Sheng *et al. J. Mater. Inf.* **2025**, *5*, 7
DOI: 10.20517/jmi.2024.75

Journal of Materials Informatics

Research Article

Open Access

Check for updates

Interpretable model of dielectric constant for rational design of microwave dielectric materials: a machine learning study

Ye Sheng[1,#], Yabei Wu[1,#,*], Chang Jiang[1,#], Xiaowen Cui[1], Yuanqing Mao[1], Caichao Ye[1,2,*], Wenqing Zhang[1,*]

[1]Department of Materials Science and Engineering & Institute of Innovative Materials, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China.
[2]Academy for Advanced Interdisciplinary Studies & Guangdong Provincial Key Laboratory of Computational Science and Material Design, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China.
[#]Authors contributed equally.

[*]**Correspondence to:** Dr. Yabei Wu, Dr. Caichao Ye, Prof. Wenqing Zhang, Department of Materials Science and Engineering & Institute of Innovative Materials, Southern University of Science and Technology, No. 1088 Xueyuan Avenue, Nanshan District, Shenzhen, Guangdong 518055, China. E-mail: wuyb3@sustech.edu.cn; yecc@sustech.edu.cn; zhangwq@sustech.edu.cn

How to cite this article: Sheng, Y.; Wu, Y.; Jiang, C.; Cui, X.; Mao, Y.; Ye, C.; Zhang, W. Interpretable model of dielectric constant for rational design of microwave dielectric materials: a machine learning study. *J. Mater. Inf.* **2025**, *5*, 7. https://dx.doi.org/10.20517/jmi.2024.75

Received: 18 Nov 2024 **First Decision:** 12 Dec 2024 **Revised:** 28 Dec 2024 **Accepted:** 31 Dec 2024 **Published:** 8 Feb 2025

Academic Editors: Shijun Zhao, Sergei Manzhos **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Machine learning (ML) has advantages in studying fundamental properties of materials and comprehending structure-property correlations. In this study, we employed sure independence screening and sparsifying operator (SISSO) method (ML technique) to explore the experimental dielectric constant, temperature coefficient of frequency resonator, and quality factor of inorganic oxide microwave dielectric materials. Among the constructed white-box models, the highest accuracy, with a coefficient of determination (R^2) of 0.8, was observed in predicting the dielectric constants of the quaternary materials. Additionally, we proposed a straightforward strategy to merge the ternary and quaternary datasets in a single training, aiming to address the issue of data scarcity in ML research. Although this strategy slightly compromises the model accuracy, it has the advantage of creating a more unified trained model for structural-property relationship understanding. Using the unified and interpretable model trained with the merged dataset, we derived a general rule governing the dielectric constant of materials. Our ML findings regarding the dielectric property provide fundamental insights for designing microwave dielectric materials with diverse dielectric constants.

Keywords: Dielectric constant, machine learning, interpretable model, merged dataset

INTRODUCTION

Microwave dielectric materials have a wide range of applications in the fields of communication, the internet, and the military[1,2]. The dielectric constant (ε_r), quality factor (Q_f), and temperature coefficient of the frequency resonator (τ_f) are three crucial dielectric properties[3] that significantly impact the performance of these dielectric materials. Therefore, designing microwave dielectric materials with optimized dielectric parameters is of utmost importance. However, the traditional trial-and-error approach requires a substantial number of experimental resources. Consequently, there is a growing interest in leveraging high-performance computer-assisted research to minimize experimental consumption and accelerate the discovery of novel materials[4-9].

In recent years, high-throughput density functional theory (DFT) calculations have been employed to determine the dielectric constant of materials[10-13]. Nevertheless, the computational requirements remain high when dealing with a large number of materials or complex systems. With the advancements in machine learning (ML) techniques, considerable efforts have been made in materials research to extract key factors influencing material properties from extensive datasets. Currently, ML methods have been applied to investigate the dielectric property[14-18] using both theoretical and experimental data. However, these investigations are classified as "black-box" methods, making physical evaluations of the resulting models for application in material design difficult. Moreover, previous studies predominantly focused on the dielectric constant, neglecting the exploration of the remaining two important parameters of microwave dielectric materials.

In this work, we gathered and organized experimental data on the dielectric properties of microwave dielectric materials. A total of 1,419 single-phase material data on dielectric properties were included. After applying several ML regression methods on the cleaned dataset, we finally employed the sure independence screening and sparsifying operator (SISSO) method[19] because of its interpretability. By considering the elemental properties, composition information, and structural details of the material, we constructed a feature space for ML analysis. To reduce the dimension of the features, we utilized the random forest (RF) method[15] to filter important features. The filtered features were used in the SISSO training process to develop a predictive model for the dielectric properties of materials. Additionally, we addressed the challenge of varying feature space sizes due to different numbers of components in materials by employing a simple chemical formula splitting approach to merge ternary and quaternary datasets. This merging strategy enabled us to train a larger dataset and develop a unified predictive model, albeit with a slight sacrifice in accuracy. Finally, through analysis of the trained model, we derived general trends regarding the influence of material features on dielectric properties, providing valuable guidance for future microwave dielectric materials design.

METHODS

Data collection and processing

The microwave dielectric properties, including ε_r, τ_f and Q_f, are gathered from previous work based on experiment measurements[1,2,18,20-23]. The dataset encompasses various types of materials, including single-phase, multiphase, and glass compounds. To streamline our analysis, we will mainly focus on a subset of 1,419 single-phase materials, comprising 1,419 ε_r, 1,188 τ_f, and 1,320 Q_f values. Within this single-phase dataset, materials were further categorized into binary, ternary, quaternary, quintuple, and six-member materials.

Feature space construction

Figure 1 represents the schematic framework of our ML approach. The feature space was constructed based on the atomic properties[24,25], compositional features, crystal structure features, and ionic properties of the materials in the database. The elemental compositions were normalized according to the chemical formulas of the materials. Crystal structure information (coordination number) was obtained from the Materials Project database[13], resulting in 563 crystal structures. Then, the atomic coordination numbers of each element in the chemical formula for these structures are analyzed. For materials without crystal structure information, we compared their chemical formulas with those having crystal structures and matched the atomic coordination number approximately if the chemical formula is similar, and the coordination number of the corresponding atomic position is the same. The elemental valence states were determined from the chemical formulas, and in combination with the coordination numbers, we derived the corresponding ionic radii[26].

