate descriptors from the XRD analysis of the injection-molded polypropylene samples, which were prepared under almost the same injection molding conditions. To this end, first, Bayesian spectral deconvolution is used to automatically create high-dimensional descriptors. Second, informative descriptors are selected to achieve highly accurate predictions by implementing the black-box optimization method using Ising machine. This approach was applied to custom-built polymer datasets containing data on homo-polypropylene and derived composite polymers with the addition of elastomers. Results show that reasonable accuracy of predictions for seven mechanical properties can be achieved using only XRD.

ARTICLE HISTORY

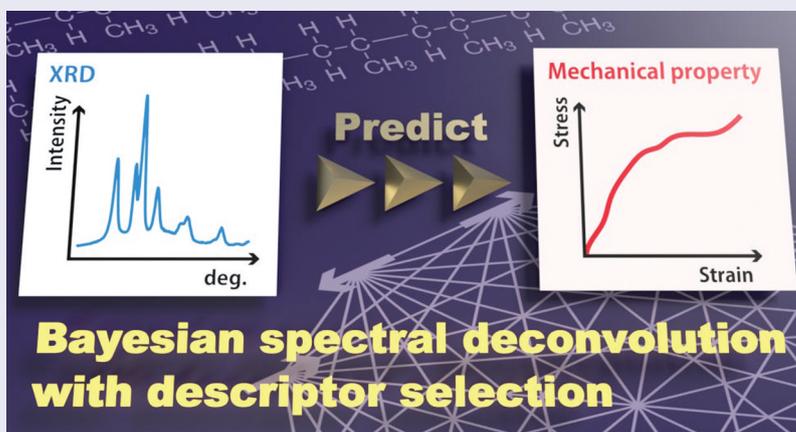
Received 28 May 2024

Accepted 30 July 2024

Revised 25 July 2024

KEYWORDS

Polypropylene; X-ray diffraction; bayesian spectral deconvolution; ising machine; machine learning



IMPACT STATEMENT

This study proposes a strategy to generate appropriate descriptors, which realize highly accurate predictions of mechanical properties via machine learning from the XRD analysis of the molded polypropylene samples.

CONTACT Ryo Tamura  tamura.ryo@nims.go.jp; Kenji Nagata  nagata.kenji@nims.go.jp; Takashi Nakanishi  nakanishi.takashi@nims.go.jp

 Materials Open Platform for Chemistry, National Institute for Materials Science, Ibaraki, Japan

*These authors equally contributed to this work.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/14686996.2024.2388016>

© 2024 The Author(s). Published by National Institute for Materials Science in partnership with Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

1. Introduction

Machine learning (ML) is powerful and indispensable for materials design [1–8]. ML can be applied to data on existing materials to predict the properties of unknown materials. However, to achieve accurate predictions, it is essential to use descriptors that correctly represent the features of these materials. Numerous ML tools have been developed to create descriptors for organic [9–11], polymer [12–14], and inorganic materials [15–17]. These descriptors are generally based on variables that researchers can easily control and generate without the need for actual synthesis, such as element concentration, materials structure, and manufacturing processes.

However, predicting the mechanical properties of polymer materials, particularly those of crystalline polymers containing higher-order structures, is challenging even with the power of ML. This is because not only the structure and degree of monomer polymerization that can be controlled but also the higher-order structure and morphology after molding significantly affect the mechanical properties of the material [18,19]. Therefore, predicting the mechanical properties of a polymer material after molding from just the monomer information and manufacturing process is typically challenging [20]. This encumbers the design of desired polymer materials based on ML. However, it is possible to predict the mechanical properties using descriptors that reflect the higher-order structural information of the polymer after molding. For example, higher-order structures can be captured by X-ray diffraction (XRD) measurements [21], and if the XRD results can be used as descriptors, the prediction of mechanical properties will be credible. However, XRD results can only be obtained after molding and the mechanical properties of unknown polymer materials cannot be predicted using these results as descriptors. An advantage of predicting mechanical properties from XRD results is the possibility of substituting complex destructive test measurements with simple non-destructive counterparts. In particular, by combining this predictor with the latest high-throughput polymer synthesis technology [22], polymer materials with desired mechanical properties can be screened using only XRD measurements, which contributes to a faster polymer design.

When using the XRD results as descriptors, the part of the diffraction pattern that captures the structural information should be selected appropriately. Certainly, using the entire diffraction pattern is possible, but typically, an excessively large descriptor dimensionality may lead to overfitting and reduce the prediction accuracy. Therefore, to achieve highly accurate predictions using XRD analysis, the extraction of parameters that express the shape of the diffraction pattern is crucial. For the important parameters, the shape of each peak

would be a better descriptor. However, extracting the number of peaks from the XRD results for polymer materials is challenging. To automatically extract this number, Bayesian spectral deconvolution has been proposed as a general method for analyzing spectra in materials science [23–25]. Using the exchange Monte Carlo approach [26], the number of peaks and their shapes are automatically obtained.

