

A Multiclass Classification Model for Predicting the Thermal Conductivity of Uranium Compounds

Y. Sun^{a,*}, M. Kumagai^{a,b,c}, M. Jin^a, E. Sato^a, M. Aoki^a, Y. Ohishi^d, and K. Kurosaki^{a,e,†}

^aInstitute for Integrated Radiation and Nuclear Science, Kyoto University, 2, Asashiro-Nishi, Kumatori, Sennan, Japan; ^bSAKURA internet Research Center, SAKURA internet Inc., Tokyo Tatemono Umeda Building 11F, 1-12-12, Umeda, Kita-ku, Osaka, Japan; ^cCenter for Advanced Intelligence Project, RIKEN, Nihonbashi 1-chome Mitsui Building, 15th floor, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, Japan; ^dGraduate School of Engineering, 2-1 Yamadaoka, Suita, Osaka, Japan; ^eResearch Institute of Nuclear Engineering, University of Fukui, 1-2-4 Kanawa-cho, Tsuruga, Fukui, Japan;

ARTICLE HISTORY

Compiled December 14, 2023

ABSTRACT

Advanced nuclear fuels are designed to offer improved performance and accident tolerance, with an emphasis on achieving higher thermal conductivity. While promising fuel candidates like uranium nitrides, carbides, and silicides have been widely studied, the majority of uranium compounds remain unexplored. To search for potential candidates among these unexplored uranium compounds, we incorporated machine learning to accelerate the material discovery process. In this study, we trained a multiclass classification model to predict a compound's thermal conductivity based on 133 input features derived from element properties and temperature. The initial training data consists of over 160,000 processed thermal conductivity records from the Starrydata2 database, but a skewed data class distribution led the trained model to underestimate compound's thermal conductivity. Consequently, we addressed the issue of class imbalance by applying Synthetic Minority Oversampling TEchnique and Random UnderSampling, improving the recall for materials with thermal conductivity higher than 15 W/mK from 0.64 to 0.71. Finally, our best model is used to identify 119 potential advanced fuel candidates with high thermal conductivity among 774 stable uranium compounds. Our results underscore the potential of machine learning in the field of nuclear science, accelerating the discovery of advanced nuclear materials.

KEYWORDS

Advanced nuclear fuels; Machine learning; Thermal conductivity

1. Introduction

After the Fukushima Daiichi nuclear power plant accident in March 2011, considerable attention has been focused on the development of accident-tolerant fuels and claddings that have greater accident tolerance and improved performance compared to the traditional UO₂-Zircaloy system[1–3] to promote safer nuclear power. Among the various

* CONTACT Y. Sun. Email: sun.yifan.7r@kyoto-u.ac.jp

† CONTACT K. Kurosaki. Email: kurosaki.ken.6n@kyoto-u.ac.jp

categories of advanced accident-tolerant fuels, high-density fuels such as uranium silicides, uranium carbides, and uranium nitrides show significant promise. Their high uranium density can offset the neutron penalties associated with the use of advanced cladding materials like FeCrAl or coated Zr-claddings[4,5]. Moreover, these uranium compounds exhibit excellent thermal conductivity, which is critical in enhancing reactor safety by reducing the centerline temperature of the fuel pellet and its temperature gradient[6–8].

Despite the promising properties and compatibility with advanced claddings, each of these high-density fuels has issues that must be addressed before their use in commercial light water reactors[4,9,10]. For instance, both UN[11,12] and UC[13,14] are known to react with water, forming hydrogen gases, while U_3Si_2 [15–18] rapidly pulverizes in H_2O -environments at elevated temperatures. These reactions pose significant safety concerns in the event of a cladding breach. Additionally, UN and UC exhibit higher swelling rates, necessitating further consideration to prevent pellet-cladding mechanical interactions[4]. Although it is crucial to continue improving the properties of these well-established advanced fuel candidates, the pursuit of new advanced fuels with high uranium density and thermal conductivity should not be overlooked. This is particularly relevant given that the vast majority of uranium compounds remain unexplored. However, it would be both time-consuming and impractical to employ the conventional trial-and-error approach to investigate the thermal conductivity of over 2000 unique uranium compounds[19].

Over the past few years, the incorporation of machine learning (ML) in materials science has revolutionized the material discovery process[20–22]. Instead of relying on human intuition and experiences, ML models, trained on extensive datasets, can efficiently predict material properties such as crystal structure[23–26], elastic properties[27–30], and thermal conductivity[31–34], contributing to the design and accelerated discovery of new materials. However, ML has seen little use in predicting the thermophysical properties of nuclear materials, likely due to the lack of a comprehensive nuclear-focused database to use as training data. To the best of our knowledge, only Kautz et al.[35] have introduced a deep learning model to predict the thermal conductivity of the U-Mo system after irradiation and Lu et al.[36] constructed a model to predict the thermal conductivity of metallic nuclear fuels limited to those with BCC structures.

To expedite the discovery of advanced fuels, this study employs an ML-based approach to efficiently identify potential fuel candidates among the vast unexplored uranium compounds. Specifically, our objective is to develop a multiclass random forest model for predicting the thermal conductivity of these compounds. In contrast to the regression model proposed by Lu et al.[36] for U-Mo-Nb-Zr-Pu alloys, which uses a limited feature set of 6 and was trained on a small dataset of 801 data points, our model leverages 133 input features, consisting of 132 Magpie descriptors[37] and temperature. Furthermore, our model was trained on a substantially larger dataset, comprised of 168,918 entries from the Starrydata2 database[38]. The availability of a comprehensive thermal conductivity database in Starrydata2 is also a factor in our decision to focus on predicting thermal conductivity, despite the importance of other properties relevant during advanced fuels’ fabrication, operation, and reprocessing.

However, it is worth noting that as the Starrydata2 database predominantly consists of materials with low thermal conductivity, it created an imbalanced training dataset, potentially affecting the accuracy of the ML model’s predictions. To address this, we employed the Synthetic Minority Oversampling Technique (SMOTE)[39] and Random Undersampling (RUS) to balance the classes in the dataset. Utilizing our classification

model, trained on a SMOTE-balanced dataset, we predicted the thermal conductivity of 774 stable uranium compounds listed on the Materials Project database[19], which provides the chemical formulas for various inorganic crystalline materials. Out of these, 119 were selected as potential advanced fuel candidates due to their high thermal conductivity, defined as exceeding 15 W/mK at temperatures between 300 and 1000 K.