Based on the feature space construction method, each element corresponded to 64 feature information, including 60 elemental features, composition, coordination number, elemental valence, and ionic radius. Therefore, the feature space for each material consists of $n \times 64$ features, where n is the number of elemental species in the material. However, the computational requirements of the SISSO method increase significantly as the feature space grows. To mitigate this issue, we employed the RF method[15] for the dimensionality reduction of the feature space. Specifically, the impurity-based feature importance measure within the RF method was employed, commonly referred to as Gini importance. This measure evaluates the significance of a feature by assessing the change in Gini impurity that results from splitting on that feature. The Gini impurity reduction across all trees in the forest provides an aggregate measure of feature importance. By leveraging RF calculations, we were able to ascertain the relative importance of each feature within our feature space. Then, only a few features with degrees of high importance were selected to construct the SISSO feature space.

Achieving an adequate model to characterize material properties often requires the use of high-dimensional descriptors. However, the computational resources required for training high-dimensional descriptors in SISSO are extremely demanding. Therefore, an alternative strategy is necessary to overcome this challenge. Here we propose an iterative running approach to enhance the complexity of SISSO output descriptors. Each execution of the SISSO generates new descriptors by mathematically combining descriptors from the feature space. These new descriptors are then added as new features to the original feature space, and the process continues iteratively. This iterative running and feature space updating method facilitates rapid increases in the dimensions of the output model generated by the SISSO method. The schematic descriptions are illustrated in Figure 1. The iterative running approach, combined with feature space updates, allows for continuous improvement in model complexity and prediction accuracy.

RESULTS AND DISCUSSIONS

ML-based dielectric properties for quaternary/ternary materials

The distribution of ε_r in the database shows that over 96% of the data falls within the range of 0 to 100, with only a small number of materials having ε_r higher than 100. To ensure balanced representation in the training set, we restrict the subsequent training to materials with ε_r below 100. Similarly, about 90% of the collected materials have the τ_f ranging from -100 to 100 ppm/K. Additionally, more than 93% of the Q_f values are within the range of 0 to 1.05×10^5 GHz. Therefore, the training set is also limited to materials within these ranges to ensure adequate representation.

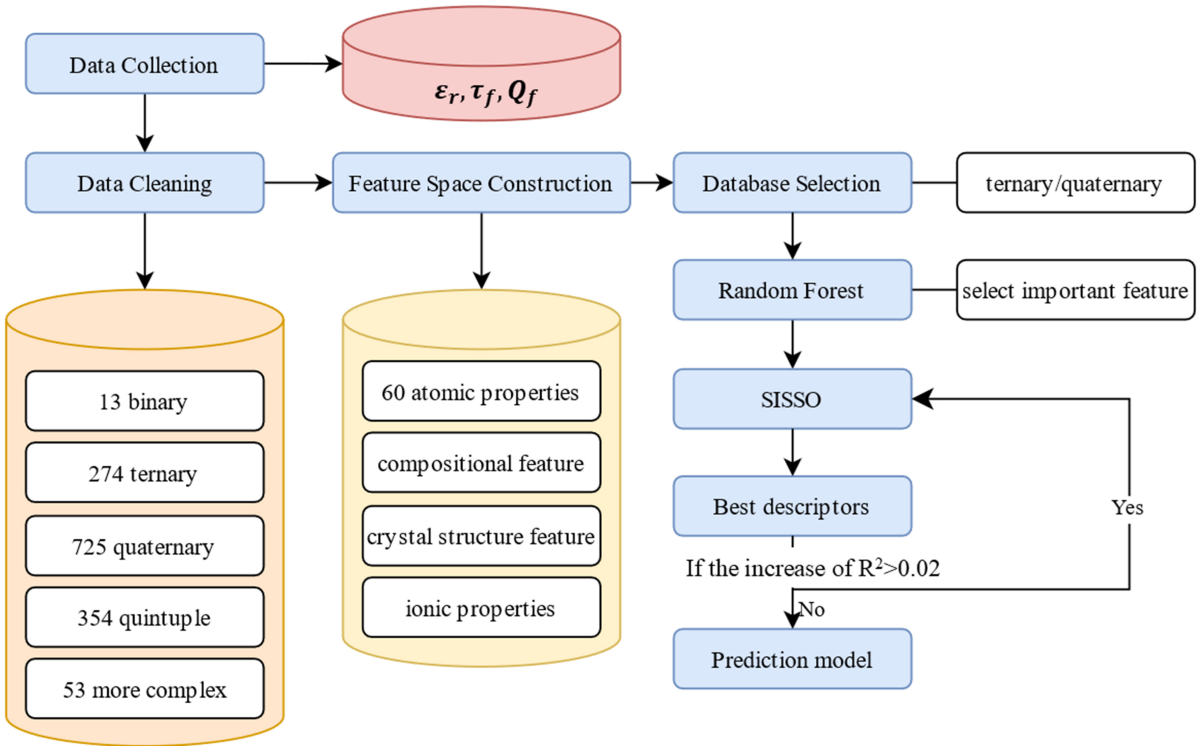

Figure 1. The schematic framework of our ML. In SISSO-based ML, an iterative strategy is chosen to enhance the complexity of the model. The iterative processing will be considered convergence when the increase of R^2 is less than 0.02. ML: Machine learning; SISSO: sure independence screening and sparsifying operator; R^2: coefficient of determination.

Given the challenges of including datasets with different numbers of components and feature space dimensions in the same training set, we initially focus on ML of dielectric properties for only quaternary materials since they make up 51% of the single-phase dataset. After restricting the ε_r values within the training set, 571 dielectric constants for the quaternary material are left, which will ultimately be utilized for training. The statistical comparisons of the distribution of the original and cleaned (for ternary and quaternary materials) datasets for the dielectric properties are represented in Figure 2.

