

Article

Prediction of Thermal and Optical Properties of Oxyfluoride Glasses Based on Interpretable Machine Learning

Yuhao Xie ² and Xiangfu Wang ^{1,3,*} 

¹ College of Electronic and Optical Engineering & College of Flexible Electronics (Future Technology), Nanjing University of Posts and Telecommunications, Nanjing 210023, China

² College of Integrated Circuit Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

³ Yunnan Key Laboratory of Electromagnetic Materials and Devices, Kunming 650091, China

* Correspondence: xfwang@njupt.edu.cn

Abstract: Based on the components of glasses, four algorithms, namely K-Nearest Neighbor, Random Forest, Support Vector Machine, and eXtreme Gradient Boosting, were used to construct an optimal machine learning model to predict the thermal and optical properties of oxyfluoride glass, namely glass transition temperature, density, Abbe number, liquidus temperature, thermal expansion coefficient, and refractive index. We perform SHAP analysis on the constructed machine learning model to explain the effects of different components on the properties. Based on the trained machine learning models, we developed several ternary system prediction maps that can effectively predict the properties of glasses composed of different proportions of components. This study provides a method to design new oxyfluoride glasses only knowing the components of glasses, which is instructive for the development of new types of oxyfluoride glasses as well as for computer-aided reverse design.

Keywords: oxyfluoride glass; machine learning; property prediction; model interpretation



Academic Editor: Francisco Torrens

Received: 13 May 2025

Revised: 30 May 2025

Accepted: 1 June 2025

Published: 3 June 2025

Citation: Xie, Y.; Wang, X. Prediction of Thermal and Optical Properties of Oxyfluoride Glasses Based on Interpretable Machine Learning. *Nanomaterials* **2025**, *15*, 860. <https://doi.org/10.3390/nano15110860>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As a unique glass material system, oxyfluoride glass is a composite of oxides and fluorides. Compared with traditional oxide glasses, oxyfluoride glasses have unique properties [1,2]. In terms of optics, some oxyfluoride glasses are suitable for working in the ultraviolet spectral region, which can be applied as a laser material; some oxyfluoride glasses enriched with a high concentration of rare-earth ions exhibit high refractive indexes, which is of great significance in the design of optical components. In addition, some oxyfluoride glasses containing rare earth elements have not only unique luminescence properties but also high thermal stability and strong paramagnetism, making them ideal for new optical and magnetic materials.

However, compared with oxide glass, there is an obvious deficiency in the current exploration of the relationship between the composition and properties of oxyfluoride glass. The traditional Edison trial-and-error method requires a large number of experimental attempts, which not only consumes a lot of time but also consumes numerous resources and largely restricts the progress of research on oxyfluoride glass [3]. In recent years, with the development of artificial intelligence technology, researchers have actively introduced machine learning (ML) algorithms, which are dedicated to exploring the intrinsic relationship between composition and properties in the field of glass [4]. In related studies,

Ravinder, R., Zaki, M., Cassar, D. R., and others have conducted in-depth studies on various properties of oxide glasses using machine learning algorithms [5–22]; Mastelini, S. M., Singla, S., and others have successfully elucidated the correlation between composition and properties of sulfur-based glasses through machine learning [23,24]; Shaik Kareem Ahmmad et al. also attempted to predict the density of fluoride glass with the help of machine learning, but failed to explain the constructed model and did not analyze the important role of fluoride-containing glass systems in optics and other fields [25]. Therefore, the study of composition-property prediction of oxyfluoride glasses through machine learning is of great significance which is expected to accelerate the development of the field of oxyfluoride glass.

In applying machine learning to the field of glass, researchers have explored numerous algorithms, attempted to compare multiple algorithms, and constructed models with optimal performance. Among these algorithms, K-Nearest Neighbors, Random Forest, Neural Networks, Support Vector Machines, and eXtreme Gradient Boosting are commonly used as analytical tools. In the study of oxide glasses, Ravinder, R. used Neural Network modeling to provide insight into the role of 37 oxides in the glass system and developed GSC diagrams [12]. Zaki, M. also used Neural Networks to focus on two key optical properties of glass, refractive index and Abbe number and introduced SHAP analysis to explain the effect of individual oxides on the optical properties [17]. Cassar, D. R. adopted a more comprehensive research strategy, using decision tree induction, K-Nearest Neighbor, and Random Forest algorithms to systematically predict and analyze six properties of oxide glasses [14]. In the study of sulfur glasses, Mastelini S. M., Singla, S., et al. adopted a similar idea to predict, compare, and interpret models for multiple properties of sulfur glasses [23,24]. The composition-property relationships of oxyfluoride glasses are therefore suitable for investigation using these mature machine learning algorithms, which are expected to drive further development in the field.

In this paper, we compare four machine learning algorithms—K-Nearest Neighbors, Random Forest, Support Vector Machines, and eXtreme Gradient Boosting—to construct predictive models for six key properties: glass transition temperature, density, Abbe number, liquidus temperature, thermal expansion coefficient, and refractive index, with respect to the composition-property relationships of oxyfluoride glasses. At the same time, we introduce the Shapley additive interpretation (SHAP) algorithm. Through this algorithm, we explain the effects of each compound composition on the six properties of oxyfluoride glass, thus revealing the intrinsic correlation between the compound compositions and the glass properties. In addition, based on the trained machine learning model, we carry out property prediction for the ternary oxyfluoride glass system. By simulating the prediction of ternary oxyfluoride glass systems with different scales, we develop a series of ternary property prediction diagrams. The plots visualize the properties of the glass in the ternary oxyfluoride glass system with different compositional ratios, which is helpful for the development of new oxyfluoride glasses and computer-aided reverse design.

