A Study on the Material Properties and Process Conditions Reasoning using Regression and Classification of Machine Learning

Hyunjong Kim
ICT-enabled Intelligent Manufacturing Research Section
Electronics and Telecommunications Research Institute
Daejeon, South Korea
hjkim23@etri.re.kr

Hyunwoo Oh
ICT-enabled Intelligent Manufacturing Research Section
Electronics and Telecommunications Research Institute
Daejeon, South Korea
hyunwoo@etri.re.kr

Suyoung Chi
ICT-enabled Intelligent Manufacturing Research Section
Electronics and Telecommunications Research Institute
Daejeon, South Korea
chisy@etri.re.kr

***Abstract*—In this paper, we explain the property prediction and manufacturing process condition inference methodology of ceramic materials with the introduction of AI models and perform PoC of the proposed method. We selected the optimal model by applying data with various input characteristics related to material properties to regression/classification machine learning models with proven performance. The accuracy of the final selected model was more than 90%, showing the possibility of drastically reducing the material development period.**

Keywords—ceramic materials, machine learning, regression, classification, property prediction, process condition inference

I. INTRODUCTION

In order for the domestic material industry to have global competitiveness, it is important not only to create materials with optimal physical properties but also to develop materials by exploring optimal material synthesis methods in a short time. In general, in order to develop a new material, several raw materials are mixed, synthesized through various complex processes, and then the performance of the material is checked. When synthesizing most materials, the physical properties of the materials vary sensitively depending on the process conditions as well as the properties of the raw materials.

For this reason, most industry / academia / research institutions directly synthesize all possible combinations of composition information and process variables to check performance or explore materials with good performance based on the experience and insights of individual researchers. However, this material search method consumes a lot of time and resources and relies heavily on the researcher's experience and ability.

To solve this problem, industries, academia, and research institutes are trying various methods, and recently, they are focusing on developing material development systems using AI(Artificial Intelligence) models (machine learning and deep learning) based on big data. In this paper, we explain the property prediction and manufacturing process condition inference methodology of ceramic materials with the introduction of AI models and experiment with the possibility of a proposed method. We experimented with property prediction and process condition inference possibilities by applying data with various input properties related to material properties to regression/classification machine learning models with proven performance.

Currently, AI is receiving a lot of attention in various fields and is being actively researched. However, in the field of material development, AI is not introduced relatively well compared to other fields, and one of the reasons is that material development researchers cannot easily use AI. In order to learn and apply AI models to processes, data related to material development must be analyzed, and knowledge of AI models is required. Therefore, it is necessary for general material researchers to develop a program that can easily apply their data to AI.

Ceramic parts are essential for related processes as high-tech manufacturing industries such as semiconductors, secondary batteries, electric vehicles, and displays have recently grown. There are many applications of fine ceramics, which are very important in modern life, and they are largely divided into four types as shown in Figure 1, including electrical/electronic ceramics, energy/environmental ceramics, mechanical/structural ceramics, and bio-ceramic.

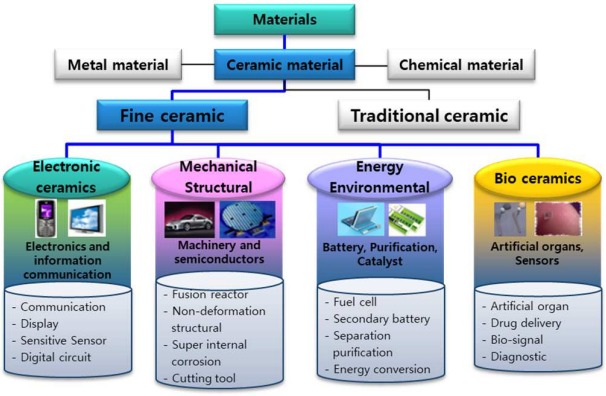

Fig. 1. Applications of Fine Ceramics as an Essential Ceramic Parts

Fine ceramic materials refer to refined minerals, artificially synthesized inorganic compounds, or materials synthesized from their compositions, and are materials that exhibit various industrial functions through precision molding, heat treatment, and processing. Repeated experiments shall be conducted under various raw material combinations and process conditions to develop materials that satisfy the required properties in various application fields. To reduce the time and cost of this process, we experimented on methodologies to predict target properties and infer compositions and process conditions that satisfy the required properties using the main machine learning models, regression models and classification models. With the introduction of the AI model, we explain the property prediction and manufacturing process conditions inference methodology of ceramic materials and perform PoC (Proof of Concept) of the proposed method.