Using the feature space construction method described earlier, we generated a total of 256 (4 × 64) features for each material in the quaternary ε_r datasets. Prior to engaging in SISSO-based model predictions, we conducted a thorough evaluation using a suite of numerical regression algorithms. These included gradient boosting regression (GBR), random forest regression (RFR), adaptive boosting regression (ABR), kernel ridge regression (KRR), and support vector regression (SVR), all aimed at predicting the quaternary dielectric constant. To ensure the robustness of our analysis, the entire dataset was randomly partitioned, allocating 80% for training and the remaining 20% for testing. The performance of the optimized models derived from these algorithms was rigorously assessed using two key metrics, as depicted in Figure 3: the coefficient of determination (R^2) to evaluate the strength of the correlation, and the root-mean-square error (RMSE) to quantify the prediction error. Figure 3 reveals that the GBR method emerged as the most effective, exhibiting the highest R^2 value of 0.80 and the lowest RMSE of 6.93 on the testing set. The predictive outcomes are graphically represented as scatter plots in Figure 4, with distinct visualizations for data from both the training and testing sets.

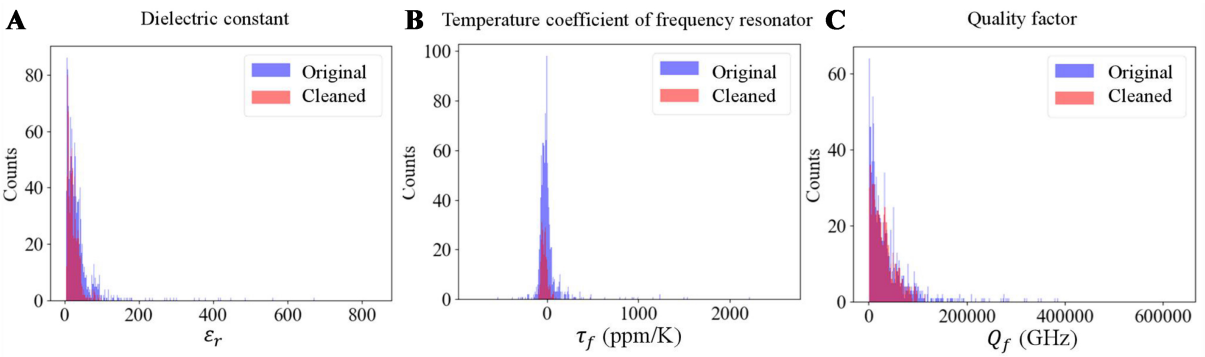

Figure 2. The distribution of the original and cleaned (selected ternary and quaternary materials) database. (A-C) represent the dielectric property of ε_r, τ_f, Q_f, respectively.

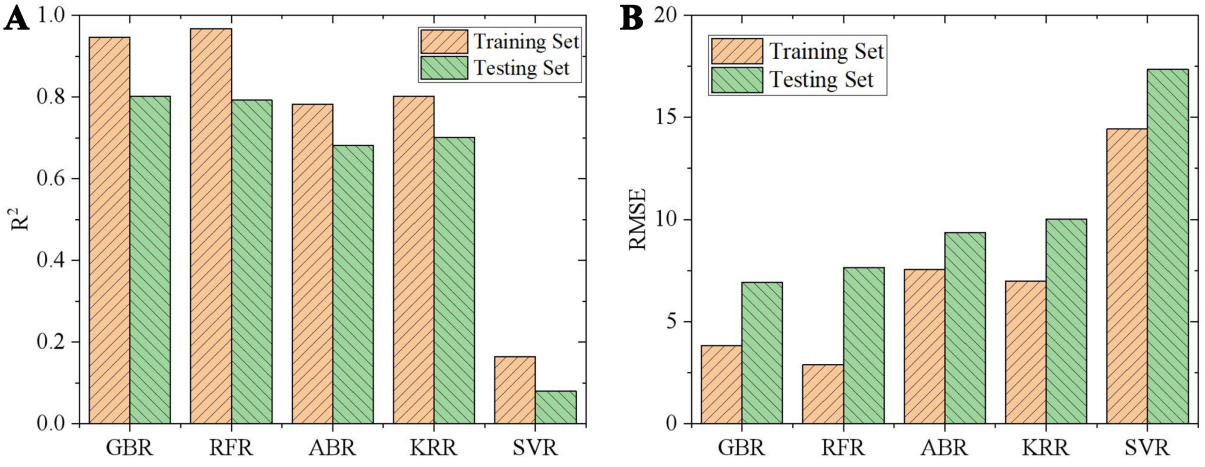

Figure 3. Comparison of the predictability of the five different regression models. (A) R^2; (B) RMSE. R^2: Coefficient of determination; RMSE: root-mean-square error.

While the GBR model demonstrated remarkable accuracy, the results from our five-fold cross-validation are unacceptable, which yielded an RMSE of 11.95 and, notably, a negative R^2 value, underscores the lack of generalization capability. These findings, coupled with the inherent limitations of "black-box" models in providing interpretable physical insights, have prompted us to seek alternative modeling strategies. Our subsequent focus on symbolic regression using the SISSO approach is driven by the need for a more transparent and interpretable model that can elucidate the underlying relationships within the data.