In this study, we prepared two types of datasets of injection-molded polypropylene samples, polypropylene (PP)s and composites of PP incorporating elastomers, under almost the same injection molding conditions. For injection molding, the same condition was used for composites of PP incorporating elastomers, while the temperature was changed only for PPs. Since the machine learning approach for polymer materials is difficult due to the hierarchical structure, the influence of the molding process on the higher-order structures was avoided as much as possible in our dataset. Bayesian spectral deconvolution was applied for obtained XRD patterns. By performing the machine learning analysis separately for two datasets, the extracted descriptors were used as ML models to predict the mechanical properties. It is shown that highly accurate predictions can be achieved by appropriately selecting the descriptors obtained through Bayesian spectral deconvolution. In addition, a black-box optimization method based on an Ising machine [27,28] was used to select the informative descriptors. With this procedure, most of the nine different mechanical properties could be predicted from the XRD results alone, excluding elongation at break and Rockwell hardness. The results indicate that machine learning can be effectively applied to the polymer materials dataset by preparing the dataset with the same conditions as possible. Thus, we believe this study becomes a test case for future data-driven approaches to polymer science. The procedure used in this study is illustrated in Figure 1. Note that, although descriptor selection is performed, the analysis of the relationship between the selected descriptors and the mechanical properties is not the purpose of this study.

2. Polypropylene dataset

Two types of custom-built polymer datasets were used to test this approach. The first dataset contained 15 types of homo-polypropylene (homo-PP) materials and is referred to as the homo-PP dataset. For each material, five temperatures (30, 45, 60, 70, and 75°C) were used for injection molding, and a total of 75 data were generated, which are summarized in Table S1. We confirmed that the XRD results do not change significantly when the temperature is changed, and the XRD results of an example are shown in Figure S1.

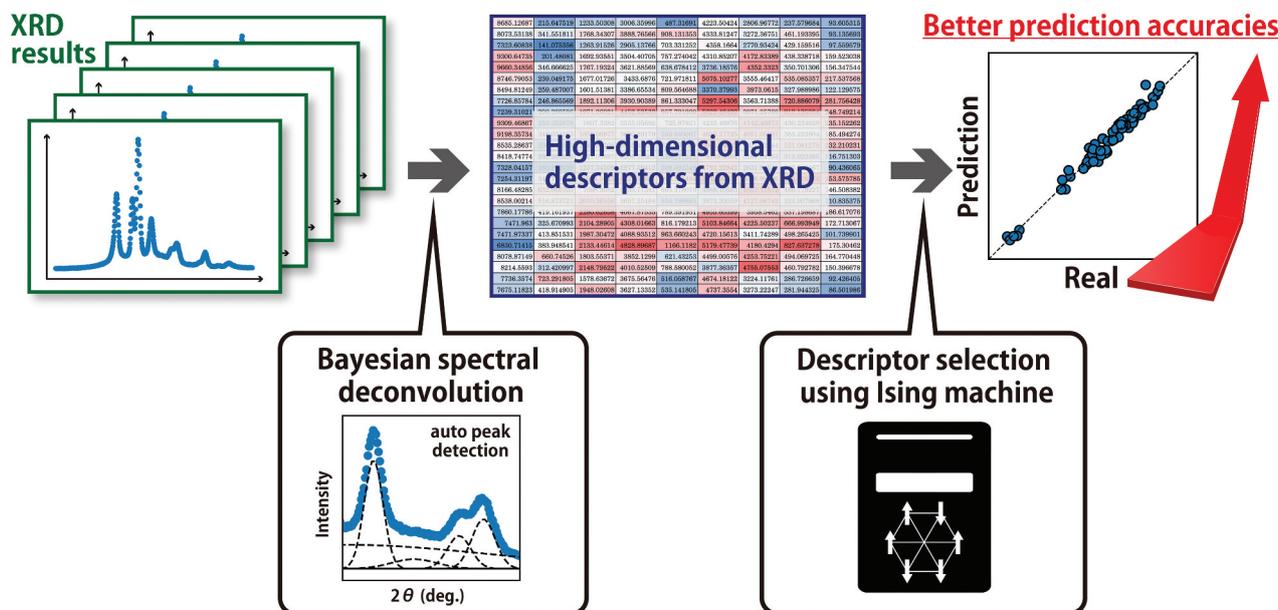


Figure 1. Flow diagram of the procedure used to provide a highly accurate machine learning prediction model using only the XRD results of polymer materials.