2. Methods

2.1. Data filtering, featurization, and binning

The Starrydata2 database was initially established to promote material discovery in the field of thermoelectric materials. It contains various thermophysical properties extracted from peer-reviewed articles, such as composition, temperature, Seebeck coefficient, thermal conductivity, electrical resistivity, etc. We processed all 372,480 data points from the Starrydata2 database (as of 2023/01/12) for ML training.

From these 372,480 data points, we retained only those with measurement temperatures between 300 and 1000 K. The lower limit of 300 K was chosen because nuclear fuels will not experience lower temperatures during operation, and the upper limit of 1000 K was selected due to the scarcity of data at higher temperatures. Additionally, we filtered out thermal conductivities below 0 or above 500 W/mK. The upper thermal conductivity limit was set to eliminate outliers that are not polycrystals, considering even the best heat-conducting metals have thermal conductivities lower than 500 W/mK[40].

Among the Starrydata2 entries with available sample information, we excluded those labeled as single crystals, powders, films, etc. The thermophysical properties of these sample forms can significantly differ from those of their bulk counterparts, and our primary interest lies in the thermal conductivity of polycrystalline bulk samples in the context of nuclear fuels. Lastly, we removed entries with human errors, such as invalid elements and floats, resulting in a filtered dataset of 168,918 entries for training. Figure 1 offers a visual presentation of the thermal conductivity data removed during the data processing step.

We utilized a total of 133 features (132 Magpie descriptors[37] and temperature) for machine learning. The Magpie descriptors from the matminer[41] featurizers module provide extensive information on the elements' properties, such as their positions on the periodic table, electronic structures, and ionic characteristics, including mean, mode, average deviation, maximum, minimum, and range. The target variable, thermal conductivity, was categorized into three classes (0–5, 5–15, and 15+ W/mK), with class 2 (15+ W/mK) representing high thermal conductivity, and class 0 (0–5 W/mK) denoting thermal conductivity similar to that of UO_2 at elevated temperatures[42]. We opted for a classification study over a regression model because our primary goal is to discover uranium compounds with thermal conductivities comparable to those of uranium carbide (UC)[6] and uranium nitride (UN)[7], acknowledging that exact values can be more precisely determined via experiments. Diagrams depicting the data preparation and model training process are provided in Figure 2.

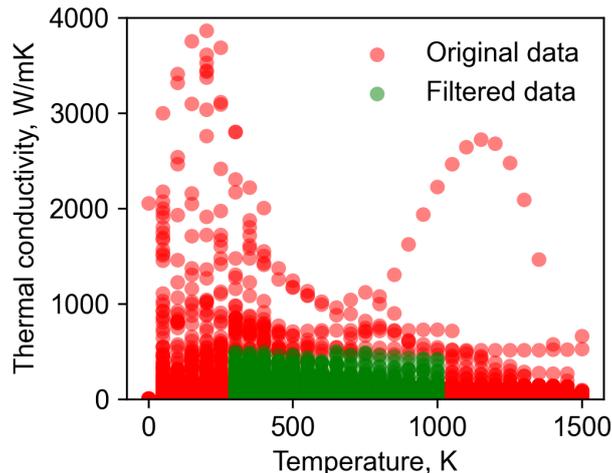


Figure 1. Distribution of thermal conductivity data before (red) and after (green) filtering. The y-axis limit is set at 4000 W/mK.

Table 1. Various class distribution and size of the training data obtained using SMOTE and RUS.

	Class 0	Class 1	Class 2	Ratio
Imbalance	139,402	21,703	7,813	18:3:1
SMOTE	139,402	139,402	139,402	1:1:1
SMOTE/RUS	69,701	69,701	69,701	1:1:1

2.2. Data balancing

Since the Starrydata2 database primarily focuses on thermoelectric materials, which typically have low thermal conductivity, its data distribution is significantly biased towards lower values, as shown in Figure 3. Without balancing the dataset, the trained model may be inclined to predict a ‘0’ value, thereby underestimating a material’s thermal conductivity.

To investigate the effect of data imbalance and data volume on the model’s performance, we utilized the Synthetic Minority Oversampling TEchnique (SMOTE)[39] and Random UnderSampling (RUS) methods from scikit-learn[43] to obtain three types of class distributions and dataset size, as shown in Table 1. SMOTE addresses imbalance by synthesizing new minority class data points. It operates by randomly selecting a data point from the minority class, finding its k -nearest neighbors in the feature space (five neighbors in this work), and creating a new data point at a randomly chosen point along the line connecting the selected data point and one of its neighbors. Conversely, RUS mitigates imbalance by randomly removing data points from the majority class, as its name suggests. Used in tandem, these techniques aim to attain various balances of class distribution and dataset sizes in the originally biased training data. For example, as shown in Table 1, in the case of SMOTE/RUS, SMOTE is used on the two minority classes 1 and 2 to each reach 50% of the size of class 0, and RUS is then applied on class 0 to randomly remove half of its data points.