2. Materials and Methods

2.1. Data Collection

All glass datasets in this study were constructed based on the processing of the SciGlass database [26] by the Python software package glasspy 0.5.3 [27]. Considering previous data processing methods, we collected statistical information for glasses with a non-zero fluoride percentage under six attributes: glass transition temperature, density, Abbe number, liquidus temperature, thermal expansion coefficient, and refractive index, and implemented the following data processing steps:

1. Ensure that the sum of the glass compositions for each group is exactly 1 in order to circumvent errors that may be introduced by manual preparation.
2. Eliminate redundant data by replacing duplicate entries with median values.
3. Remove extreme values, defined as data points outside the 0.05% and 99.95% percentile ranges. Previous research has shown that some of the extreme values can lead to deterioration in model performance.
4. Removing components with standard deviations less than 10^{-3} , i.e., characteristics with very low variance.
5. Apply Variance Inflation Factor (VIF) to remove features with a high degree of multicollinearity.
6. Remove glass with a low fluoride content by setting an appropriate fluoride content threshold for each attribute dataset, a step that allows the model to focus more on the fluoride component and simplifies the number of features, reducing model complexity.
7. Select only those compound components present in at least 10 glass compositions to ensure that the data in the training and test sets are representative.

It is worth noting that when dealing with the dataset of thermal expansion coefficient, we found that the difference in the order of magnitude between the lowest and the highest values of the original dataset amounted to two orders of magnitude, so we took the past literature's treatment [14] and preprocessed the dataset of this property with a logarithmic function with a base of 10. After the processing, the model performance of this property is greatly improved in three algorithms: KNN, RF, and XGBoost. After the above processing flow, the final statistical results of database information for each property are presented in Table 1.

Table 1. Descriptive statistics of the dataset used.

	T _g	Density	AbbeNumber	T _{liq}	Log ₁₀ (TEC)	RefractiveIndex
Number of oxides	29	38	27	11	39	41
Number of fluorides	43	41	19	28	29	48
count	4955	3806	640	2669	2502	5209
max	1087.15	8.09	107.6	1818.15	−4.44	2.29
min	223.15	1.79	15.53	569.15	−5.43	1.27
mean	580.46	3.95	44.35	1009.56	−4.91	1.52

2.2. Machine Learning Algorithms

In this study, we have chosen four algorithms that are commonly used in related research in this field and have excellent performance, namely the K-Nearest Neighbors algorithm (KNN) [28], the Random Forest algorithm (RF) [29], the Support Vector Machines algorithm (SVM) [30] and the eXtreme Gradient Boosting algorithm (XGBoost) [31].

The K-Nearest Neighbors algorithm is based on the principle of local approximation. For a given target sample to be predicted, the algorithm searches for the K neighboring samples that are closest to the target sample in the existing training dataset based on specific distance measures, such as Euclidean distance, Manhattan distance, and so on. Subsequently, the actual output values of these K neighboring samples are integrated by means of weighted averaging, etc., to determine the predicted value of the target sample. In weighted averaging, the closer the neighbor samples are to the target sample, the higher the weight is usually given.

The Random Forest algorithm combines the Bagging technique with Decision Trees. The bagging technique generates multiple distinct sub-datasets by performing a release sampling of the original training dataset. Each sub-dataset is used to train a decision tree

independently, and some features are randomly selected during the decision tree construction process to further increase the diversity of the model. Ultimately, the prediction result of the Random Forest is the arithmetic mean of the prediction results of all decision trees. This integration strategy gives full play to the advantages of multiple decision trees and effectively avoids the overfitting problem that is prone to occur in a single decision tree.

The core idea of the Support Vector Machines algorithm is to explore an optimal regression hyperplane in the feature space such that the sum of distances from all sample points to this hyperplane is minimized while allowing for some degree of error. Different from traditional regression methods, SVR introduces the ε -insensitive loss function. That is, the loss is not counted for sample points within the “insensitive band” of width ε centered on the regression hyperplane; the loss is counted only when the sample points fall outside the “insensitive band”. This feature makes SVR more robust to noise and outliers.

The eXtreme Gradient Boosting algorithm builds powerful regression models by iteratively training a series of decision trees. Each newly generated decision tree is trained based on the prediction errors of all previous trees, with the goal of gradually reducing the overall prediction error. Each decision tree in XGBoost is a regression tree with leaf nodes storing predicted values. During the decision tree construction process, XGBoost uses a greedy algorithm to select features and split points that minimize the loss function each time. By iteratively fitting the gradient and introducing regularization terms, XGBoost is able to efficiently capture complex patterns in the data and has a good tolerance for noise and outliers in the data due to its integrated nature based on multiple decision trees, where the error of a single decision tree has less impact on the overall result.

In our work, we divided the training set and test set in the ratio of 80:20 for training the four algorithm models. Considering that hyperparameters have a significant impact on the performance of the prediction models, we used the grid search method and combined it with a 5-fold cross-validation strategy to optimize the hyperparameters of the four algorithms mentioned above using the R^2 scores as an evaluation metric [32,33]. In Tables 2 and 3, we recorded the hyperparameter optimization results of the four algorithms for each attribute. After completing the model optimization, we compared the performance of the four algorithms and constructed the optimal prediction model, on the basis of which we carried out the subsequent SHAP analysis as well as the ternary system prediction.

Table 2. Hyperparameters of the algorithms used for glass transition temperature, density, and Abbe number.

Model	Hyperparameter	T _g	Density	AbbeNumber
KNN	n_neighbors	3	3	3
	p	1	1	1
	weights	distance	distance	distance
RF	max_depth	none	none	none
	min_samples_leaf	1	1	1
	min_samples_split	2	2	4
	n_estimators	200	400	400
SVM	C	100	100	100
	gamma	scale	1	scale
	kernel	poly	rbf	rbf
XGBoost	colsample_bytree	0.9	0.8	0.9
	learning_rate	0.1	0.1	0.05
	max_depth	9	7	11
	n_estimators	300	300	300
	subsample	0.9	0.8	0.8

Table 3. Hyperparameters of the algorithm used for liquidus temperature, thermal expansion coefficient, and refractive index.

Model	Hyperparameter	T _{liq}	Log ₁₀ (TEC)	RefractiveIndex
KNN	n_neighbors	3	3	3
	p	1	1	1
	weights	distance	distance	distance
RF	max_depth	none	none	none
	min_samples_leaf	1	1	1
	min_samples_split	2	2	2
	n_estimators	300	400	300
SVM	C	100	1	100
	gamma	scale	scale	0.1
	kernel	rbf	rbf	rbf
XGBoost	colsample_bytree	0.9	0.8	1
	learning_rate	0.2	0.05	0.2
	max_depth	11	11	7
	n_estimators	300	300	300
	subsample	0.8	0.9	0.9

The algorithms used in this study are provided by Python packages, including NumPy [34], pandas [35], matplotlib [36], Scikit-learn [37], and xgboost. The packages NumPy and pandas are used to process the dataset; matplotlib is used for data visualization; and Scikit-learn and xgboost provide the code for machine learning.