Thus, we considered that the samples with different temperatures can be treated as samples obtained under the same injection molding conditions. For the second dataset, referred to as the homo-PP-with-elastomer dataset, four types of homo-PPs with five different elastomers were targeted. The injection molding temperature was fixed at 45°C. Polymer materials were prepared with different mixing ratios (homo-PP:elastomer = 9:1, 8:1, and 7:1), using three types of screws (S1: weak, S2: medium, and S3: strong) for the twin screw extruder (Berstorff, ZE40A, Germany) and two different rotational speeds (100 and 300 rpm). A total of 120 samples were prepared, the details of which are summarized in Table S2. For each polymer material, a dumbbell-shaped specimen with a thickness of 4 mm was prepared by injection molding, and nine types of mechanical properties were measured for each sample: tensile modulus (MPa), yield stress (MPa), yield strain (%), elongation at break (%), flexural strength (MPa), flexural modulus (MPa), Charpy impact value (kJ/m²), heat distortion temperature (°C), and Rockwell hardness (HRC). Tensile tests (5966 Universal Testing Systems, Instron®, US) were performed at a tensile speed of 50 mm/min and a chuck distance of 115 ± 1 mm. In addition, the Charpy impact test (Digital Impact Testing Machine DG-UB Type, Toyo Seiki Seisaku-sho, Ltd., Japan) was performed at 23°C, the heat distortion temperature (HDT Tester 3M-2V, Toyo Seiki Seisaku-sho, Ltd., Japan) was measured at 0.45 MPa, and the Rockwell hardness (Rockwell Hardness Tester Model 3R, Imai Testing Machine MFG., Co., Ltd., Japan) was measured while overlapping two samples. Each test was repeated five times and the mean values of the results were used as the

prediction target. XRD measurements in the machine direction (MD; flow direction of the resin) and transverse direction (TD; width direction of the resin) were performed using a Rigaku SmartLab (Rigaku, Japan), which has a PhotonMax high-flux 9 kW rotating-anode X-ray source coupled with a HyPix-3000 high-energy-resolution semiconductor detector.

3. Results

3.1. Bayesian spectral deconvolution of XRD results

Bayesian spectral deconvolution was used to extract descriptors from the XRD results. Initially, the target peak function was defined as a Gaussian function, and the number of peaks was established to minimize the Bayesian free energy calculated using the exchange Monte Carlo method. Consequently, for each direction, that is, MD and TD, 12 and 11 peaks, including an amorphous halo, were extracted from the XRD data for the homo-PP and the homo-PP-with-elastomer datasets, respectively. Subsequently, a Gaussian function is estimated for each peak. Five parameters characterizing each peak – that is, the height, location, full width at half maximum (FWHM), area, and area ratio of the peaks – were extracted. Thus, for the homo-PP dataset, 120-dimensional descriptors were obtained, including MD and TD, and the dimensions of the descriptors for the homo-PP with the elastomer dataset were 110. Examples of the Bayesian spectral deconvolution results for homo-PP and homo-PP with an elastomer are shown in Figure 2. Although the five parameters in each Gaussian peak are not independent, we do not know in advance which are better

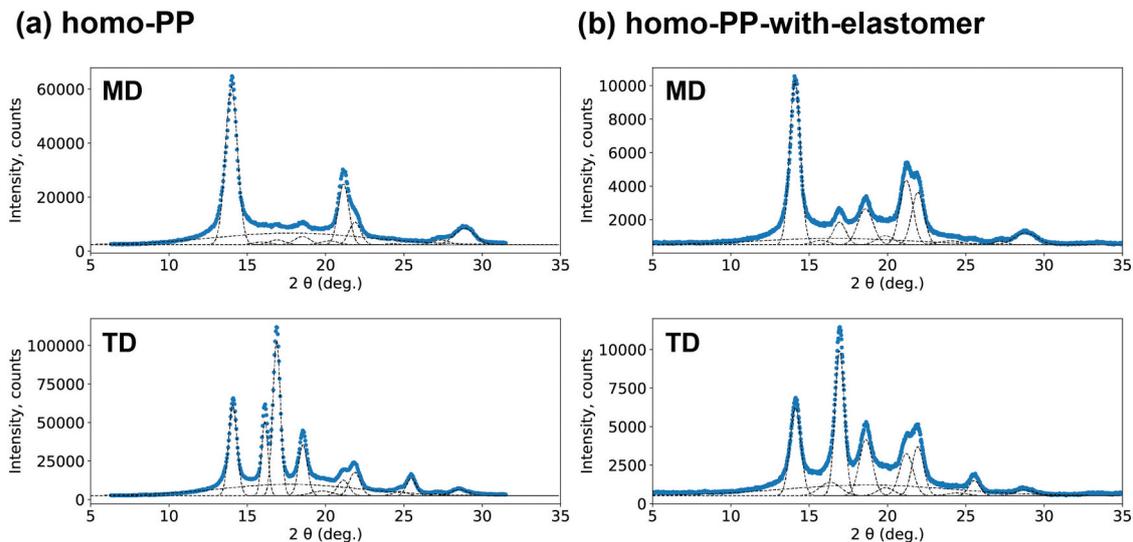


Figure 2. Examples of the Bayesian spectral deconvolution results for (a) homo-pp and (b) homo-pp with an elastomer when the XRD results of MD and TD, respectively, are targeted. The target samples are the ID 1 for each dataset. The blue points are the measured data, and the dotted lines are the fitting results.

parameters to improve the prediction accuracy of mechanical properties. Thus, after preparing such redundant features, the strategy of selecting important features, which will be explained in Sec. 3.2, was adopted in this study.