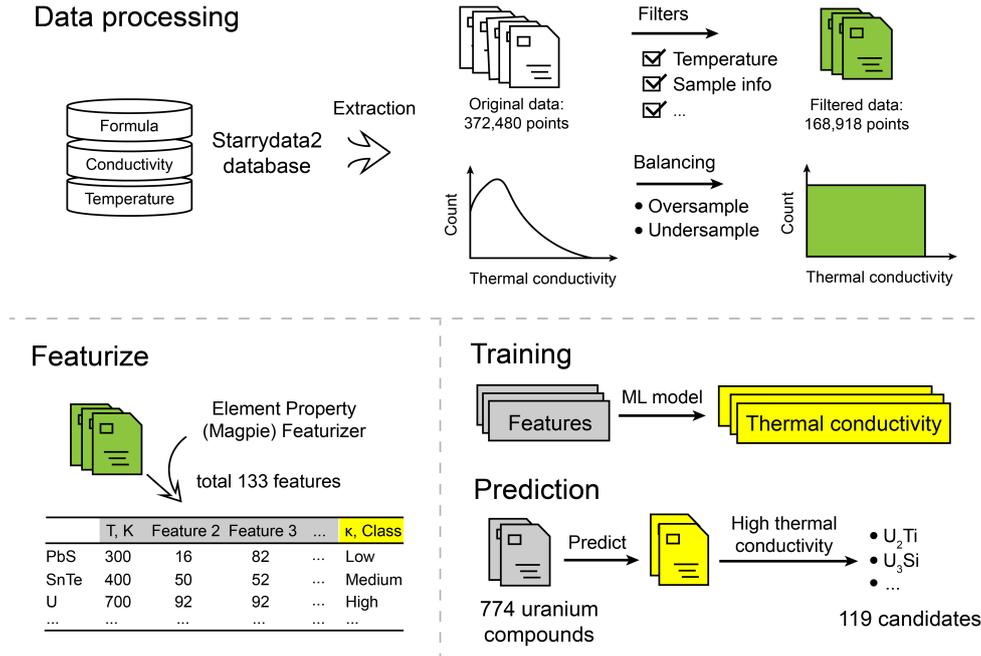


Figure 2. Schematic showing the data processing, featurizing, and model training/prediction processes.

2.3. Grouping, training, and evaluation

The Starrydata2 database, which inherently includes doped materials, is susceptible to data leakage. This is a scenario where some information present in the training data might also exist in the test data. This situation is illustrated in Figure 4, where, if no grouping is applied, both the training and test data contain information on uranium oxides. This arrangement enables a model to predict the properties of $\text{U}_{0.8}\text{Zr}_{0.2}\text{O}_2$ accurately, as it is trained with similar compounds. However, such models, trained with this form of data leakage, could underperform significantly when applied to truly unseen data.

To mitigate data leakage, we grouped the data based on similar sample compositions. This was achieved by first identifying the main elements (one or two elements with the highest molar percentages) in the sample compositions. In cases where multiple elements have the same molar percentage, the selection is done in alphabetical order. As provided in the examples in Figure 4, the main elements of UO_2 , U, U_3O_8 , $\text{U}_{0.8}\text{Zr}_{0.2}\text{O}_2$, and FeCrO_3 are (U, O), (U), (U, O), (U, O), and (Cr, O), respectively. Consequently, samples with the same main elements are placed in the same group. In this example, three groups with distinct main elements are formed. During machine learning, data within a specific group is only present either in the training or validation sets. This strategy effectively minimizes data leakage that would be present prior to grouping. After applying the grouping strategy, uranium oxides, the (U, O) group, only appear in the training data, as shown in Figure 4.

As for the machine learning algorithm, we employed a random forest classifier with cross-validation executed using the StratifiedGroupKFold technique, provided by scikit-learn[43]. Data balancing with SMOTE and RUS was applied to the training data within each cross-validation fold, leaving the validation data untouched. After 10 cross-validation folds, the performance of the models was evaluated per class based on

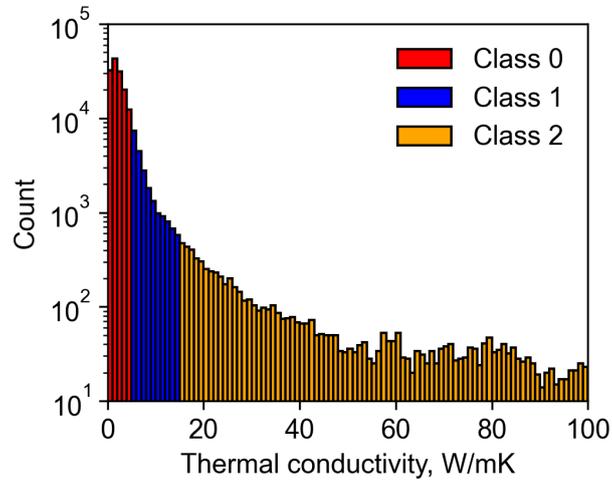


Figure 3. Semilog plot showing the data number within each class. Class 0: 0–5 W/mK, Class 1: 5–15 W/mK, Class 2: 15 + W/mK. The x-axis limit is set at 100 W/mK.

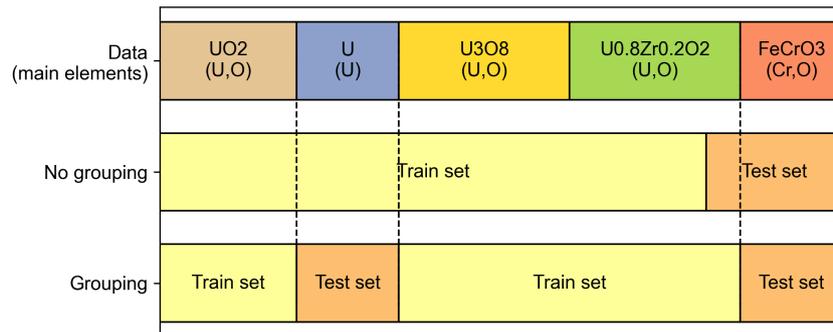


Figure 4. Data grouping by the main elements of each compound reduces data leakage during training.

the aggregated precision, recall, and F1 scores, defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

Where TP, FP, and FN represent true positives, false positives, and false negatives, respectively. Precision measures the accuracy of the model’s positive predictions, recall assesses the model’s ability to find all positive instances, and the F1 score provides a balanced evaluation. To supplement these metrics, we employed the Matthew’s correlation coefficient (MCC) to provide a comprehensive evaluation of each model’s performance. Unlike the class-specific metrics such as precision, recall, and F1 score, MCC incorporates all elements of the confusion matrix to determine the quality of the classification model. For multiclass classification, MCC can be calculated using the equation provided by Gorodkin[44]:

$$MCC = \frac{c \times s - \sum_k^K p_k \times t_k}{\sqrt{(s^2 - \sum_k^K p_k^2) \times (s^2 - \sum_k^K t_k^2)}} \quad (4)$$

In this equation, c represents the number of correctly predicted samples, s is the total number of samples, p_k indicates the total number of times class k is predicted, and t_k is the number of samples actually in class k .