2.3. SHAP Analysis

In the field of machine learning, model interpretability is crucial. SHAP, a powerful model interpretation tool whose theory is based on the Shapley value in cooperative game theory, is now widely used in many fields [38–40]. In this study, we computed the Shapley value of each glass component with the help of Python’s shap module in order to interpret the six constructed models.

Due to the additivity property of SHAP values, the predicted values of the models can be obtained by adding the SHAP values of all the feature components in a given prediction to a base value, which is usually the mean of the target value. This additivity property provides a clear picture of how much each feature component contributes to the final property prediction and which compounds contribute to the increase or decrease of a given property. In order to visualize the results of the SHAP analysis, we have visualized the top ten components in terms of importance and their positive or negative influence and magnitude on the target property in the form of a beeswarm plot in Section 3.3.

3. Results and Discussion

3.1. Analysis of the Datasets Used in This Study

Figures 1–3 and Table 1 summarize a number of statistical metrics for each attribute dataset, including the size of the dataset, the distribution of values, the frequency of occurrence of each compound, and the number of components in each glass. Figure 1 shows the data distribution under each property, and it can be seen that not all properties conform to the normal distribution, which poses a challenge for the optimization and selection of models. Figure 2 shows the number of occurrences of different compounds for each property. In terms of Abbe number, for example, the number of occurrences of B₂O₃ and SiO₂ is high, about 400, much higher than the other compounds. This indicates a certain sparsity in the data set, again challenging the model. Figure 3 shows the number of compounds contained in a group of glasses for each property. As an example, a group

of glasses tends to contain no more than 13 compound species in terms of Abbe number. This provides a reference for the development of new oxyfluoride glasses. Table 1 shows the Abbe number is the smallest of the datasets, with 640 instances, while the refractive index dataset is the largest, with 5209 instances. Despite the relatively small size of the dataset for each property compared to the dataset in the oxide glass study, the dataset in this study still shows significant research value and potential for application in the field of machine learning.

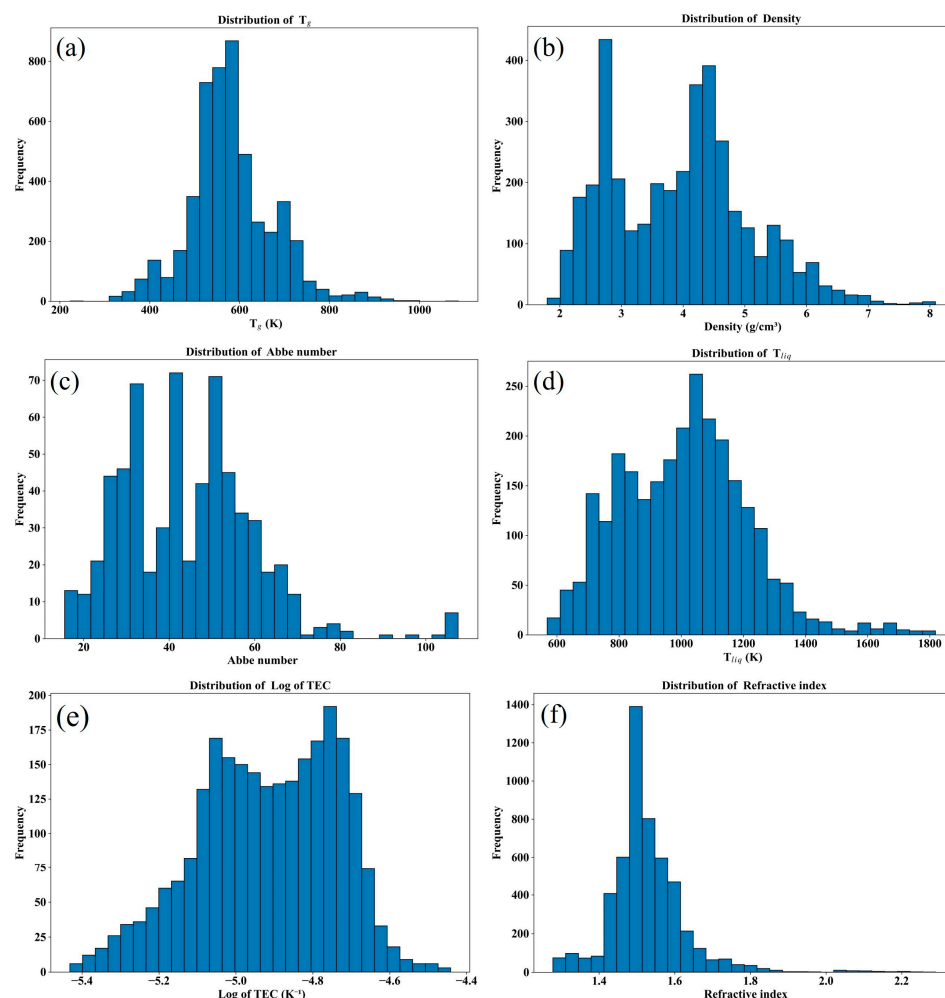


Figure 1. Data distribution of six properties of oxyfluoride glass, (a) glass transition temperature, (b) density, (c) Abbe number, (d) liquidus temperature, (e) thermal expansion coefficient after taking logarithmic values, (f) refractive index.

Figure 1 shows that the distribution of values for some of the glass properties (e.g., Abbe number and density) is more asymmetric compared to other properties, a phenomenon that may be related to the number of examples available for study. It is worth highlighting that Table 1, Figures 2 and 3 together show the uniqueness of the frequency of occurrence, the number of components, and the number of features considered for each type of compound under each property. This abundance of statistical data provides an extremely rich source of information for glass studies and helps to target specific analyses for different properties.