3.2. Descriptor selection

The Bayesian spectral deconvolution technique allowed the extraction of descriptors with more than 100-dimensions from the XRD results. However, this dimensionality is too high, resulting in overfitting and poor prediction accuracy at this rate. The prediction performances using all the extracted descriptors are shown in Table S3, where the failure in the predictions of almost all the mechanical properties is revealed. Therefore, the appropriate selection of the descriptors that can be used is essential to improve the prediction accuracy. The least absolute shrinkage and selection operator (LASSO) method is a well-known feature selection method used to avoid overfitting in linear regression; however, LASSO is often trapped in local minimum [29], and is seldom a suitable method. The exhaustive-search (ES) method [30,31] is useful for finding better combinations of descriptors as it exhaustively searches for the optimal combination that achieves the highest prediction accuracy. However, in the case of the homo-PP, $2^{120} - 1$ combinations existed when considering all the combinations of descriptors, and the ES method could not be used. Therefore, this study used a different approach to searching for better combinations. Finding a better combination to improve the prediction accuracy can be considered as a binary black-box optimization. A factorization machine with quantum annealing

(FMQA) algorithm [27] using Ising machines was proposed as a binary black-box optimization method that can be used to select appropriate XRD descriptors. The procedure for selecting descriptors using FMQA is as follows:

- (1) The dimension of the descriptors is N . Ten types of binary bits of length N are randomly generated. If the binary bit for each bit string is 1, the corresponding descriptor is used, and the mean squared error (MSE), evaluated by 10-fold cross-validation, is calculated. Linear regression is used as the ML prediction model. These 10 types of data are the initial data for black-box optimization.
- (2) The factorization machine (FM) is trained with 10 data points to predict the MSE when a binary string is input. The FM is defined as follows:

$$MSE = \sum_{i=1}^N w_i^* q_i + \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^K v_{ik}^* v_{jk}^* q_i q_j, \quad (1)$$

where q_i is the 0/1 binary bit, and the value of K is fixed at 8. The trained parameters are written as w_i^* and v_{ik}^*

- (3) By solving this FM with an Ising machine, (in this study, Fixstars Amplify AE[32] is used), a set of descriptors that would reduce the MSE is searched.
- (4) For the selected descriptors, we evaluate the actual value of the MSE by training a linear regression model.
- (5) Steps 2-4 are repeated for 290 times, and this independent trial is performed five times.