In this study, the goal in data balancing is to maximize the trained model’s recall while maintaining a relatively constant F1 score, especially for class 2, which correlates with the model’s proficiency at identifying all uranium compounds with potential high thermal conductivity. This emphasis arises from our goal to discover as many such compounds as possible for further experimental verification. Maximizing recall ensures broad discovery, albeit at the risk of including some false positives. This approach is sensible in scenarios like ours, where experimental verification is feasible, and its difficulty is much less than the initial discovery. Any false positives can be eliminated in the subsequent experimental verification phase. Prioritizing precision, on the other hand, would result in fewer false positives but might lead to overlooked potential positive cases that cannot be recovered later. Therefore, it is more beneficial to cast a wider net in the prediction phase to identify as many potential candidates as possible. On the other hand, less emphasis is placed on improving the recall of class 0 (0–5 W/mK) as it contains materials with thermal conductivity on par with that of UO_2 , which are not the primary focus of this study. Finally, it is crucial that enhancing the recall and F1 score for class 2 does not adversely affect the model’s overall performance, as assessed by the MCC value. This is important because a model that excels in predicting one specific class but performs poorly overall could indicate overfitting and may be unreliable for predicting unseen data.

Three final random forest classification models are trained using the three different data sets illustrated in Table 1. The reason why data on uranium compounds from the Starrydata2 database are not removed from the final training set is that our goal

Table 2. Aggregated (KFold = 10) classification reports of the random forest models trained with various datasets. The highest value of each metrics is highlighted in green.

	Class	Precision	Recall	F1 score
Imbalanced	0	0.898	0.935	0.916
	1	0.513	0.333	0.376
	2	0.746	0.642	0.647
SMOTE	0	0.924	0.889	0.905
	1	0.459	0.489	0.445
	2	0.699	0.709	0.680
SMOTE/RUS	0	0.920	0.897	0.907
	1	0.459	0.457	0.434
	2	0.660	0.684	0.649

Table 3. Aggregated (KFold = 10) Matthew’s correlation coefficients (MCCs) of the random forest models trained with various datasets.

	Imbalanced	SMOTE	SMOTE/RUS
MCC	0.45	0.48	0.47

is to search for uranium compounds with excellent thermal conductivity with as much knowledge as possible. However, the uranium compounds present in the training set are not included in the final predictions when searching for uranium compounds with high thermal conductivity.

3. Results

3.1. Effect of balancing and size on the model’s performance

The aggregated classification reports, which detail the precision, recall, and F1 score of random forest classification models trained with three different data balances and sizes, are presented in Table 2. The MCC values for the three models are provided in Table 3. The tabulated values in both tables are the averages over 10 cross validation folds.

A comparison of the performances of the three models reveals the typical trade-off between precision and recall. Specifically, as a model becomes more precise with its predictions, it is also less likely to discover all of the positive cases. For the baseline model, trained on the original imbalanced dataset from the Starrydata2 database, the recall for classes 1 and 2 are 0.333 and 0.642, respectively. These values improved significantly to 0.489 and 0.709 after the application of SMOTE to the training data for data balancing, as illustrated in Table 2. Furthermore, the model’s MCC value in Table 3 also showed an improvement from 0.45 to 0.48 after the application of SMOTE. When the size of the training data was reduced with random undersampling in the SMOTE/RUS dataset, the model’s performance metrics either remained roughly constant or declined. This trend highlights the importance of data size, even that of the synthesized oversampling data, in improving the trained model’s performance. Overall, compared to the baseline model trained on imbalanced data, the application of SMOTE led to significant improvements in the recall and F1 score for class 2, along with a higher MCC value. This validates the use of SMOTE in data balancing as it led to the best-performing model for predicting the thermal conductivity of uranium

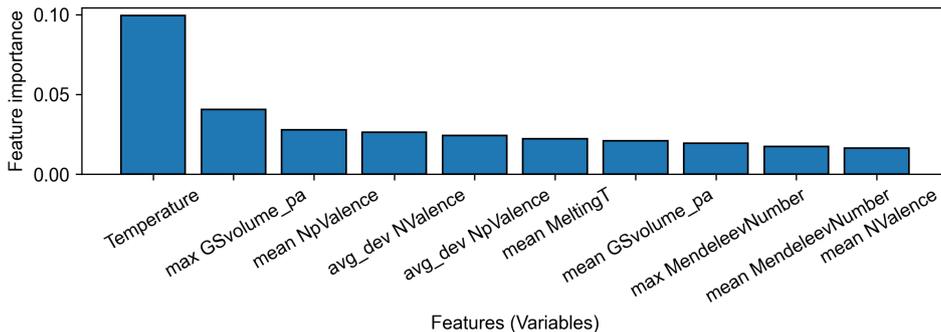


Figure 5. Top 10 features of importance in the classification model trained using the SMOTE-balanced dataset.

compounds.

3.2. Feature importance

While the complexity and non-linear nature of random forest models make fully unraveling their inner workings a challenging task, understanding feature importance provides crucial insights into the model’s decision-making process. This evaluation allows us to confirm whether the model’s prediction mechanism aligns with established principles in materials science. Figure 5 displays the top 10 most important features of the classification model, trained using the SMOTE-balanced dataset. These features are derived from temperature and element properties of the compound’s constituent atoms. These properties include ground state volume (GSvolume), number of valence electrons in the p-orbital (NpValence), number of valence electrons (NValence), melting point, and the Mendeleev numbers. In the Magpie descriptors, valence electrons are defined as those beyond the nearest noble gas configuration. For example, the NValence values for Fe, Te, and Au are 8, 16 and 25, respectively.

Temperature, as expected, emerges as the most influential feature. It governs phonon-phonon and phonon-electron scattering processes, thereby impacting a compound’s lattice and electronic thermal conductivity. GSvolume, which represents the volume per atom at its ground state (0K) as determined by Density-functional theory (DFT) calculations, offers insight into the compound’s lattice volume. This is an important factor in the phonon dispersion relations and, subsequently, the lattice thermal conductivity of the compound.