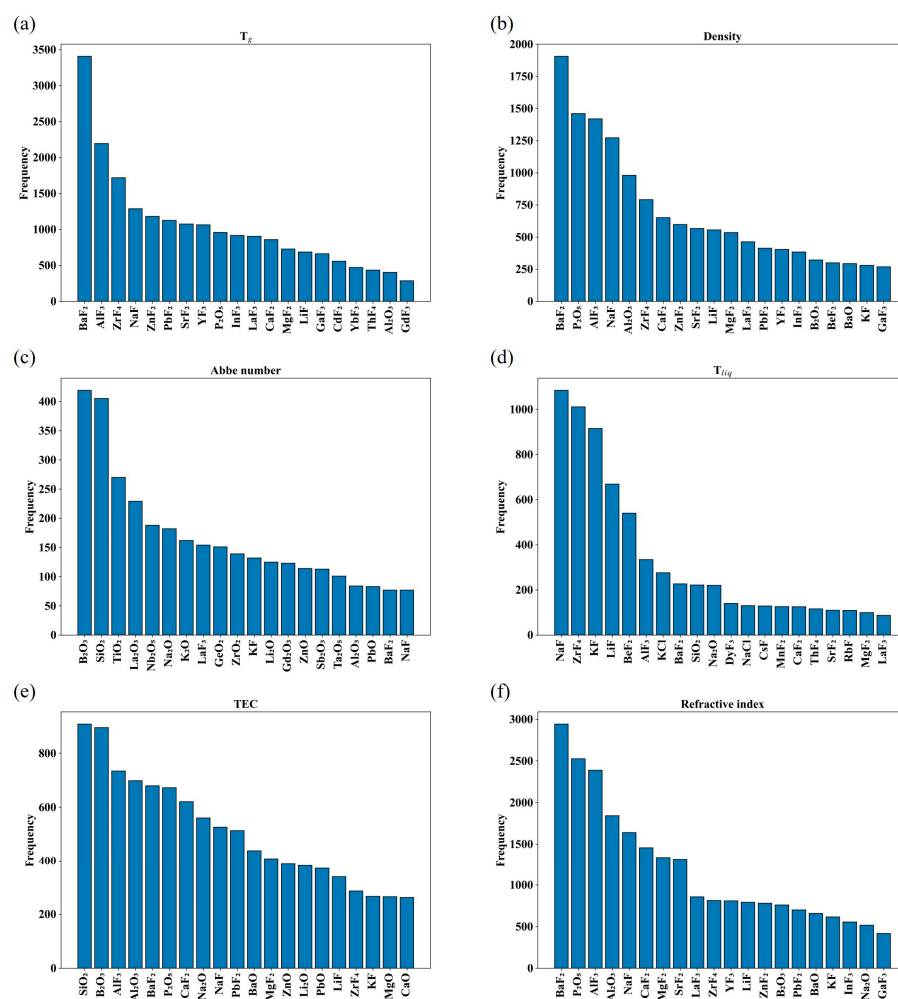


Figure 2. Frequency of occurrence of various compounds for six properties of oxyfluoride glass, (a) glass transition temperature, (b) density, (c) Abbe number, (d) liquidus temperature, (e) thermal expansion coefficient, (f) refractive index.

3.2. Predictive Performance Measures

Figures 4–9 show the subplots of the predicted and experimental deviations, R^2 scores, and probability density functions of the errors (insets located in the upper left corner of the figures) for the six properties: glass transition temperature, density, Abbe number, liquidus temperature, thermal expansion coefficient, and refractive index for the four algorithms on both the training and test sets. It is important to note that the four algorithms are carefully hyper-parameterized for each attribute (the hyper-parameters are recorded in Tables 2 and 3), and thus, the optimal model for each algorithm is presented in the figure. Meanwhile, a 45-degree straight line representing the ideal state is marked in each figure, and the closer the data points are to this line, the closer the predicted values are to the experimental values and the better the prediction effect of the model is. In addition, the probability density function (PDF) of the errors and their corresponding 90% confidence intervals (presented on a light background) are attached in the upper left corner of each figure to visualize the distribution of the errors of the algorithms. We also count the training time and other performance metrics, including MAE and RMSE, of each optimized model in Tables 4 and 5, which allow us to visually and comprehensively observe the differences between different algorithms.

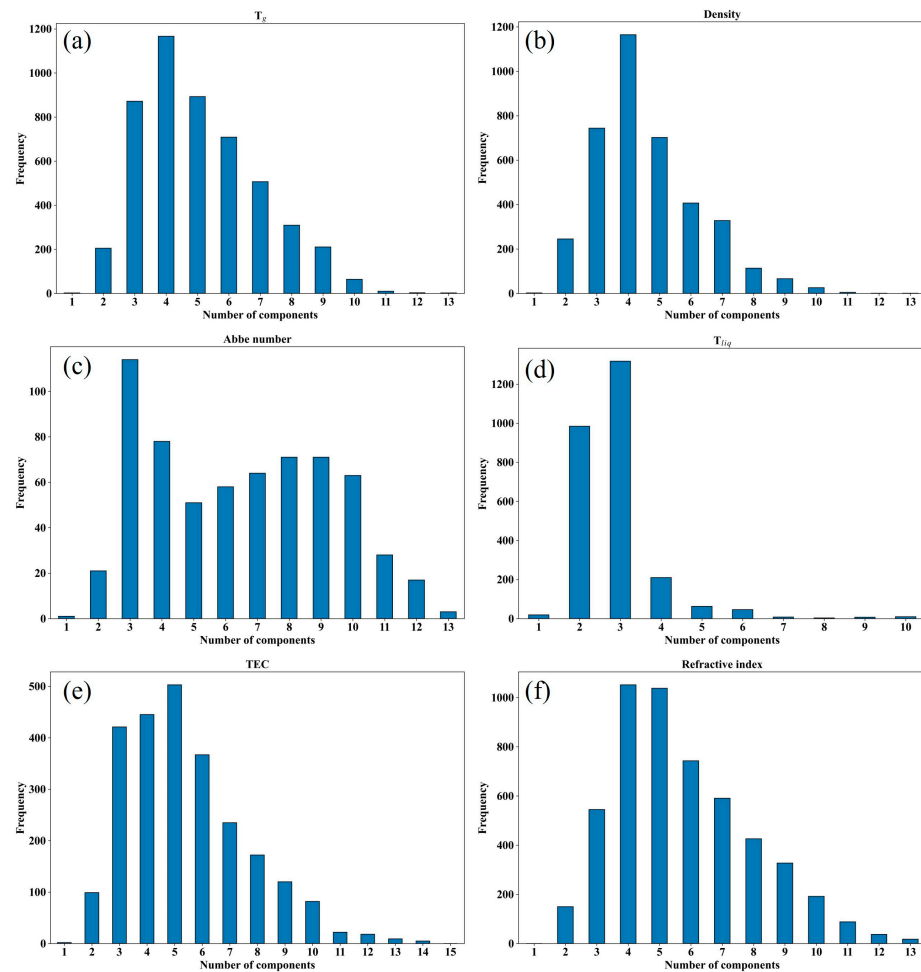


Figure 3. Frequency of the number of glass components for six properties of oxyfluoride glass, (a) glass transition temperature, (b) density, (c) Abbe number, (d) liquidus temperature, (e) thermal expansion coefficient, (f) refractive index.