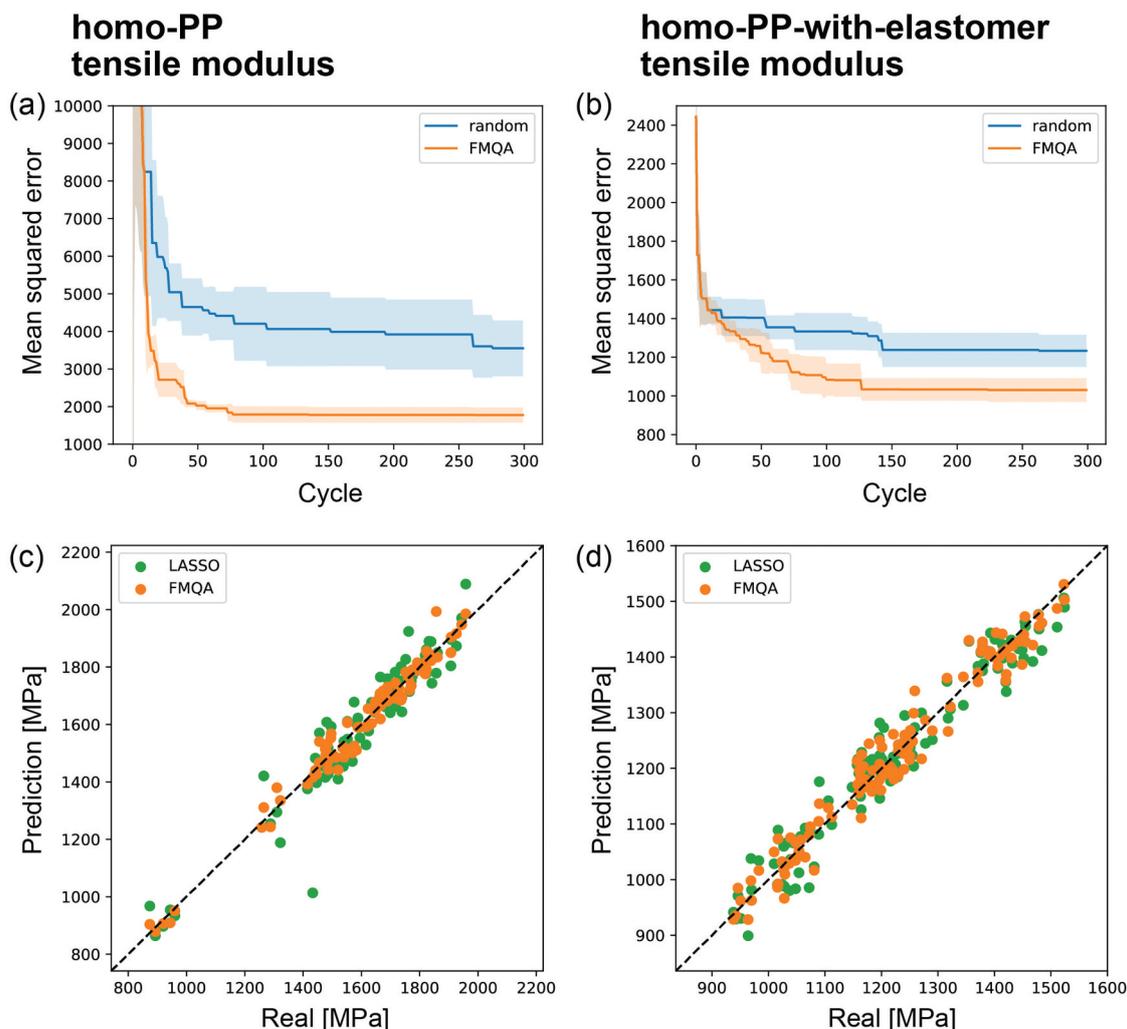


Figure 3. Mean squared error (MSE) as a function of the number of cycles for the tensile modulus of (a) homo-pp and (b) homo-pp-with-elastomer datasets. As the 10-fold cross-validation evaluates the MSE, this value indicates the prediction accuracy. The lines and shaded areas are the mean and standard deviation of five independent trials, respectively. Prediction results using 10-fold cross-validation for the (c) homo-pp and (d) homo-pp-with-elastomer datasets. The lowest MSE case in the FMQA search and LASSO results are shown for comparison.

Since black-box optimization can perform a global search, better combinations can be found compared to LASSO by avoiding entrapment in local minima.

The results of the descriptor selection for the tensile modulus case are shown in Figure 3(a,b). The horizontal axis represents the number of cycles and the vertical axis represents the corresponding minimum MSE value. The mean MSE values of five independent trials are plotted. For comparison, the results for a randomly generated bit string are also presented. Compared with random sampling, the MSE value of FMQA decreased with each cycle. Thus, FMQA determines a set of descriptors to achieve better prediction accuracy. Figures 3(c,d) present the prediction results of the 10-fold cross-validation of the lowest MSE in the FMQA search. These data points were not included in the training data when predicting each point. Most data points were distributed around the 45-degree line, and better predictions were achieved. The results of LASSO are also plotted for comparison,

revealing that FMQA allowed the selection of descriptors to build an ML model with better prediction accuracy. Here, `sklearn.linear_model.LassoCV` with default settings was used to perform LASSO [33]. We confirmed that the FMQA obtained better descriptors for the other eight mechanical properties (see Figures S2 and S3). As a reference, the selected descriptors depending on the mechanical properties using FMQA and LASSO are summarized in Figures S4–S7. The extraction of additional information regarding the relationship between the descriptors and predicted targets using these results is cumbersome.

3.3. Prediction accuracies of mechanical properties by XRD results

To illustrate the prediction accuracies for the nine mechanical properties, Figure 4 plots the R^2 values for the predictions when a 10-fold cross-validation is performed for the lowest MSE case in the FMQA search.