We believe that the mean and average deviation of NpValence are used by the model to distinguish between compounds with and without metallic bonds. Metallic compounds, with a significant number of free electrons participating in heat transport, exhibit superior electronic thermal conductivity. Most metal elements, with their valence electrons located in the s and d orbitals, do not have valence electrons in the p orbitals (NpValence = 0). Some minor exceptions include post-transition metals such as aluminum. Consequently, both the mean and average deviation of NpValence are generally zero for metal alloys. In addition, the mean melting point of the constituent atoms can also provide insights into the nature of the compound’s bonds. Ionic compounds, usually containing chalcogen and halogen elements, tend to have lower mean melting points compared to metal alloys. While it’s not immediately obvious how the model uses the NValence features, we speculate that they also help differentiate among

Table 4. Top 50 out of 119 predicted uranium compounds with thermal conductivity higher than 15 W/mK between 300 and 1000 K, ranked in descending uranium density order.

Rank	Compound	Rank	Compound	Rank	Compound	Rank	Compound	Rank	Compound
1	U ₃ Si	11	U ₂ RuC ₂	21	U ₂ PtC ₂	31	UFeC ₂	41	U ₃ Co ₇ B ₂
2	U ₂ Ti	12	U ₃ (SiC) ₂	22	UCo ₂	32	UCrC ₂	42	U ₂ B ₆ Mo
3	U ₂ CN	13	U ₂ OsC ₂	23	U ₂ PN ₂	33	UVN ₂	43	U ₂ B ₆ W
4	UCo	14	U ₂ NiC ₃	24	UIr	34	U ₂ SbN ₂	44	UTcC ₂
5	U ₂ C ₃	15	U ₄ N ₇	25	U ₅ Re ₃ C ₈	35	U ₂ Mn ₃ Si	45	UReC ₂
6	UBN	16	UC ₂	26	U ₂ Co ₃ Si	36	UMnFe	46	UMoC ₂
7	UBC	17	U ₂ RhC ₂	27	UMn ₂	37	UMnC ₂	47	UWC ₂
8	UN ₂	18	U ₂ IrC ₂	28	U ₂ AsN ₂	38	UVC ₂	48	USiRh
9	U ₁₁ Ni ₁₆	19	U ₂ MnN ₃	29	UCoC ₂	39	U ₂ ReB ₆	49	UTaN ₂
10	UH ₃	20	U ₂ CrN ₃	30	UFe ₂	40	U ₂ Re ₂ C ₃	50	UTaC ₂

various types of compounds.

Lastly, the Mendeleev numbers, while not being thermophysical parameters, correspond to an element’s position on the periodic table, indirectly reflecting its physical and chemical properties. In this context, the maximum Mendeleev number could be leveraged by the model to differentiate between various types of materials, such as metallics and oxides. The mean Mendeleev number may indicate the dominant elements in the material, enabling the model to infer combinations of elements with characteristic thermal conductivities.

In conclusion, the top ten features employed in the random forest classification model are variables that are closely connected to a material’s thermal conductivity. This validates the model’s decision-making process and its alignment with established materials science principles, thereby ensuring robust and accurate predictions of the thermal conductivity of unseen uranium compounds.

3.3. Predicted uranium compounds with high thermal conductivity

A list of 783 stable uranium compounds for prediction is extracted from the Materials Project database, and 9 duplicate formulas that are already present in the training data are removed to prevent data leakage. The classification model trained with a SMOTE-balanced dataset is used for the prediction, and only the 146 stable uranium compounds predicted to belong in class 2 (15+ W/mK) at some point between 300 and 1000 K are further processed. To narrow down the search for potential fuel candidates, we remove compounds containing non-uranium metal elements with melting temperature lower than 1500 K. After sorting by the compound’s uranium density in a descending order, the top 50 fuel candidates are listed in Table 4, and a full list of the 119 predicted potential fuel candidates is provided in Table A1 in Appendix A.

4. Discussion

4.1. Evaluation of model performance and prediction trends

The recall, F1 score, and MCC value of the models, as shown previously in Tables 2 and 3, all confirm improved model performance after data balancing. However, these metrics do not fully illustrate the shifts in the model’s prediction trend. The confusion matrices of both models, displayed in Figure 6, indicate a decrease in the model’s tendency to underestimate thermal conductivity following data balancing. After data balancing with SMOTE, the probability of the model incorrectly classifying materials

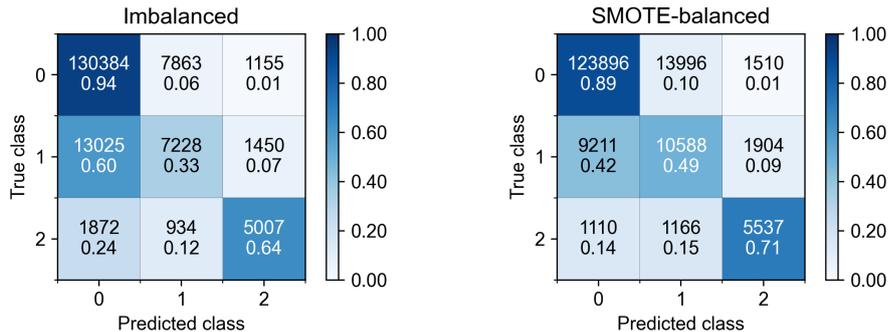


Figure 6. Confusion matrices of the models trained with the imbalanced dataset (left) and the SMOTE-balanced dataset (right).

from classes 1 and 2 as class 0 decreased from 60% to 42% and from 24% to 14%, respectively. Moreover, the model trained with a SMOTE-balanced dataset now shows a higher tendency to assign larger thermal conductivity values, which contributes to the increased recall of classes 1 and 2. As demonstrated in Figure 6, the number of data points predicted to be in classes 1 and 2 increased from 16025 and 7612 to 25750 and 8951, respectively. This increase led to the correct identification of an additional 3360 and 530 data points in classes 1 and 2, respectively.