We selected the optimal model by applying data with various input characteristics related to material properties to

regression/classification machine learning models with proven performance. The accuracy of the final selected model is more than 90%, and we have confirmed the possibility of drastically reducing the material development period through the proposed method.

II. Related Works

Recently, interest in artificial intelligence technology in the field of material development can be said to be explosive, and the biggest reason is the rapid prediction speed. For example, existing computational science technology (first principle-calculation) took months to predict material properties, but artificial intelligence can be shortened to minutes. Basically, unlike the existing computational science method, artificial intelligence's prediction method is a method of finding a correlation between material (composition and structure) information and physical properties in the material property DB and predicting the properties of new materials based on this. Principles such as quantum mechanics, fixed mechanics, and statistical mechanics are not typically used. For this reason, artificial intelligence-based material design technology can provide a new paradigm in the field of material development.

In [1], material artificial intelligence technology is identified as the fourth paradigm for material development. In other words, it is classified into a first-order paradigm that relies on experience and trial and error tests, a second-order paradigm based on thermodynamics, a third-order paradigm based on computational science, and a fourth-order paradigm based on artificial intelligence. Currently, artificial intelligence is being used in various ways in the field of material design. In particular, research is actively underway in the field of developing new theoretical methods to overcome the shortcomings of existing computational science and technology, the field of predicting physical properties from material information quickly (forward prediction technology), and the field of reverse prediction technology.

However, the prediction performance of these models is a limitation in improving the accuracy of the prediction model due to the complexity of space and time between the environment and the source and the problem that the prediction value converges to an average concentration as the prediction time increases in the artificial neural network [2].

Recently, research results that predict the characteristics of materials using machine learning based on a large amount of material data obtained based on density functional theory (DFT) are drawing attention[3-10]. Material data based on quantum calculation DFT include molecular electrode material development[3], energy band gap prediction[4], 2-dimensional magnetic material characteristic prediction[5], thermoelectric material development[6], and superhard material development[7-10].

Among them, bulk modulus and shear modulus related to the mechanical properties of the material correspond to representative properties used to develop materials with excellent hardness and incompressibility. Expectations for the development of new ultra-hard materials are growing as machine learning is recently used to identify material candidates with excellent mechanical properties using volume elasticity and shear elasticity coefficient data obtained through DFT quantum calculations.

In [10], they uses three machine learning models: Support Vector Machine Regression(SVR), Random Forest Regression(RF), and Xgboost Regression(XGB), but each model was optimized through cross validation(CV) and hyper-parameter tuning. Based on the results of the regression analysis of each model, the predicted material characteristics of each model were compared with each other using two performance indicators: the coefficient of determination (R^2-score) value and the root-mean-square-error (RMSE).

We used regression models and classification models for property prediction and process condition inference in the development of ceramic dielectric materials, respectively. To analyze the performance of the experimental model, MSE, RMSE, and R^2-score were used as regression models, and the optimal model was selected using accrual indicators for classification models.

III. System Design for Properties Prediction and Process Inference of Ceramic Material

We simplified the ceramic process into four major processes: ball mill, spray drying, molding, and sintering, and summarized data related to quality and physical properties considered in each process as shown in Fig. 2. The ball mill process is a process of making slurry by mixing raw materials, and a mixing ratio, a ball mill rotation speed, and a mixing time are important factors. The slurry is transformed into powder through a spray drying process, and in this process, slurry visibility, atomizer rotational speed, and inlet/outlet temperature have an important influence on material properties. In the molding process, tap density, molding pressure, and pressure retention time related to the density of the product are correlated with quality, and in the sintering process, the heating rate, temperature retention time, and retention temperature that affect the contraction of the product were considered as important input data.

Fig. 2. List of data related to fine ceramics production procedures and material characteristics for each process

A number of experiments shall be conducted to find product manufacturing conditions that satisfy the required properties because of the property determined not only by combining raw materials but also by various conditions set in the manufacturing process.

TABLE I. Number of Raw Material Combinations and Experimental Conditions for The Development of Dielectric Materials for Future Vehicles

Sortation	Raw combination	Composition	Processes	Properties
No. Dataset	*37~49*	*49*	*49*	*49*

For example, as shown in Table 1, up to 67,620 experiments in raw material combinations, composition, process conditions, and physical properties are required to

develop genetic materials for future vehicles, which cause numerous time and huge cost losses. The experimental conditions for optimal composition and process exploration of genetic materials for future vehicles are up to 67,620 cases, and assuming that the time required per experiment is one hour, the time required to conduct all experiments is well over seven years.