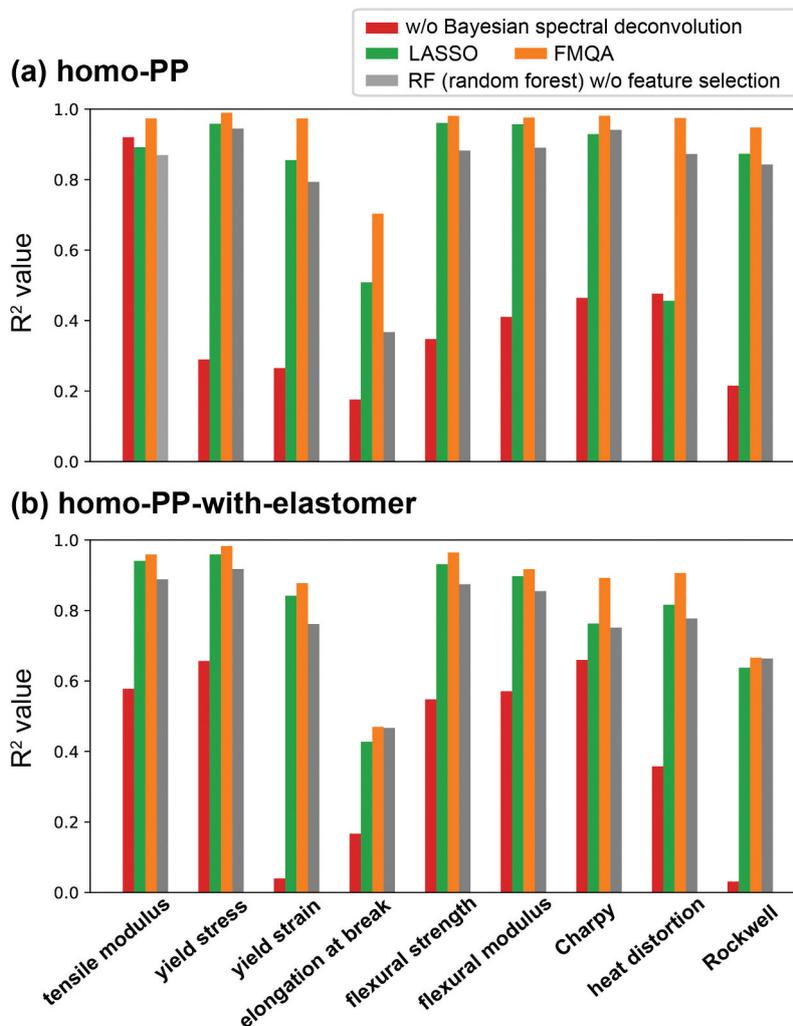


Figure 4. R^2 values for the predictions when the 10-fold cross-validation is performed for the lowest MSE case in the FMQA search for (a) homo-pp and (b) homo-pp-with-elastomer datasets. We also plot the results obtained using LASSO, the six-dimensional descriptors without using Bayesian spectral deconvolution, and the random forest regression without feature selection. As the R^2 values approach 1, the prediction accuracy increases.

The LASSO results are also shown. For comparison purposes, the $\alpha(110)$ peak height, FWHM, and crystallinity, which are usually extracted by researchers without using Bayesian spectral deconvolution, were derived from the XRD results for MD and TD, and the prediction results using these six-dimensional descriptors were obtained. In addition, for comparison with the nonlinear model, the results of the random forest model without the feature selection are also evaluated. Clearly, our Bayesian spectral deconvolution technique increased the prediction accuracy for all the mechanical properties. The selection of descriptors using FMQA ensures that most mechanical properties can be predicted with high accuracy based solely on XRD results (see Table S3). Furthermore, the prediction accuracy of the linear model and the feature selection by FMQA outperform those of the nonlinear random forest model. However, the prediction accuracy for the elongation at break was low for both homo-PP and homo-PP-with-elastomer datasets. We confirmed that

the experimental accuracy for the elongation was poor, and that the values changed significantly over the five experiments. For example, in the homo-PP sample with the highest variability, the elongation values over the five experiments were 87.9%, 108.2%, 476.7%, 483.5%, and 491.3%. Thus, the prediction targets have a high uncertainty, rendering accurate predictions impossible. In addition, the Rockwell hardness for homo-PP-with-elastomer dataset has a lower prediction accuracy than the others, although it was not as inaccurate as the elongation. Similar to elongation, this is a property with high variation between samples. These large variations between samples for elongation and Rockwell hardness are due to the structural defects in the samples. However, the XRD data has information on crystal structure and little information on structural defects. Thus, we believe it is difficult to predict the elongation and Rockwell hardness from the XRD results. It should be emphasized that the prediction of the other seven mechanical properties, excluding elongation at break

and Rockwell hardness, was achieved with high accuracy based solely on the results of XRD, which is a relatively simple measurement.

4. Conclusions

This study attempted to predict the mechanical properties of PPs using XRD analysis alone. To achieve highly accurate predictions, a strategy was proposed using Bayesian spectral deconvolution and an informative descriptor selection method by employing FMQA. Consequently, most of the mechanical properties considered in this study, such as the tensile modulus, yield stress, yield strain, flexural strength, flexural modulus, Charpy impact value, and heat distortion temperature, were predicted with high accuracy using XRD results alone. However, it was challenging to predict the elongation at break. Therefore, our approach enables the prediction of the mechanical properties of PPs, typically assessed through destructive testing, using only nondestructive testing, specifically, XRD. In this study, each dataset was prepared under almost the same injection molding conditions, and the machine learning analysis was performed separately for each dataset. Thus, our trained model cannot be applied to predict the mechanical properties of all PPs. If new PP samples are considered, we need to collect the mechanical properties and prepare the machine learning model according to our proposed procedure.