Both the recall values and the confusion metrics point to the same conclusion: with the proper class balancing, we are able to address the bias towards predicting lower thermal conductivity, which was prevalent with data sourced from the Starrydata2 database. While the field of nuclear materials lacks a comprehensive thermophysical properties database, our study illustrates an effective method for transforming non-nuclear focused databases for machine learning applications. However, there is significant potential for enhancing the performance of the proposed classification model, especially considering that the recall for class 2 (15+ W/mK) is currently only at 0.71. The model presently incorporates temperature and various elemental properties as features, yet thermal conductivity is intrinsically related to the crystal and electronic structure of a material. Therefore, introducing crystal structure features such as lattice parameters, angles, and site numbers, or electronic information such as band information and density of states could potentially improve the model’s recall. Tools like the Materials Project API allow for the extraction of these crystal structure and electronic structure features for compounds. However, the prevalence of doped materials in the Starrydata2 database imposes substantial challenges to the incorporation of these features in this work. Moreover, determining how to effectively featurize the electronic structure information is another crucial issue.

4.2. Analysis of predicted uranium compounds

The thermal conductivity of most materials presented in Table 4 has not been previously reported. Among the materials with known high thermal conductivity, the model accurately identified U_3Si [45], U_2RuC_2 [46], and U_2RhC_2 [46]. However, it overestimated the thermal conductivity of UC_2 [47], U_2C_3 [48], and $UMoC_2$ [46], even though their thermal conductivity ultimately surpasses 15 W/mK at temperatures higher than 1000 K.

Among the 119 selected candidates, only UN_2 has a calculated non-zero band

gap[49]. This suggests that these uranium compounds display metallic-like electrical conductivity where delocalized free electrons can contribute to a high electronic thermal conductivity. For example, the calculated electronic thermal conductivity of UMn_2 and UFe_2 from their reported electrical resistivity[50] both exceed 15 W/mK below 1000 K. For U_2PtC_2 [51], while its electrical resistivity is reported, it is only reported for room temperature, providing insufficient information to estimate its electronic thermal conductivity at higher temperatures.

Overall, the concurrence between the predicted thermal conductivity and existing literature underscores the predictive capability of the classification model trained with the SMOTE-balanced dataset. However, it should be noted that, despite these promising candidates featuring an energy above hull of zero, it does not necessarily ensure the synthesizability of single-phase materials or their stability at high temperatures, warranting further experimental validation. For instance, the fabrication of U_2C_3 results in UC_{2-x} impurities[48,52], and UN_2 has been shown to convert to U_2N_3 and UN at temperatures higher than 948 K under an inert argon atmosphere[53].

5. Conclusion

To accelerate the discovery of advanced nuclear fuels with high thermal conductivity, we introduced a multiclass random forest classification model to predict the thermal conductivity of uranium compounds. The model, trained on a dataset of 168,918 thermal conductivity records from the Starrydata2 database and Magpie + temperature features, successfully improved the recall from 0.33 and 0.64 to 0.49 and 0.71 for classes 1 (5–15 W/mK) and 2 (15+ W/mK), respectively, after applying SMOTE for data balancing. The results offer valuable insights for the application of machine learning in nuclear materials research, particularly where comprehensive nuclear-specific databases are lacking.

Our classification model predicted that 146 out of 774 stable uranium compounds would exhibit a thermal conductivity of 15 W/mK or higher within the temperature range of 300 to 1000 K. From this subset, we further narrowed down to 119 potential advanced fuel candidates by filtering based on the elements' melting points and arranging them by their uranium density. Our classification model, which can be trained in just a few minutes, offers a far more efficient and feasible technique for searching advanced nuclear fuels among the unexplored uranium compounds, compared to the traditional trial-and-error approach. While our ML-assisted approach still necessitates follow-up experiments to verify the thermal conductivity, high-temperature stability, and synthesizability of the selected candidates, it significantly reduces the number of samples requiring fabrication and measurements.

Finally, it should be emphasized that the development of advanced nuclear fuels is influenced by a multitude of factors, ranging from manufacturing and operation to re-processing. While our model is specifically focused on thermal conductivity predictions, it serves as a significant stepping stone for the development of more comprehensive ML frameworks. These advanced models could enable a wider range of thermophysical and mechanical properties predictions and significantly accelerate the discovery of well-rounded advanced nuclear fuels.

Disclosure statement

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

Funding

This work was supported by MEXT Innovative Nuclear Research and Development Program Grant Number JPMXD0220354330 and JPMXD0222682541.

Data availability statement

The data that support the findings of this study are openly available at <https://github.com/AzarashiYifan/classification-uranium-thermal-conductivity>.

References

- [1] Terrani KA. Accident tolerant fuel cladding development: Promise, status, and challenges. *J Nucl Mater.* 2018;501:13–30.
- [2] Kurata M. Research and development methodology for practical use of accident tolerant fuel in light water reactors. *Nucl Eng Technol.* 2016;48(1):26–32.
- [3] Zinkle SJ, Terrani KA, Gehin JC, et al. Accident tolerant fuels for LWRs: A perspective. *J Nucl Mater.* 2014;448(1-3):374–379.
- [4] Pasamehmetoglu K, Massara S, Costa D, et al. State-of-the-art report on light water reactor accident-tolerant fuels. Boulogne-Billancourt, France: Nuclear Energy Agency; 2018.
- [5] Yun D, Lu C, Zhou Z, et al. Current state and prospect on the development of advanced nuclear fuel system materials: A review. *Mater Today Energy.* 2021;1(1). 100007.
- [6] Vasudevamurthy G, Nelson AT. Uranium carbide properties for advanced fuel modeling—a review. *J Nucl Mater.* 2022;558. 153145.
- [7] Ross SB, El-Genk MS, Matthews RB. Thermal conductivity correlation for uranium nitride fuel between 10 and 1923 K. *J Nucl Mater.* 1988;151(3):318–326.
- [8] White JT, Nelson AT, Dunwoody JT, et al. Thermophysical properties of U_3Si_2 to 1773 K. *J Nucl Mater.* 2015;464:275–280.
- [9] Watkins JK, Gonzales A, Wagner AR, et al. Challenges and opportunities to alloyed and composite fuel architectures to mitigate high uranium density fuel oxidation: Uranium mononitride. *J Nucl Mater.* 2021;553. 153048.
- [10] Watkins JK, Wagner AR, Gonzales A, et al. Challenges and opportunities to alloyed and composite fuel architectures to mitigate high uranium density fuel oxidation: Uranium diboride and uranium carbide. *J Nucl Mater.* 2022;560. 153502.
- [11] Jolkkonen M, Malkki P, Johnson K, et al. Uranium nitride fuels in superheated steam. *J Nucl Sci Technol.* 2017;54(5):513–519.
- [12] Sugihara S, Imoto S. Hydrolysis of uranium nitrides. *J Nucl Sci Technol.* 1969;6(5):237–242.
- [13] Bradley MJ, Ferris LM. Hydrolysis of uranium carbides between 25 and 100°. I. uranium monocarbide. *Inorg Chem.* 1962;1(3):683–687.
- [14] Hori Y, Mukaibo T. Kinetic study of the reaction between uranium monocarbide and water vapor. *Bull Chem Soc Jpn.* 1967;40(8):1878–1883.
- [15] Migdisov A, Nisbet H, Li N, et al. Instability of U_3Si_2 in pressurized water media at elevated temperatures. *Commun Chem.* 2021;4. 65.