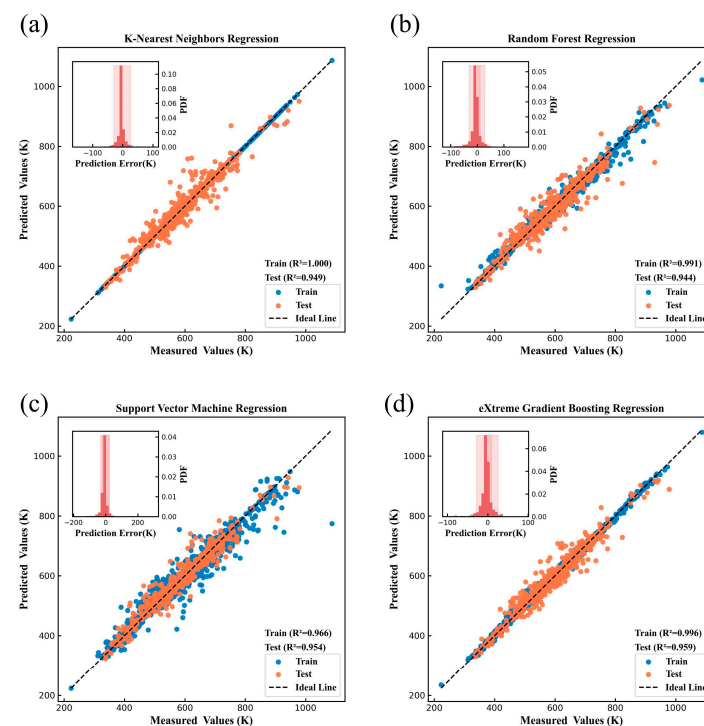


Figure 4. Prediction of glass transition temperature under (a) KNN, (b) RF, (c) SVM, (d) XGBoost.

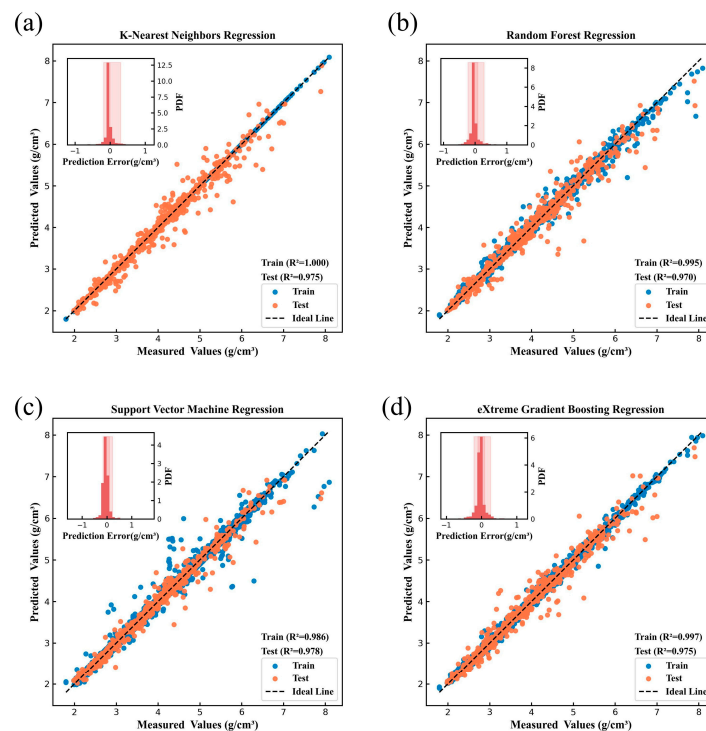


Figure 5. Prediction of density under (a) KNN, (b) RF, (c) SVM, (d) XGBoost.

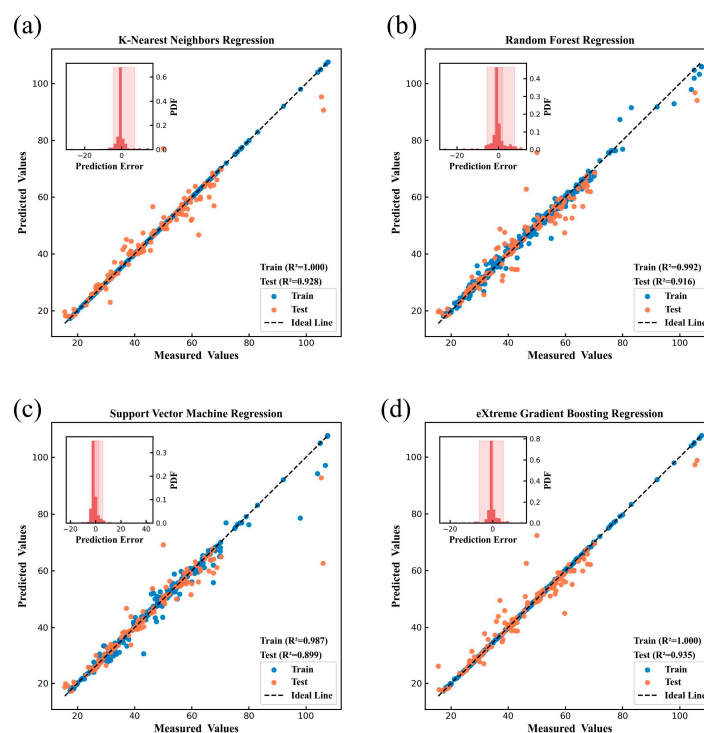


Figure 6. Prediction of Abbe number under (a) KNN, (b) RF, (c) SVM, (d) XGBoost.