In pursuit of this goal, we turned our attention to feature selection, a critical step in preparing for SISSO analysis. To reduce the dimensionality for SISSO and identify the most influential features, we employed an RF-based importance ranking, as depicted in Figure 5A. This ranking is essential for understanding how each feature correlates with the dielectric constant ε_r and for guiding the selection of key predictors in our symbolic regression model. The symbols "A", "B", "C", and "D" represent the four components present in the quaternary material system. In our database, the quaternary material components A-B-C-D are arranged according to the following rules, which may differ from the traditional arrangement in the chemical formula. For instance, A is typically alkaline-earth metals or rare-earth elements, C is transition metals, B can be any element found in A or C, and D is always oxygen. Furthermore, the order of elements in the chemical formula is usually sorted from left to right based on the electronegativity, with lower

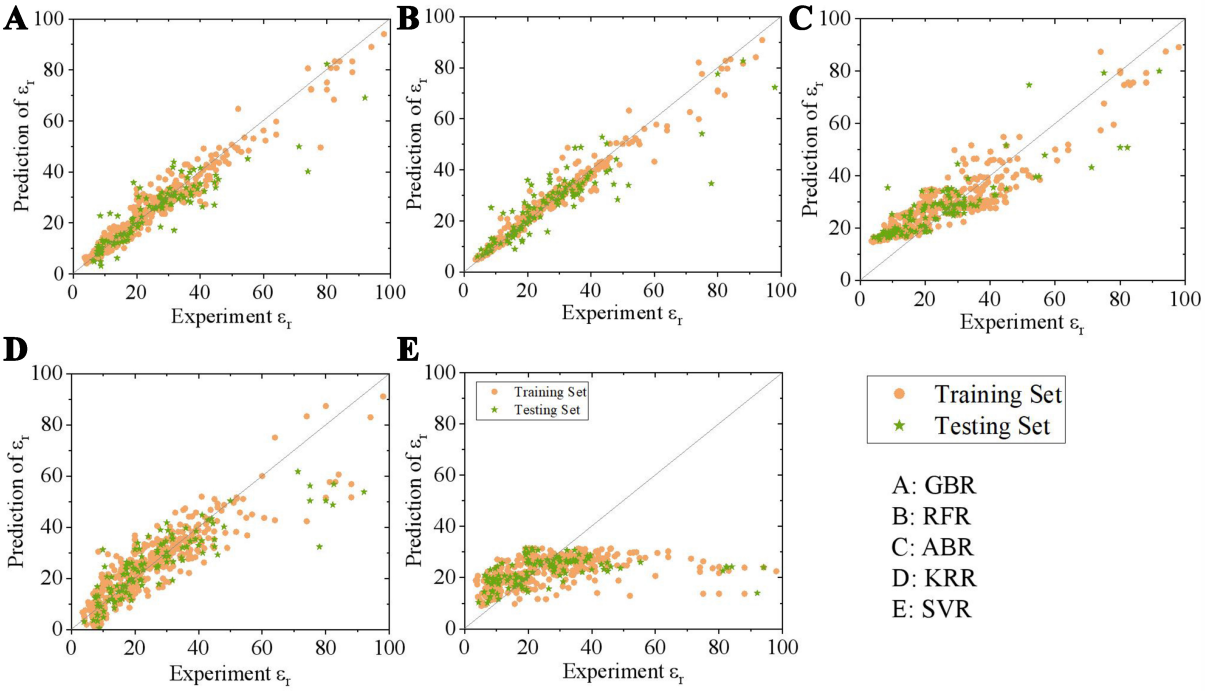

Figure 4. Scatter plot of the predicted dielectric constant versus the experiment reports using different ML models. (A) GBR; (B) RFR; (C) ABR; (D) KRR; and (E) SVR. ML: Machine learning; GBR: gradient boosting regression; RFR: random forest regression; ABR: adaptive boosting regression; KRR: kernel ridge regression; SVR: support vector regression.

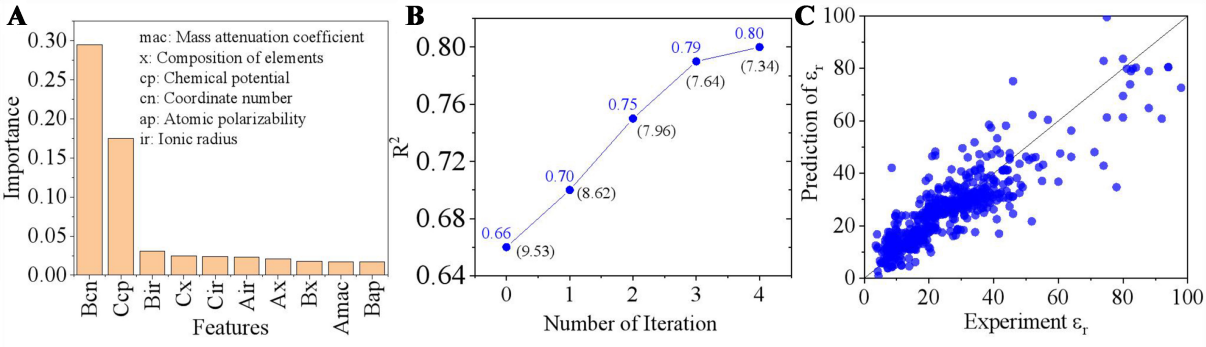

Figure 5. ML results of ε_r for quaternary materials. (A) RF ranking of the importance of material features for ε_r in the quaternary material dataset; (B) The variation of R^2 and RMSE (in the black bracket) as a function of SISSO iterations; (C) Scatter plot of the predicted dielectric constant versus experiment reports after four iterations. ML: Machine learning; RF: random forest; R^2: coefficient of determination; RMSE: root-mean-square error; SISSO: sure independence screening and sparsifying operator.

electronegativity on the left and higher electronegativity on the right.

Six most significant features are selected for constructing the SISSO training feature space ($4 \times 6 = 24$ for each material), including element ratio (x), coordination number (cn), ionic radius (ir), atomic chemical potential (cp), atomic mass attenuation coefficient for CrK_α (mac), and atomic polarizability (ap), as shown in Figure 5A. These features are used to train the SISSO model iteratively, with new descriptors generated in each round of training. The training results are shown in Figure 5B and C. After four iterations, the prediction model achieved an R^2 accuracy of 0.8. The equation representing the prediction model can be found in Supplementary Equation (1). We will delve into a detailed discussion later.

A five-fold cross-validation was also performed. The results of these cross-validation iterations are presented in Table 1. They demonstrate that our model not only achieves high accuracy on the training data but also maintains consistent performance across different validation sets. This consistency indicates that our SISSO-based model is not overfitted with the training data and has a strong generalization capability. Based on this foundation, we further investigate the impact of utilizing a newly optimized polarizability database[27] on our model's accuracy. Incorporating this database, we observed a comparable level of accuracy, with an R^2 of 0.78 and an RMSE of 7.68. This additional analysis, while not altering our primary conclusions, provides valuable insights into the sensitivity of our model to different polarizability values and further underscores its reliability.