To solve this problem, we predicted ceramic material properties by inputting various conditions as shown in Figure 3, designed a methodology that can backward raw material composition and process conditions to meet the required properties, and tested the possibility using genetic material experimental data.

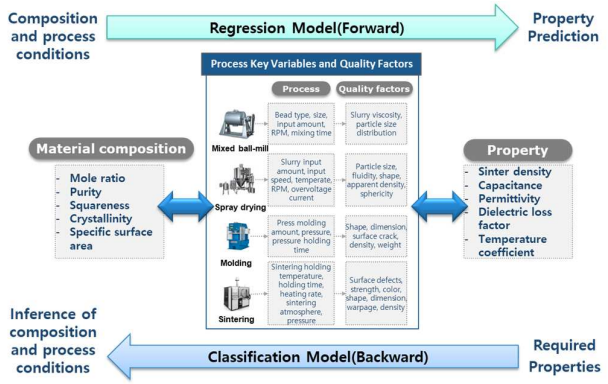

Fig. 3. Forward & Backward Methodology Schematic Diagram for Predicting the Properties of Ceramic Materials and Inferring Conditions Satisfying the Required Properties

Figure 4 shows how to select a material property and condition inference model through the concepts of input features and target features switching between machine learning regression models and classification models. If the combination of ceramic raw materials and process conditions are entered into the regression model and the property to be predicted is set as the target feature, the regression model provides the material property prediction as the result value. At this time, only the prediction model of the set target feature is generated. Conversely, the composition and condition inference model uses the required property, that is, the target feature, as an input, and a classification model is created for each feature used as an input to predict material properties.

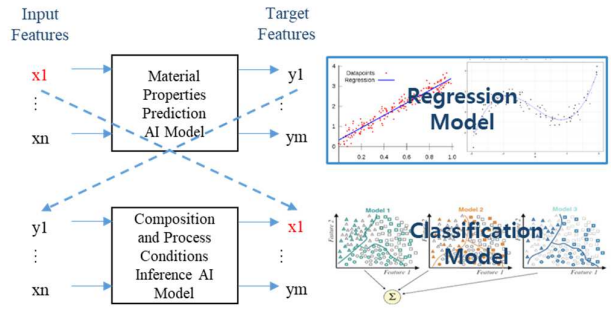

Fig. 4. Material Properties and Conditions Inference Model Selection Method using Input and Target Features Switching Concepts between Regression and Classification Models

Figure 5 shows the flowchart for material properties and condition inference methods through input feature switching between regression models and classification models. Even if the optimal model is selected according to the proposed procedure, the data preprocessing, feature selection process, and fine tuning process through hyperparameter setting are repeated if the reference performance is not satisfied.

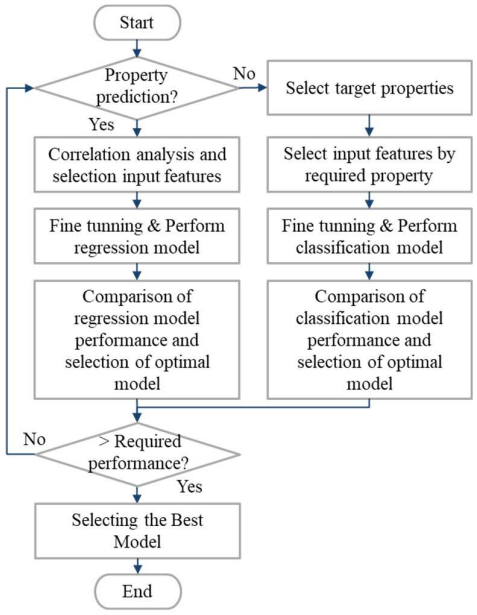

Fig. 5. Flowchart for Predicting the Properties of Ceramic Materials and Driving the Process Condition Inference System

IV. Experiment and Analysis of Machine Learning Model for Predicting Properties of Ceramic Dielectric Materials and Inferencing Process Conditions

We used genetic material experimental data (MDF: Miniature Data Factory) with 37 characteristics and selected 21 features correlated with target features through correlation analysis. Among the 21 features, target features to predict material properties used barium/titanium molar ratio, purity, tetragonality, crystallinity, specific surface area, density, dielectric constant, and dielectric loss coefficients. Only purity, specific surface area, dielectric constant, and dielectric loss coefficient met the requirements.