- [16] Sweet RT, Yang Y, Terrani KA, et al. Performance of U_3Si_2 in an LWR following a cladding breach during normal operation. *J Nucl Mater.* 2020;539. 152263.
- [17] Nelson AT, Migdisov A, Wood ES, et al. U_3Si_2 behavior in H_2O environments: Part II, pressurized water with controlled redox chemistry. *J Nucl Mater.* 2018;500:81–91.
- [18] Wood ES, White JT, Grote CJ, et al. U_3Si_2 behavior in H_2O : Part I, flowing steam and the effect of hydrogen. *J Nucl Mater.* 2018;501:404–412.
- [19] Jain A, Ong SP, Hautier G, et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* 2013;1(1). 011002.
- [20] Liu Y, Zhao T, Ju W, et al. Materials discovery and design using machine learning. *J Materiomics.* 2017;3(3):159–177.
- [21] Butler KT, Davies DW, Cartwright H, et al. Machine learning for molecular and materials science. *Nature.* 2018;559(7715):547–555.
- [22] Schmidt J, Marques MR, Botti S, et al. Recent advances and applications of machine learning in solid-state materials science. *npj Comput Mater.* 2019;5. 83.
- [23] Behara S, Poonawala T, Thomas T. Crystal structure classification in ABO_3 perovskites via machine learning. *Comput Mater Sci.* 2021;188. 110191.
- [24] Kusaba M, Liu C, Yoshida R. Crystal structure prediction with machine learning-based element substitution. *Comput Mater Sci.* 2022;211. 111496.
- [25] Oliynyk AO, Adutwum LA, Rudyk BW, et al. Disentangling structural confusion through machine learning: structure prediction and polymorphism of equiatomic ternary phases ABC. *J Am Chem Soc.* 2017;139(49):17870–17881.
- [26] Park WB, Chung J, Jung J, et al. Classification of crystal structure using a convolutional neural network. *IUCrJ.* 2017;4(4):486–494.
- [27] De Jong M, Chen W, Notestine R, et al. A statistical learning framework for materials science: application to elastic moduli of k-nary inorganic polycrystalline compounds. *Sci Rep.* 2016;6. 34256.
- [28] Evans JD, Coudert FX. Predicting the mechanical properties of zeolite frameworks by machine learning. *Chem Mater.* 2017;29(18):7833–7839.
- [29] Furmanchuk A, Agrawal A, Choudhary A. Predictive analytics for crystalline materials: bulk modulus. *RSC Adv.* 2016;6(97):95246–95251.
- [30] Xie T, Grossman JC. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys Rev Lett.* 2018;120(14). 145301.
- [31] Chen L, Tran H, Batra R, et al. Machine learning models for the lattice thermal conductivity prediction of inorganic materials. *Comput Mater Sci.* 2019;170. 109155.
- [32] Pal K, Park CW, Xia Y, et al. Scale-invariant machine-learning model accelerates the discovery of quaternary chalcogenides with ultralow lattice thermal conductivity. *npj Comput Mater.* 2022;8. 48.
- [33] Jaafreh R, Kang YS, Hamad K. Lattice thermal conductivity: an accelerated discovery guided by machine learning. *ACS Appl Mater Interfaces.* 2021;13(48):57204–57213.
- [34] Miyazaki H, Tamura T, Mikami M, et al. Machine learning based prediction of lattice thermal conductivity for half-Heusler compounds using atomic information. *Sci Rep.* 2021; 11. 13410.
- [35] Kautz EJ, Hagen AR, Johns JM, et al. A machine learning approach to thermal conductivity modeling: A case study on irradiated uranium-molybdenum nuclear fuels. *Comput Mater Sci.* 2019;161:107–118.
- [36] Lu Y, Huang X, Ren Z, et al. A prediction model for thermal conductivity of metallic nuclear fuel based on multiple machine learning models. *J Nucl Mater.* 2023;154553.
- [37] 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. 16028.
- [38] Katsura Y, Kumagai M, Kodani T, et al. Data-driven analysis of electron relaxation times in PbTe-type thermoelectric materials. *Sci Technol Adv Mater.* 2019;20(1):511–520.
- [39] Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* 2002;16:321–357.
- [40] Brandes EA, Brook G. *Smithells metals reference book.* 7th ed. Oxford (MA):