The same approach is applied to study the τ_f and Q_f of the quaternary materials. The feature space is expanded to include the dielectric constant as an additional feature, resulting in a total of $4 \times 64 + 1 = 257$ features. The RF importance ranking of these features for τ_f and Q_f are shown in Supplementary Figure 1A and 2A, respectively. Prediction models for τ_f (Q_f) are developed using this feature space, and their performance is depicted in Supplementary Figure 1B and C (Supplementary Figure 2B and C), as given in Supplementary Equations (2 and 3).

We observe that the prediction models for τ_f and Q_f have lower accuracies than the model for ε_r. This discrepancy is likely due to the influence of extrinsic factors, such as sintering temperature, which can cause variations in τ_f and Q_f, leading to inconsistencies across different studies[28-35]. Currently, these external factors are not reflected in our feature space, which may also be incomplete, missing critical information pertinent to τ_f and Q_f. Consequently, these limitations can affect the models' ability to accurately fit the data.

Similar investigations were performed for the ε_r, τ_f and Q_f ternary dielectric materials. The results and formulas of these investigations can be found in Supplementary Figures 3-5 and Supplementary Equations (4-6). The observations made in the ternary materials align with those in the quaternary materials, leading us to concentrate on dielectric constant predictions due to their more accurate prediction model. It was observed that the dielectric constant formulas for quaternary and ternary dielectric materials [Supplementary Equations (1 and 4)] exhibit a comparable trend in terms of the influence of atomic characteristics, such as mass attenuation coefficient and atomic polarizability of elements at different positions, on the calculation results. However, there are differences in the impact of material composition information and coordination numbers on the prediction of ε_r between the ternary and quaternary models. These differences suggest variations in the data distribution of dielectric constant datasets between ternary and quaternary datasets.

These findings emphasize the significance of merging the datasets to create a unified prediction model capable of capturing common trends and characteristics across multi-component material systems. By combining data from ternary and quaternary materials, it becomes possible to develop a more comprehensive and robust prediction model that can generalize well to different compositions and structures. This would enhance our understanding of the structure-property relationships in dielectric materials and enable more accurate predictions for a wider range of material systems.

ML of dielectric constant for merged multi-component dataset

Considering the characteristics of our dataset, which primarily consists of inorganic oxide materials, it is worth noting that most quaternary materials are obtained by doping ternary materials at specific cationic sites. Therefore, we can see the ternary materials as quaternary by splitting the cation sites in the chemical formula. The ternary oxide ($A_iB_jO_k$) can be split in two different ways: $A_iB_jO_k \rightarrow A_{xi}A_{(1-x)i}B_jO_k$ or $A_iB_jO_k \rightarrow$

Table 1. Cross-validation results for dielectric constant using the quaternary dataset

Training groups	Testing group	Training R^2	Training RMSE	Testing R^2	Testing RMSE
1-4	5	0.77	7.75	0.75	8.56
1-3, 5	4	0.79	7.35	0.72	9.40
1, 2, 4, 5	3	0.77	7.84	0.76	7.67
1, 3-5	2	0.76	8.16	0.75	6.94
2-5	1	0.78	7.55	0.71	8.99
Average		0.77	7.73	0.74	8.31

R^2: Coefficient of determination; RMSE: root-mean-square error.

$A_iB_{xj}B_{(1-x)j}O_k$. By employing the splitting methods, the ternary oxide materials are transformed into quaternary-like materials. In the feature space construction, we duplicate the feature information of the corresponding site. This allows us to merge the ternary oxides with the original quaternary oxides, thereby expanding the dataset in a single training.

Before proceeding with the ML analysis of the merged dataset, it is crucial to understand the potential impact of splitting methods and splitting ratio (x) on the training results. To address this, we perform ML studies on the quaternary-like dataset (the split ternary dataset) using the two distinct splitting methods and varying splitting ratios. We found that neither the splitting methods nor the splitting ratio had an effect on the results when compared to those obtained from the original ternary dataset. The prediction accuracy of R^2 remains consistent with that using the original training model. This observation provides a robust basis for merging the quaternary-like dataset with the original quaternary dataset, thus supporting further analysis.

In the subsequent ML training using the merged dataset (quaternary and quaternary-like), we adopted a splitting ratio of $x = 0.5$. It is worth noting that the choice of splitting methods, $A_iB_jO_k \rightarrow A_{0.5i}A_{0.5i}B_jO_k$ or $A_iB_jO_k \rightarrow A_iB_{0.5j}B_{0.5j}O_k$, may potentially influence the training model for the merged dataset. Consequently, we consider three different splitting strategies in our simulations: (1) exclusively splitting the A sites (Strategy-A); (2) exclusively splitting the B sites (Strategy-B); and (3) randomly selecting either A or B sites for splitting (Strategy-AB).

Figure 6A compares the training results of different merging strategies with the original training model. The models trained using the combined datasets are consistently less accurate than those trained using either dataset alone. Furthermore, we observed differences in accuracy among the different merging strategies when no iteration was performed. Notably, Strategy-B achieved the highest accuracy. This is likely because, in the quaternary dataset, the main doping site is at the B site, which aligns with the actual doping situation for the B site in the quaternary materials. However, after a few iterations, we found that the accuracy of all three strategies became much closer. This indicates that as the model complexity increases, the choice of merging strategies no longer affects the overall accuracy.