In the field of machine learning, the R^2 score is used as a widely used evaluation metric to measure the proportion of the variance in the predicted variable that can be explained by the characteristics of the independent variable and the model. The closer its value tends to 1, it means the better the model fits the data and the better the model's explanatory and predictive performance. MAE (Mean Absolute Error), as an intuitive measure of error, directly reflects the average level of prediction deviation by calculating the mean of the absolute difference between the predicted value and the true value, is insensitive to outliers,

is robust, and facilitates an intuitive understanding of the actual size of the prediction error. RMSE (Root Mean Square Error) focuses on the measurement of the degree of dispersion of the predicted value by taking the root of the square of the error, highlighting the effect of extreme errors, and can sensitively reflect the model's ability to deal with outliers.

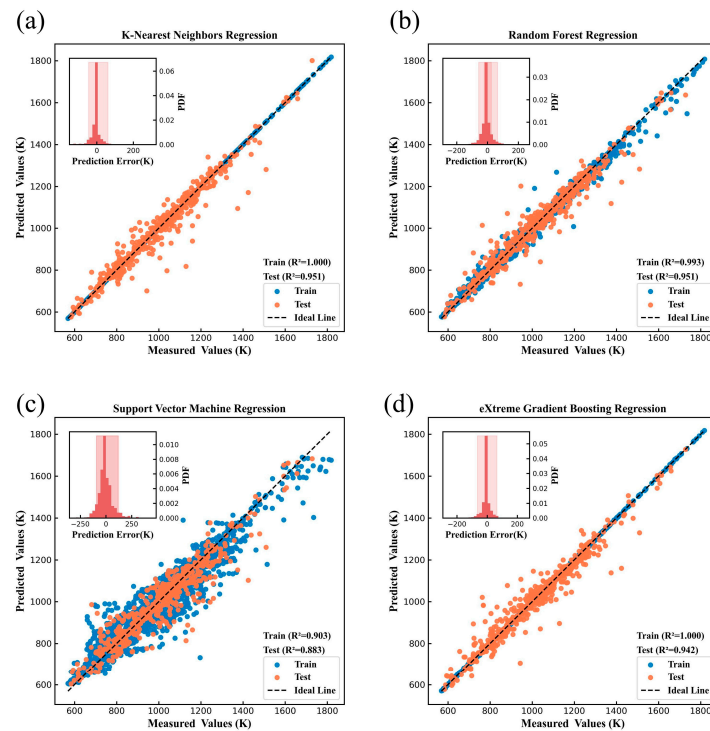


Figure 7. Prediction of liquidus temperature under (a) KNN, (b) RF, (c) SVM, (d) XGBoost.

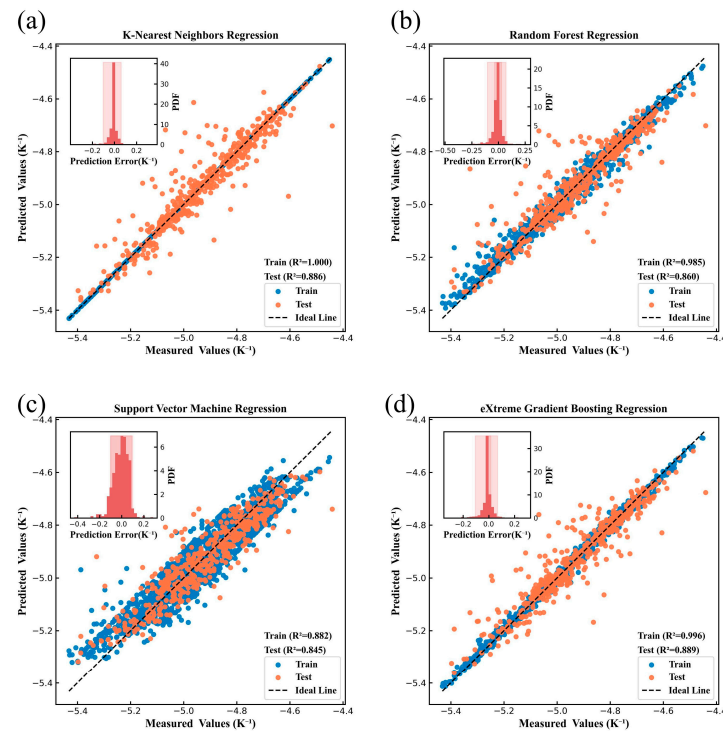


Figure 8. Prediction of thermal expansion coefficient under (a) KNN, (b) RF, (c) SVM, (d) XGBoost.

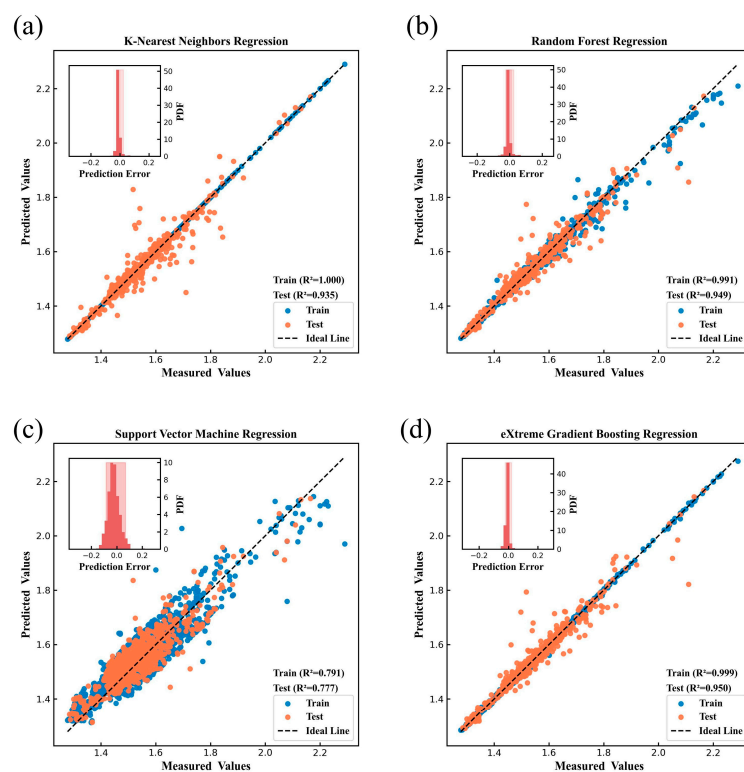


Figure 9. Refractive index prediction under (a) KNN, (b) RF, (c) SVM, (d) XGBoost.