The combination of Bayesian spectral deconvolution and the informative descriptor selection method using FMQA can be applied to other spectral analyses such as X-ray photoelectron spectroscopy and Mössbauer spectroscopy measurements. Therefore, our approach contributes to the highly accurate prediction of properties that cannot be expressed without incorporating complex information from measurements.

Acknowledgments

We thank Miharu Shimizu for performing the simulations. Furthermore, we thank all members of the Materials Open Platform for Chemistry, composed of NIMS, Asahi Kasei Corporation, Mitsubishi Chemical Corporation, Mitsui Chemicals, Inc., and Sumitomo Chemical Co., Ltd. for their valuable discussions. T. N. acknowledges the World Premier International Research Center Initiative (WPI), MEXT, Japan.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Ryo Tamura  <http://orcid.org/0000-0002-0349-358X>
Keitaro Sodeyama  <http://orcid.org/0000-0002-9228-0729>
Kazuki Hammura  <http://orcid.org/0009-0000-9671-5059>
Masanobu Naito  <http://orcid.org/0000-0001-7198-819X>
Masahiko Demura  <http://orcid.org/0000-0002-7308-3041>
Takashi Nakanishi  <http://orcid.org/0000-0002-8744-782X>

Data availability statement

Due to the nature of the research, due to commercial supporting data is not available.

Impact statement

This study proposes a strategy to generate appropriate descriptors, which realize highly accurate predictions of mechanical properties via machine learning from the XRD analysis of the molded polypropylene samples.

References

- [1] Ramprasad R, Batra R, Pilania G, et al. Machine learning in materials informatics: recent applications and prospects. *Npj Comput Mater.* 2017;3(1):1–13. doi: [10.1038/s41524-017-0056-5](https://doi.org/10.1038/s41524-017-0056-5)
- [2] Audus DJ, de Pablo JJ. Polymer informatics: opportunities and challenges. *ACS Macro Lett.* 2017;6(10):1078–1082. doi: [10.1021/acsmacrolett.7b00228](https://doi.org/10.1021/acsmacrolett.7b00228)
- [3] Wu S, Kondo Y, Kakimoto M, et al. Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *Npj Comput Mater.* 2019;5(1):1–11. doi: [10.1038/s41524-019-0203-2](https://doi.org/10.1038/s41524-019-0203-2)
- [4] Doan Tran H, Kim C, Chen L, et al. Machine-learning predictions of polymer properties with polymer genome. *J Appl Phys.* 2020;128(17):171104. doi: [10.1063/5.0023759](https://doi.org/10.1063/5.0023759)
- [5] Terayama K, Sumita M, Tamura R, et al. Black-box optimization for automated discovery. *Acc Chem Res.* 2021;54(6):1334–1346. doi: [10.1021/acs.accounts.0c00713](https://doi.org/10.1021/acs.accounts.0c00713)
- [6] Tamura R, Osada T, Minagawa K, et al. Machine learning-driven optimization in powder manufacturing of Ni-Co based superalloy. *Mater Des.* 2021;198:109290. doi: [10.1016/j.matdes.2020.109290](https://doi.org/10.1016/j.matdes.2020.109290)
- [7] Hayashi Y, Shiomi J, Morikawa J, et al. RadonPy: automated physical property calculation using all-atom classical molecular dynamics simulations for polymer informatics. *Npj Comput Mater.* 2022;8(1):1–15. doi: [10.1038/s41524-022-00906-4](https://doi.org/10.1038/s41524-022-00906-4)
- [8] Ma R, Zhang H, Xu J, et al. Machine learning-assisted exploration of thermally conductive polymers based on high-throughput molecular dynamics simulations. *Mater Today Phys.* 2022;28:100850. doi: [10.1016/j.mtphys.2022.100850](https://doi.org/10.1016/j.mtphys.2022.100850)
- [9] Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci.* 1988;28(1):31–36. doi: [10.1021/ci00057a005](https://doi.org/10.1021/ci00057a005)