To analyze the reason behind the decrease in accuracy of the merged dataset, we speculate that it is primarily due to the differing distribution characteristics of the ternary and quaternary material dielectric constant datasets. To demonstrate this, we present the prediction accuracies for the corresponding ternary/quaternary database using different models (trained by Strategy-AB, the original quaternary and ternary datasets) in Figure 6B and C. For the single quaternary dataset, the model trained with Strategy-AB generally exhibits lower prediction accuracy than the model trained solely on the quaternary dataset. However, the overall trend with the number of iterations aligns with the expectation, as indicated by the

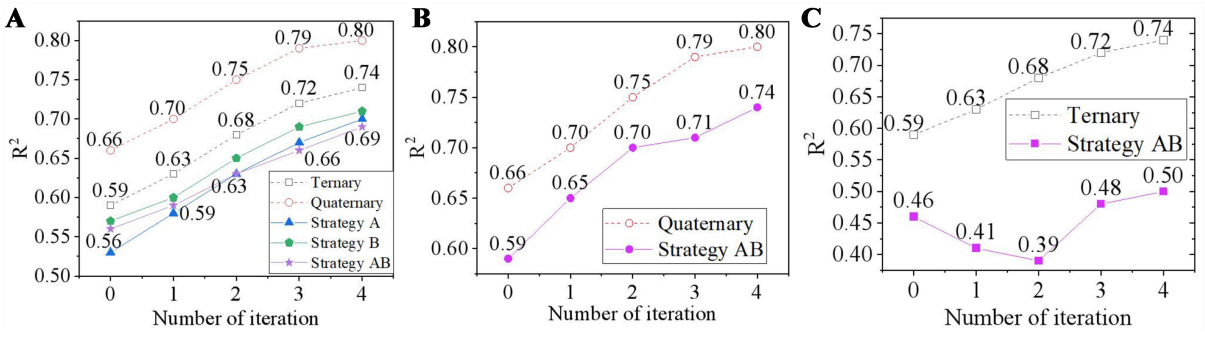

Figure 6. R^2 values for the merged dataset of ternary and quaternary. (A) The comparison of R^2 obtained by the merged dataset with three different strategies and the training model for the ternary (quaternary) dataset alone; (B) Comparison of the predicted R^2 for the original quaternary dataset by using the models obtained by strategy-AB and the original quaternary dataset; (C) Comparison of the predicted R^2 for the quaternary-like (ternary) dataset by using the models obtained by Strategy-AB and the original ternary dataset. R^2: Coefficient of determination.

purple stars in Figure 6A. On the other hand, the prediction accuracies of the quaternary-like (ternary) dataset using the Strategy-AB trained model are consistently worse overall, and they do not show a consistent pattern with iterations. This demonstrates that the ε_r for the ternary material dataset used in this study has a distinct distribution compared to the ε_r for the quaternary material dataset. Given that the amount of quaternary material data used is 2.5 times larger than the ternary material data, merging these differently distributed datasets and training a model on them results in a final model that tends to be closer to the quaternary dataset and deviates from the ternary dataset. As a result, this leads to a decrease in the prediction accuracy of the model for either single quaternary or single ternary datasets.

Interpretation of the model and rational design

We employed an iterative training approach to update the feature space in our research. However, as we progressed through several iterations, the resulting trained model became excessively intricate, which presented difficulties in terms of analysis. Nonetheless, we observed that the model's prediction accuracy improved after two iterations compared to a non-iterative model. Therefore, to facilitate a more thorough investigation of the underlying physics, we opted to analyze the model after two training iterations. A generalized fitting model, developed by Strategy-AB, for the merged ternary-quaternary oxide system (the corresponding parameters are listed in Supplementary Table 1), is given as:

$$\varepsilon_r = 2.8 \times 10^{-4} \frac{A_{ap} \cdot A_{cn} \cdot B_{cn} \cdot C_{ap} \cdot C_{mac} \cdot D_x \cdot cos\,(C_x)}{A_{ap} + C_{ap}} + 0.2 \frac{A_{cp} \cdot C_{cp} \cdot D_x^2 \cdot (A_{mac} + B_{mac}) \cdot cos\,(C_x)}{e^{C_{cp}}} + 6.4 D_x \cdot cos\,(C_x) - 1.4 \times 10^{-4}\, cos\left(A_{ap}\right) \cdot e^{B_{cn}} + 1.16$$

In order to gain a clearer understanding of the equation mentioned above, the normalized contribution of each term to the dielectric constant is estimated. The results are depicted in Figure 7, clearly illustrating the distinct contribution ratios of each term. In materials with high dielectric constants, the first two terms have a dominant influence on the dielectric constant. Therefore, when designing systems with high dielectric constants, it is crucial to consider the characteristics and variations represented by these terms. Conversely, the third term becomes the primary factor influencing the dielectric constant in materials with low dielectric constants (e.g., less than 25). Hence, when designing systems with lower dielectric constants, it is essential to consider the characteristics encompassed by all three terms.

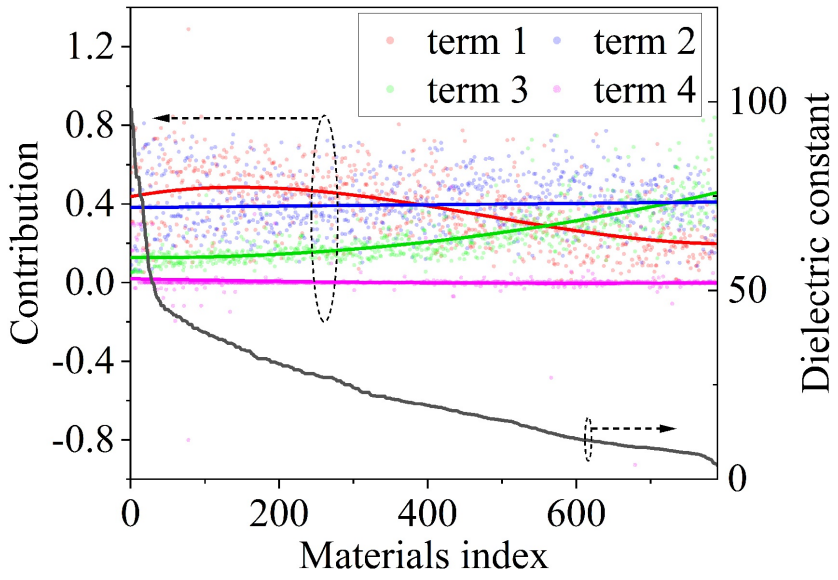

Figure 7. Normalized contribution of the dielectric constant prediction given by each term in the trained model using the Strategy-AB dataset after two iterations. The black solid line shows the variation of the dielectric constant from large to small. The colored dots correspond to the contribution of the first to the fourth term (red-blue-green-magenta) in the equation to the dielectric constant, while the corresponding solid lines represent the fitting of these respective contributions.