Table 4. The training time of each optimized model (unit: seconds).

	T_g	Density	AbbeNumber	T_{liq}	$\text{Log}_{10}(\text{TEC})$	RefractiveIndex
KNN	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
RF	7.7798	6.9369	2.0694	4.3618	8.2211	16.2200
SVM	1.5781	0.5985	0.0380	0.1833	0.0370	0.0812
XGBoost	0.6149	0.3840	0.4191	0.5659	0.9105	0.5850

Table 5. The performance of different algorithms under various evaluation indicators.

		T_g	Density	AbbeNumber	T_{liq}	$\text{Log}_{10}(\text{TEC})$	RefractiveIndex
R^2	KNN	0.949	0.975	0.928	0.951	0.886	0.935
	RF	0.944	0.970	0.916	0.951	0.860	0.949
	SVM	0.954	0.978	0.899	0.883	0.845	0.777
	XGBoost	0.959	0.975	0.935	0.942	0.889	0.950
MAE	KNN	11.588	0.094	2.223	25.111	0.034	0.011
	RF	12.290	0.102	2.652	27.253	0.038	0.011
	SVM	11.150	0.094	2.070	46.643	0.052	0.037
	XGBoost	10.939	0.097	2.176	29.622	0.036	0.010
RMSE	KNN	20.104	0.176	4.189	44.177	0.059	0.026
	RF	21.149	0.192	4.530	44.365	0.065	0.023
	SVM	19.094	0.167	4.980	68.118	0.068	0.048
	XGBoost	18.024	0.175	3.982	47.863	0.058	0.023

Figures 4–9, Tables 4 and 5 show that different algorithms can achieve different optimal performance on datasets with different attributes and can construct the best model for the attributes. In terms of glass transition temperature, XGBoost has the best performance, followed by SVM. In terms of density, all four algorithms show high performance, with KNN and SVM standing out. In terms of Abbe number, XGBoost shows the best performance, followed by KNN. In terms of liquid phase temperature, KNN and RF show the best performance. In this study, the coefficient of thermal expansion is the more difficult

property to predict, and the four algorithms perform slightly lower than the other properties in this property, with KNN showing the best performance, followed by XGBoost. In refractive index, XGBoost and RF perform well. It is also worth noting that in terms of training time, KNN has a significant advantage, XGBoost and SVM are moderate, and RF consumes the longest time, which requires extra attention when dealing with larger and more complex datasets.

Overall, the KNN algorithm is simple and efficient, with outstanding performance in the prediction of multiple properties, but the algorithm itself has low complexity and weak ability to capture nonlinear relationships; RF improves the model complexity through integrated learning, but the training time is relatively long, and it is weak in dealing with high-dimensional sparse datasets; the SVM algorithm is robust, and it performs well in the prediction of some properties, but the algorithm requires higher requirements for the selection and optimization of the kernel function; XGBoost algorithm is relatively more complex, but the training time is moderate, and the performance achieved is outstanding in almost all properties. Therefore, we choose the model constructed based on XGBoost for the subsequent research work to ensure the consistency of the evaluation system in the subsequent SHAP analysis and ternary system prediction process.

3.3. Interpreting the Induced Models

As mentioned above, we applied the SHAP algorithm to describe the optimal models we constructed for the six properties and analyzed the contribution of each compound to the predicted properties [41]. In Figure 10, we visualize the top ten compounds in terms of importance in each property in the form of beeswarm plots, ranked from top to bottom according to the level of importance.

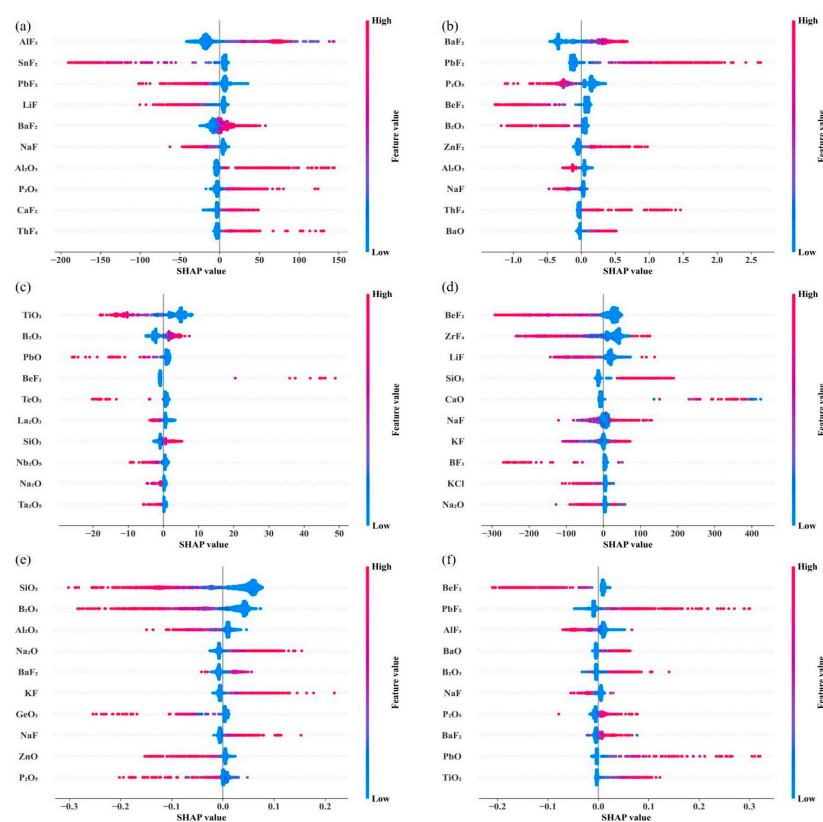


Figure 10. Beeswarm plot investigating SHAP values for six properties. (a) glass transition temperature, (b) density, (c) Abbe number, (d) liquidus temperature, (e) thermal expansion coefficient, and (f) refractive index. The labels on the left show the ten most important compounds identified by SHAP analysis in decreasing order of importance.