- [10] Huang B, von Lilienfeld OA. Communication: understanding molecular representations in machine learning: the role of uniqueness and target similarity. *J Chem Phys.* 2016;145(16):161102. doi: 10.1063/1.4964627
- [11] Krenn M, Häse F, Nigam A, et al. Self-referencing embedded strings (SELFIES): a 100% robust molecular string representation. *Mach Learn Sci Technol.* 2020;1:045024. doi: 10.1088/2632-2153/aba947
- [12] Kim C, Chandrasekaran A, Huan TD, et al. Polymer genome: a data-powered polymer informatics platform for property predictions. *J Phys Chem C.* 2018;122(31):17575–17585. doi: 10.1021/acs.jpcc.8b02913
- [13] Lin T-S, Coley CW, Mochigase H, et al. BigSMILES: a structurally-based line notation for describing macromolecules. *ACS Cent Sci.* 2019;5(9):1523–1531. doi: 10.1021/acscentsci.9b00476
- [14] Wu S, Lambard G, Liu C, et al. iQSPR in XenonPy: a bayesian molecular design algorithm. *Mol Inform.* 2020;39(1–2):1900107. doi: 10.1002/minf.201900107
- [15] Ward L, Agrawal A, Choudhary A, et al. A general-purpose machine learning framework for predicting properties of inorganic materials. *Npj Comput Mater.* 2016;2(1):1–7. doi: 10.1038/npjcom-pumats.2016.28
- [16] Isayev O, Oses C, Toher C, et al. Universal fragment descriptors for predicting properties of inorganic crystals. *Nat Commun.* 2017;8(1):15679. doi: 10.1038/ncomms15679
- [17] Ward L, Dunn A, Faghaninia A, et al. Matminer: an open source toolkit for materials data mining. *Comput Mater Sci.* 2018;152:60–69. doi: 10.1016/j.commatsci.2018.05.018
- [18] Popli R, Mandelkern L. Influence of structural and morphological factors on the mechanical properties of the polyethylenes. *J Polym Sci B Polym Phys.* 1987;25(3):441–483. doi: 10.1002/polb.1987.090250301
- [19] Mimura K, Ito H, Fujioka H. Improvement of thermal and mechanical properties by control of morphologies in pes-modified epoxy resins. *Polymer.* 2000;41(12):4451–4459. doi: 10.1016/S0032-3861(99)00700-4
- [20] Tamura R, Takei Y, Imai S, et al. Experimental design for the highly accurate prediction of material properties using descriptors obtained by measurement. *Sci Tech Adv Mater Methods.* 2021;1(1):152–161. doi: 10.1080/27660400.2021.1963641
- [21] Trovati G, Sanches EA, Neto SC, et al. Characterization of polyurethane resins by FTIR, TGA, and XRD. *J Of Appl Polym Sci.* 2010;115(1):263–268. doi: 10.1002/app.31096
- [22] Oliver S, Zhao L, Gormley AJ, et al. Living in the Fast Lane—high throughput Controlled/living radical polymerization. *Macromolecules.* 2019;52(1):3–23. doi: 10.1021/acs.macromol.8b01864
- [23] Nagata K, Sugita S, Okada M. Bayesian spectral deconvolution with the exchange Monte Carlo method. *Neural Networks.* 2012;28:82–89. doi: 10.1016/j.neunet.2011.12.001
- [24] Shinotsuka H, Nagata K, Yoshikawa H, et al. Development of spectral decomposition based on bayesian information criterion with estimation of confidence interval. *Sci Technol Adv Mater.* 2020;21(1):402–419. doi: 10.1080/14686996.2020.1773210
- [25] Matsumura T, Nagamura N, Akaho S, et al. High-throughput XPS spectrum modeling with autonomous background subtraction for 3d5/2 peak mapping of SnS. *Sci Technol Of Adv Mater Methods.* 2023;3(1):2159753. doi: 10.1080/27660400.2022.2159753
- [26] Hukushima K, Nemoto K. Exchange monte carlo method and application to spin glass simulations. *J Phys Soc Jpn.* 1996;65(6):1604–1608. doi: 10.1143/JPSJ.65.1604
- [27] Kitai K, Guo J, Ju S, et al. Designing metamaterials with quantum annealing and factorization machines. *Phys Rev Res.* 2020;2(1):013319. doi: 10.1103/PhysRevResearch.2.013319
- [28] Koshikawa AS, Ohzeki M, Kadowaki T, et al. Benchmark test of black-box optimization using D-Wave quantum annealer. *J Phys Soc Jpn.* 2021;90(6):064001. doi: 10.7566/JPSJ.90.064001
- [29] Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B (Methodol).* 1996;58(1):267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x
- [30] Nagata K, Kitazono J, Nakajima S, et al. An exhaustive search and stability of sparse estimation for feature selection problem. *IPSJ Online Trans.* 2015;8:25–32. doi: 10.2197/ipsjtrans.8.25
- [31] Igarashi Y, Takenaka H, Nakanishi-Ohno Y, et al. Exhaustive search for sparse variable selection in linear regression. *J Phys Soc Jpn.* 2018;87(4):044802. doi: 10.7566/JPSJ.87.044802
- [32] Fixstars amplify [internet]. Available from: <https://amplify.fixstars.com/en/>
- [33] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011;12:2825–2830.