In detail, the following findings were observed:
(a) A positive correlation between the dielectric constant and the mass attenuation coefficient of elements A, B, and C.
(b) The dielectric constant is positively correlated with the chemical potential of the element in the A site and negatively correlated with the chemical potential of the element in the C site.
(c) The dielectric constant is negatively correlated with the atomic polarizability of the element in the C position.
(d) The dielectric constant is positively correlated with the normalized number of elements in C and D sites in the chemical formula.

These findings provide some general guidelines. To achieve lower dielectric constants, the A position can be filled with elements such as Li, Na, and Mg, which have lower elemental mass attenuation coefficients. Additionally, Li and Na possess suitable chemical potential values for achieving low dielectric constants. In the C position, choosing B (boron) is advantageous for minimizing the dielectric constant due to its lower elemental mass attenuation coefficient, which consists of previous experimental reports very well[30,36,37]. Despite boron has low atomic polarizability, it is the mass attenuation coefficient that predominantly influences the reduction of the dielectric constant. This subtlety underscores the intricate balance of atomic properties that collectively dictate dielectric performance. Conversely, for materials with higher dielectric constants, the A position can be occupied by elements such as Ag, Pb, and Bi, which have large mass attenuation coefficients and chemical potential. In the C position, Zr is the most favorable choice for achieving high dielectric constants due to its chemical potential and atomic polarizability values. Additionally, Ti is also more likely to result in large dielectric constants.

Our results offer valuable insights and guidelines for the rational design and selection of elements in the different positions to create new dielectric materials with desired dielectric constants. By considering the elemental properties, such as mass attenuation coefficients, chemical potentials, and atomic polarizability,

researchers can make informed choices in the design process. These guidelines provide a framework for developing dielectric materials with tailored properties, contributing to advancements in material science and technology.

CONCLUSION

In this study, we employed the SISSO method to investigate the dielectric properties of microwave dielectric materials, specifically focusing on the dielectric constant, quality factor, and temperature coefficient of the frequency resonator. A predictive model with high interpretability was developed. While the model's accuracy was influenced by varying experimental conditions, the results regarding the dielectric constants were particularly valuable. To address the challenge of merging datasets with different numbers of components, a straightforward splitting chemical formula approach was proposed. This method allows for training on a combined (larger) dataset of dielectric constants for multi-component compounds. The analysis of the prediction model also revealed a general guideline for tuning the dielectric constant of materials. The proposed approach of combining datasets with different numbers of elements for dielectric properties can be applied to other properties in future ML investigations of multi-component materials. Moreover, the insights acquired from evaluating the prediction model offer valuable guidance for developing future microwave dielectric materials.

DECLARATIONS

Acknowledgment

We acknowledge the computational support provided by the Center for Computational Science and Engineering at Southern University of Science and Technology.

Authors' contributions

Made substantial contributions to conception and design of the study and performed data analysis and interpretation: Sheng, Y., Wu, Y., Jiang, C., Cui, X., Mao, Y.

Performed data acquisition and provided administrative, technical, and material support: Wu, Y., Ye, C., Zhang, W.

Availability of data and materials

The corresponding data have been uploaded to the author's GitHub repository: https://github.com/yabeiwu/DielectricProperties/tree/main/ExpDiele. Further data are available from the corresponding authors upon request.

Financial support and sponsorship

This work is supported in part by the National Natural Science Foundation of China (No. 12104207), Guangdong Provincial Key Laboratory of Computational Science and Material Design (2019B030301001), and High level of special funds (G03050K002). Zhang, W. also acknowledges support from the Guangdong Innovation Research Team Project (2017ZT07C062).

Conflicts of interest

Zhang, W. serves as an Advisory Editor for *Journal of Materials Informatics* but was not involved in any part of the editorial process, including the selection of reviewers, manuscript handling, or decision-making. The other authors declare no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