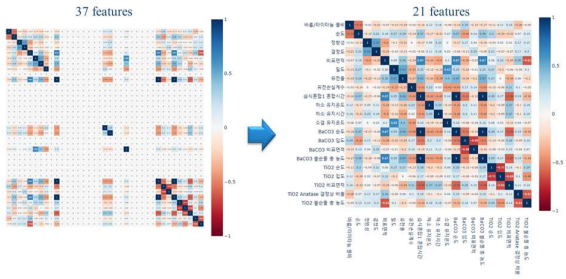

Fig. 6. Data Preprocessing Results of Selecting 21 Features through Correlation Analysis between Input Features

A. *Material Properties Prediction Test using Regression Models*

There were 370 datasets used to learn the physical properties prediction model of the four dielectric materials, and the train set and test set ratios were set to 0.7 and 0.3. The machine learning regression model used to predict physical properties performed AutoML on 20 subjects, including the Gradient Boosting Stressor, Extreme Gradient Boosting, Extra Tree Stressor, Bayesian Ridge, and K-Neighbors Stressor.

Figure 7 shows the learning results of the optimal model for four target features.

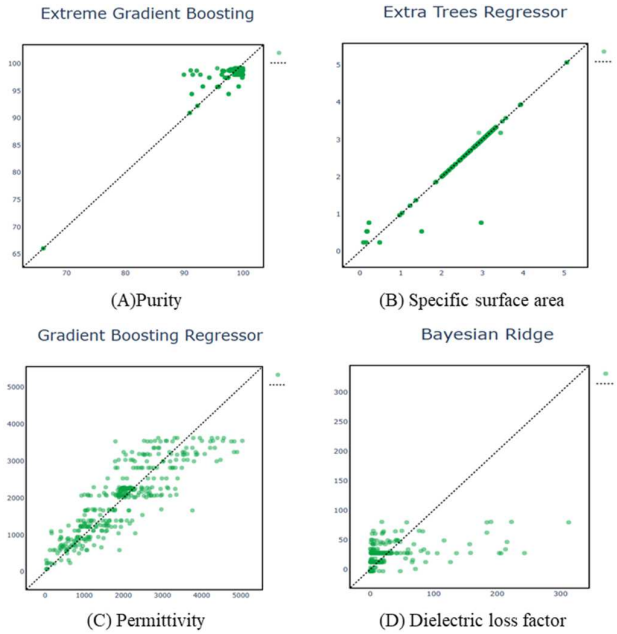

Fig. 7. Regression Model Training Results for 4 Properties(Purity, Specific surface area, Permittivity, Dielectric loss factor)

It can be seen that a machine learning regression model suitable for predictive properties is selected differently depending on the correlation and data distribution between input and target features. Through machine learning model experiments, we were able to confirm that the specific surface area among the properties of genetic materials had the best prediction performance (R^2-score), and in the case of the dielectric loss factor, it was inappropriate to predict properties through a machine learning regression model. Table 2 shows the fit model and the performance of the model for each feature.

TABLE II. PERFORMANCE ANALYSIS OF MATERIAL PROPERTIES PREDICTION BY REGRESSION MODEL

Target Property	Regression Model	MSE	RMSE	R^2-Score
Purity	Extra Trees	**3.53**	**1.87**	**0.66**
	Decision Tree	3.61	1.89	0.66
	Extreme Gradient Boosting	3.61	1.89	0.66
Specific surface area	Extra Trees	**0.06**	**0.21**	**0.93**
	Extreme Gradient Boosting	0.06	0.21	0.93
	Decision Tree	0.06	0.22	0.93
Permittivity	Gradient Boosting	**373,521**	**609.69**	**0.68**
	Light Gradient Boosting	397,299	628.86	0.66
	Random Forest	407,371	636.94	0.65
Dielectric loss factor	Bayesian Ridge	**1,077**	**31.17**	**0.25**
	Lasso Regression	1,078	31.2	0.24
	Elastic Net	1,078	31.2	0.24

B. Inference Test of Composition and Process Conditions using Classification Models

Thirteen features were considered for composition and process inference by property, but nine features were selected, excluding features with one or less classes(conditions). Here we used nine classification models: Gradient Boosting, Random Forest, Logistic Regression, Ada Boost, Decision Tree, and Support Vector Machine (SVM). Table 3 shows the top three optimal models for each of the six required physical properties feature with high model accuracy. Due to the nature of the classification model, the model accuracy was 1.0 for $BaCO_3$ Purity and Ball mill mixing time with only two classes, and the compliance model performance was 0.9 or higher for features with four classes.