- Butterworth-Heinemann; 1992.
- [41] Ward L, Dunn A, Faghaninia A, et al. Matminer: An open source toolkit for materials data mining. *Comput Mater Sci.* 2018;152:60–69.
 - [42] Gibby R. The effect of plutonium content on the thermal conductivity of (U,Pu)O₂ solid solutions. *J Nucl Mater.* 1971;38(2):163–177.
 - [43] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res.* 2011;12:2825–2830.
 - [44] Gorodkin J. Comparing two k-category assignments by a k-category correlation coefficient. *Comput Biol Chem.* 2004;28(5-6):367–374.
 - [45] White J, Nelson A, Byler D, et al. Thermophysical properties of U₃Si to 1150 K. *J Nucl Mater.* 2014;452(1-3):304–310.
 - [46] Arai Y, Ohmichi T, Fukushima S, et al. Thermal conductivity of UMoC₂, UMoC_{1.7}, U₂RuC₂ and U₂RhC₂. *J Nucl Mater.* 1985;132(3):284–287.
 - [47] Lewis, HD and Kerrisk, JF. Electrical and thermal transport properties of uranium and plutonium carbides. Los Alamos (NM): Los Alamos National Lab.(LANL) (US); 1976.
 - [48] De Coninck R, Van Lierde W, Gijs A. Thermal diffusivity and conductivity of U₂C₃ up to 2200 K. *J Nucl Mater.* 1973;46(2):213–216.
 - [49] Jain A, Hautier G, Moore CJ, et al. A high-throughput infrastructure for density functional theory calculations. *Comput Mater Sci.* 2011;50(8):2295–2310.
 - [50] Hamaguchi Y, Kunitomi N, Komura S, et al. Electrical resistivity of UEe₂ and UMn₂. *J Phys Soc Jpn.* 1962;17(2):398–398.
 - [51] Meisner G, Giorgi A, Lawson A, et al. U₂PtC₂ and systematics of heavy fermions. *Phys Rev Lett.* 1984;53(19):1829–1832.
 - [52] Nuñez UC, Eloirdi R, Prieur D, et al. Structure of UC₂ and U₂C₃: XRD, ¹³C NMR and EXAFS study. *J Alloys Compd.* 2014;589:234–239.
 - [53] Silva GC, Yeaman CB, Sattelberger AP, et al. Reaction sequence and kinetics of uranium nitride decomposition. *Inorg Chem.* 2009;48(22):10635–10642.

Appendix A. List of predicted advanced fuel candidates with high thermal conductivity.

Table A1. 119 predicted uranium compounds with thermal conductivity higher than 15 W/mK between 300 and 1000 K, ranked in descending uranium density order.

Rank	Compound	Rank	Compound	Rank	Compound	Rank	Compound	Rank	Compound
1	U ₃ Si	25	U ₅ Re ₃ C ₈	49	UTaN ₂	73	UFe ₃ B ₂	97	U ₅ (B ₁₂ Mo ₅) ₂
2	U ₂ Ti	26	U ₂ Co ₃ Si	50	UTaC ₂	74	URu ₃	98	U(SiRh) ₂
3	U ₂ CN	27	UMn ₂	51	UNbN ₂	75	URu ₂ Rh	99	U(SiRu) ₂
4	UCo	28	U ₂ AsN ₂	52	USiRu	76	UB ₄ W	100	USi ₂ RuRh
5	U ₂ C ₃	29	UCoC ₂	53	UNbC ₂	77	URh ₃	101	U(SiPd) ₂
6	UBN	30	UFe ₂	54	HfUN ₂	78	UCo ₃ B ₂	102	U(SiPt) ₂
7	UBC	31	UFeC ₂	55	URE ₂	79	UNi ₂ B ₂ C	103	U(CrC) ₄
8	UN ₂	32	UCrC ₂	56	ZrUN ₂	80	USi ₃	104	U ₆ P ₁₃ Rh ₂₀
9	U ₁₁ Ni ₁₆	33	UVN ₂	57	UOs ₂	81	UFe ₂ SiC	105	UB ₁₂
10	UH ₃	34	U ₂ SbN ₂	58	U ₆ Co ₁₂ Ge ₄ C	82	U ₆ Fe ₁₆ Si ₇ C	106	UPt ₅
11	U ₂ RuC ₂	35	U ₂ Mn ₃ Si	59	HfUC ₂	83	U ₃ Fe ₂ Si ₇	107	U(BRu) ₄
12	U ₃ (SiC) ₂	36	UMnFe	60	ZrUC ₂	84	UIr ₃	108	UPdPt ₄
13	U ₂ OsC ₂	37	UMnC ₂	61	UFeB ₄	85	UMn ₂ SiC	109	U(Mn ₂ P) ₂
14	U ₂ NiC ₃	38	UVC ₂	62	UCoB ₄	86	U ₂ (PdRh) ₃	110	UMn ₅ P ₃
15	U ₄ N ₇	39	U ₂ ReB ₆	63	UREB ₃	87	ThUC ₂	111	UCr ₅ P ₃
16	UC ₂	40	U ₂ Re ₂ C ₃	64	UGeRh	88	U ₄ Si ₆ Tc ₇	112	UV ₅ P ₃
17	U ₂ RhC ₂	41	U ₃ Co ₇ B ₂	65	UMnB ₄	89	U ₄ Tc ₇ Ge ₆	113	U ₄ Be ₅₁ B
18	U ₂ IrC ₂	42	U ₂ B ₆ Mo	66	UCrB ₄	90	UNi ₅	114	UBe ₁₃
19	U ₂ MnN ₃	43	U ₂ B ₆ W	67	UGePd	91	U ₂ Mn ₃ Si ₅	115	U(Cr ₃ P ₂) ₂
20	U ₂ CrN ₃	44	UTcC ₂	68	UNiB ₄	92	UB ₂ Ru ₃	116	U ₂ (Co ₇ B ₂) ₃
21	U ₂ PtC ₂	45	UREC ₂	69	GdUN ₂	93	U(FeB ₃) ₂	117	U ₂ Si ₇ Ru ₁₂
22	UCo ₂	46	UMoC ₂	70	UB ₄ Ru	94	UB ₂ Os ₃	118	U ₂ (Ni ₇ B ₂) ₃
23	U ₂ PN ₂	47	UWC ₂	71	UREB ₄	95	U ₂ Nb ₃ Si ₄	119	YUCo ₁₀
24	UIr	48	USiRh	72	UB ₄ Os	96	UB ₂ Ir ₃		