REFERENCES

1. Sebastian, M. T. Dielectric materials for wireless communication. 1st Edition. Elsevier: 2008. DOI
2. Sebastian, M. T.; Ubic, R.; Jantunen, H. Low-loss dielectric ceramic materials and their properties. *Int. Mater. Rev.* **2015**, *60*, 392-412. DOI
3. Reaney, I. M.; Iddles, D. Microwave dielectric ceramics for resonators and filters in mobile phone networks. *J. Am. Ceram. Soc.* **2006**, *89*, 2063-72. DOI
4. Schmidt, J.; Marques, M. R. G.; Botti, S.; Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *npj. Comput. Mater.* **2019**, *5*, 221. DOI
5. Wang, A. Y.; Murdock, R. J.; Kauwe, S. K.; et al. Machine learning for materials scientists: an introductory guide toward best practices. *Chem. Mater.* **2020**, *32*, 4954-65. DOI
6. Gomes, C. P.; Fink, D.; van, D. R. B.; Gregoire, J. M. Computational sustainability meets materials science. *Nat. Rev. Mater.* **2021**, *6*, 645-7. DOI
7. Barr, J. A.; Lin, F.; Ashton, M.; Hennig, R. G.; Sinnott, S. B. High-throughput density functional calculations to optimize properties and interfacial chemistry of piezoelectric materials. *Phys. Rev. Mater.* **2018**, *2*, 025002. DOI
8. Schleder, G. R.; Padilha, A. C. M.; Acosta, C. M.; Costa, M.; Fazzio, A. From DFT to machine learning: recent approaches to materials science - a review. *J. Phys. Mater.* **2019**, *2*, 032001. DOI
9. Jain, A.; Hautier, G.; Moore, C. J.; et al. A high-throughput infrastructure for density functional theory calculations. *Comput. Mater. Sci.* **2011**, *50*, 2295-310. DOI
10. Petousis, I.; Mrdjenovich, D.; Ballouz, E.; et al. High-throughput screening of inorganic compounds for the discovery of novel dielectric and optical materials. *Sci. Data.* **2017**, *4*, 160134. DOI PubMed PMC
11. Petousis, I.; Chen, W.; Hautier, G.; et al. Benchmarking density functional perturbation theory to enable high-throughput screening of materials for dielectric constant and refractive index. *Phys. Rev. B.* **2016**, *93*, 115151. DOI
12. Umeda, Y.; Hayashi, H.; Moriwake, H.; Tanaka, I. Prediction of dielectric constants using a combination of first principles calculations and machine learning. *Jpn. J. Appl. Phys.* **2019**, *58*, SLLC01. DOI
13. Jain, A.; Ong, S. P.; Hautier, G.; et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL. Mater.* **2013**, *1*, 011002. DOI
14. Takahashi, A.; Kumagai, Y.; Miyamoto, J.; Mochizuki, Y.; Oba, F. Machine learning models for predicting the dielectric constants of oxides based on high-throughput first-principles calculations. *Phys. Rev. Mater.* **2020**, *4*, 103801. DOI
15. Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE. Trans. Pattern. Anal. Mach. Intell.* **1998**, *20*, 832-44. DOI
16. Gopakumar, A.; Pal, K.; Wolverton, C. Identification of high-dielectric constant compounds from statistical design. *npj. Comput. Mater.* **2022**, *8*, 832. DOI
17. Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *JOM.* **2013**, *65*, 1501-9. DOI
18. Qin, J.; Liu, Z.; Ma, M.; Li, Y. Machine learning approaches for permittivity prediction and rational design of microwave dielectric ceramics. *J. Materiomics.* **2021**, *7*, 1284-93. DOI
19. Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M.; Ghiringhelli, L. M. SISSO: a compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Mater.* **2018**, *2*, 083802. DOI
20. Neil Alford Home Page. Available online: https://www.imperial.ac.uk/people/n.alford. (accessed on 2024-01-25).
21. Shannon, R. D. Dielectric polarizabilities of ions in oxides and fluorides. *J. Appl. Phys.* **1993**, *73*, 348-66. DOI
22. Chen, X. M.; Qin, N.; Li, Y. Microstructures and microwave dielectric characteristics of $Ba_{6-3x}(Sm_{1-y}La_y)_{8+2x}Ti_{18}O_{54}$ solid solutions (x = 2/3 and 0.75). *J. Electroceram.* **2002**, *9*, 31-5. DOI
23. Okawa, T.; Imaeda, M.; Ohsato, H.; Harada, A. Site occupancy of Bi ions and microwave dielectric properties in Bi-doped $Ba_{6-3x}R_{8+2x}Ti_{18}O_{54}$ (R = rare earth, x = 2/3) solid solutions. *Mater. Chem. Phys.* **2003**, *79*, 199-203. DOI
24. A. A. Baikov Institute of Metallurgy and Materials Science. Available online: https://phases.imet-db.ru/elements (accessed on 2024-01-25).
25. Choudhary, S.; Ranjan, P.; Chakraborty, T. Atomic polarizability: a periodic descriptor. *J. Chem. Res.* **2020**, *44*, 227-34. DOI
26. Shannon, R. D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta. Cryst. A.* **1976**, *32*, 751-67. DOI
27. Qin, J.; Liu, Z.; Ma, M.; Li, Y. Optimizing and extending ion dielectric polarizability database for microwave frequencies using machine learning methods. *npj. Comput. Mater.* **2023**, *9*, 1093. DOI

28. Guo, M.; Li, Y.; Dou, G.; Lin, J. A new microwave dielectric ceramics for LTCC applications: $Li_2Mg_2(WO_4)_3$ ceramics. *J. Mater. Sci. Mater. Electron.* **2014**, *25*, 3712-5. DOI
29. Zhang, J.; Zuo, R. Synthesis and microwave dielectric properties of $Li_2Mg_2(WO_4)_3$ ceramics. *Mater. Lett.* **2015**, *158*, 92-4. DOI
30. Wu, J.; Huang, H. Effect of crystallization on microwave dielectric properties of stoichiometric cordierite glasses containing B_2O_3 and P_2O_5. *J. Mater. Res.* **2000**, *15*, 222-7. DOI
31. Wang, F.; Lai, Y.; Zeng, Y.; et al. Enhanced microwave dielectric properties in $Mg_2Al_4Si_5O_{18}$ through Cu^{2+} substitution. *Eur. J. Inorg. Chem.* **2021**, *2021*, 2464-70. DOI
32. Yao, G. G.; Liu, P.; Zhang, H. W. Microwave dielectric properties of $Li_2MgTi_3O_8$ ceramics produced by reaction-sintering method. *J. Mater. Sci. Mater. Electron.* **2013**, *24*, 1128-31. DOI
33. George, S.; Sebastian, M. T. Microwave dielectric properties of novel temperature stable high Q $Li_2Mg_{1-x}Zn_xTi_3O_8$ and $Li_2A_{1-x}Ca_xTi_3O_8$ (A = Mg, Zn) ceramics. *J. Eur. Ceram. Soc.* **2010**, *30*, 2585-92. DOI
34. Wu, S.; Chen, D.; Jiang, C.; Mei, Y.; Ma, Q. Synthesis of monoclinic $CaSnSiO_5$ ceramics and their microwave dielectric properties. *Mater. Lett.* **2013**, *91*, 239-41. DOI
35. Du, K.; Song, X.; Li, J.; et al. Optimised phase compositions and improved microwave dielectric properties based on calcium tin silicates. *J. Eur. Ceram. Soc.* **2019**, *39*, 340-5. DOI
36. Došler, U.; Kržmanc, M. M.; Suvorov, D. The synthesis and microwave dielectric properties of $Mg_3B_2O_6$ and $Mg_2B_2O_5$ ceramics. *J. Eur. Ceram. Soc.* **2010**, *30*, 413-8. DOI
37. Chang, S.; Pai, H.; Tseng, C.; Tsai, C. Microwave dielectric properties of ultra-low temperature fired Li_3BO_3 ceramics. *J. Alloys. Compd.* **2017**, *698*, 814-8. DOI