Unlike the property prediction regression model, the composition and process inference model by property is characterized by a large difference in performance for each classification model type. This means that the selection of highly correlated features through data preprocessing and the selection of optimal models through AutoML are important.

TABLE III. PROCESS CONDITION INFERENCE PERFORMANCE ANALYSIS BY CLASSIFICATION MODEL

Target Condition	Classification Model	Accuracy
$BaCO_3$ Composition	Gradient Boosting	**0.938**
	Random Forest	0.902
	Logistic Regression	0.696
$BaCO_3$ Purity	Ada Boost	**1.0**
	Decision Tree	0.973
	Logistic Regression	0.964
Ball mill mixing time	Ada Boost	**1.0**
	Gradient Boosting	1.0
	Decision Tree	0.973
Sintering retention temperature	SVM(Support Vector Machine)	**0.813**
	Decision Tree	0.795
	Gradient Boosting	0.795
Calcination holding temperature	Gradient Boosting	**0.991**
	Random Forest	0.902
	Decision Tree	0.893
Calcination holding time	Gradient Boosting	**0.946**
	Decision Tree	0.911
	Random Forest	0.893

V. RESULTS AND FURTHER STUDY

With the introduction of the AI model, we explained the methodology for predicting the properties of ceramic materials and inferring manufacturing process conditions, and performed PoC of the proposed method. We selected the optimal model by applying data with various input characteristics related to material properties to regression/classification machine learning models with proven performance. The accuracy of the final selected model was more than 90%, showing the possibility of drastically reducing the material development period. In the future, we plan to apply a deep neural network to features that have not obtained the required performance by applying a machine learning model, and to evaluate the impact of property

prediction by feature and apply the results to the prediction/inference model by applying an explainable artificial intelligence (XAI) model.

ACKNOWLEDGMENT

This work was supported by Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE)(No. P0022332, Digital data platform for material development)

This work was supported by Korea Evaluation Institute of Industrial Technology(KEIT) grant funded by the Korea government(MOTIE)(RS-2023-00259964, Development of Auto Re-learning Common Service Platform Technology for Ceramic Industry DX based on Cloud)

REFERENCES

[1] A. Agrawal and A. Choudhary, "APL Materials", 2016.

[2] Md. M. Rahman and F. H. Siddiqui, "Multi-layered attentional peephole convolutional LSTM for abstrctive text summarization," ETRI Journal, vol. 43, Issue 2, pp. 288-298, April 2021.

[3] S. Muller, C. Sauter, R. Shunmugasundaram, N. Wenzler, V. De Andrade, F. D. Carlo, E. Konukglu, and V. Wood, "Deep learning-based segmentation of lithuum-ion battery microstructures enhanced by artificially generated electrodes", Nature communications vol. 12, October 2021.

[4] Y. Zhuo, A. M. Tehrani, and J. Brgoch, "Predicting the Band Gaps of Inorganic Solids by Machine Learning", Phys. Chem. Lett. vol. 9, 7, pp.1668-1673, March 2018.

[5] T. D. Rhone, W. Chen, S. Desai, S. B. Torrisi, D. T. Larson, A. Yacoby, and E. Kaxiras, "Data-driven studies of magnetic two-dimensional materials", Scientific Reports, vol. 10, September 2020.

[6] T. Wang, C. Zhang, H. Snoussi, and G. Zhang, "Machine Learning Approaches for Thermoelectric Materials Research", Advanced Functional Materials, vol. 30, January 2020.

[7] A. M. Tehrani, A. O. Oliynyk, M. Parry, Z. Rizvi, S. Couper, F. Lin, L. Miyagi, T. D. Sparks, and J. Brgoch, "Machine Learning Directed Search for Ultraincompressible, Superhard Materials", Journal of the American Chemical Society, vol. 140, pp.9844-9853, July 2018.

[8] Y. Zuo, M. Qin, C. Chen, W. Ye, Xi. Li, J. Luo, and S. P. Ong, "Accelerating materials discovery with Bayesian optimization and graph deep learning", Materials Today, vol. 51, pp.126-135, December 2021.

[9] P. Avery, X. Wang, C. Oses, E. Gossett, D. M. Proserpio, C. Toher, S. Curtarolo, and E. Zurek, "Perdicting superhard materials via a machine learning informed evolutionary structure search", Nature communications vol. 5, September 2019.

[10] W.-C. Chen, J. N. Schmidt, D. Yan, Y. K. Vohra, and C.-C. Chen, "Machine learning and evolutionary prediction of superhard B-C-N compounds", npj Computational Materials, vol. 7, July 2021.