With the help of these beeswarm plots, researchers were able to qualitatively analyze the effect of compounds on properties. Taking the glass transition temperature as an example, the components can be categorized into two groups based on the direction of their influence on the predicted values: those that contribute to an increase in the predicted values, including AlF_3 , BaF_2 , Al_2O_3 , P_2O_5 , CaF_2 , ThF_4 , and those that lead to a decrease in the predicted values, which are, in order of importance, SnF_2 , PbF_2 , LiF , and NaF . It is worth noting that the importance of the components changes across the different properties analyzed, and some of the components may show a shift from increasing to decreasing model output. In addition, the synergistic effect of components on multiple properties should not be overlooked. For example, the presence of BeF_2 causes a decrease in the values of density, liquidus temperature, and refractive index, while NaF drives an increase in the thermal expansion coefficient while attenuating the glass transition temperature and density. This complex relationship needs to be emphasized in the development of new oxyfluoride glass designs as well as in computer-aided reverse design.

3.4. Model Application

To further explore the model performance, Figure 11 shows the ternary diagrams of different oxyfluoride glass systems [42–44], which cover the prediction results of glass transition temperature, density, Abbe number, liquidus temperature, thermal expansion coefficient, and refractive index. As seen from the plots, the properties of oxyfluoride glass vary with the ratio of different components, which can help to rationalize the design of the composition of oxyfluoride glass. Taking liquidus temperature as an example, when a high liquidus temperature is required, the proportion of MgF_2 in the glass should be controlled at 85–95%, AlF_3 at 0–10%, and SiO_2 at 0–10%; if a low liquidus temperature is expected, the proportion of MgF_2 in the glass should be controlled at 0–10%, AlF_3 at 80–90%, and SiO_2 at 0–10%. It can be seen that the model is capable of making predictions for the complete compositional domain, finding better solutions for the requirements, and exploring new types of glasses that have not been exploited so far.

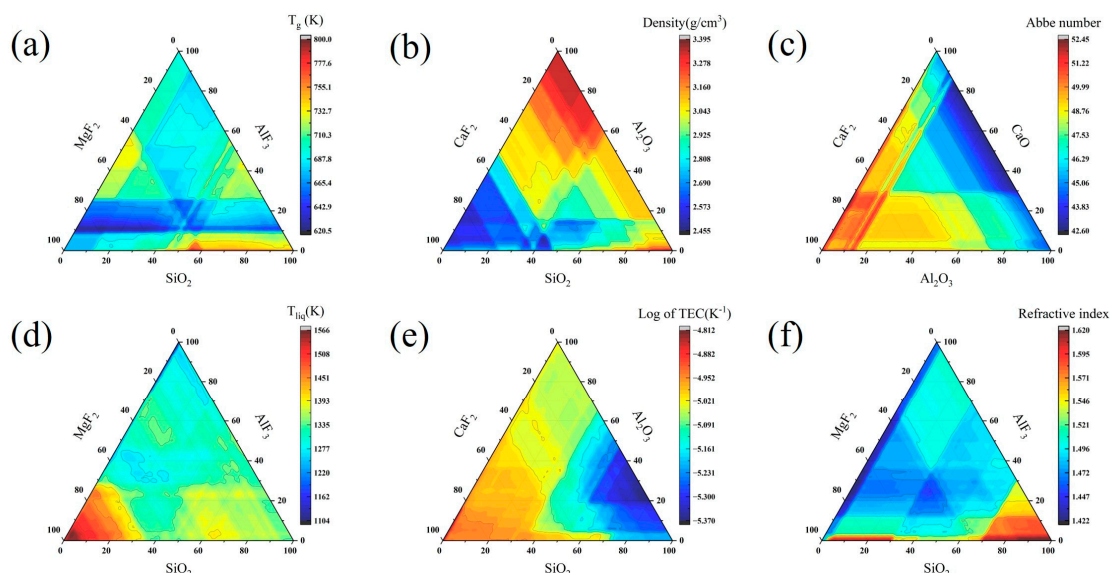


Figure 11. Ternary diagrams for different systems: (a) glass transition temperature of MgF_2 , AlF_3 , and SiO_2 systems, (b) density of CaF_2 , Al_2O_3 , and SiO_2 systems, (c) Abbe number of CaF_2 , CaO , and Al_2O_3 systems, (d) liquidus temperature of MgF_2 , AlF_3 , SiO_2 systems, (e) thermal expansion coefficient (logarithmic form) of CaF_2 , Al_2O_3 , SiO_2 systems, (f) refractive index of MgF_2 , AlF_3 , SiO_2 systems.

However, it is important to note that some of the components shown in the ternary diagrams may not be able to form glass under actual experimental conditions. Therefore, in regions with missing or sparse data points, the predictions of the model may not be reliable. In this case, expert knowledge plays a crucial role in analyzing the validity of the predictions. Even so, the model shows a strong potential for application in the field of computer-aided design and can provide technical support for the design and development of oxyfluoride glass.

4. Conclusions

In this research work, we collected an extensive dataset of oxyfluoride glasses and carried out the training and study of four machine-learning algorithms for six glass properties. Specifically, we employed K-Nearest Neighbors, Random Forests, Support Vector Machines, and eXtreme Gradient Boosting algorithms to construct prediction models for glass transition temperature, density, Abbe number, liquidus temperature, thermal expansion coefficient, and refractive index, respectively. In order to achieve the optimal performance of each model, we applied the grid search method to adjust the model hyperparameters. The results show that the XGBoost algorithm exhibits overall excellent performance in the task of predicting the above six glass properties, and the rest of the algorithms perform well for some of the properties. In addition, with the help of the SHAP algorithm, we have explained the effect of individual components on the properties. This can help in the development of new oxyfluoride glasses as well as computer-aided reverse design. Finally, we have predicted the properties of oxyfluoride glass for different ternary systems. Although the relatively limited number of dataset instances used to train the model results in a dip in the model's performance in the data-sparse region, our model is still important for screening or designing novel glass materials with the right combination of properties. In summary, our study bridges the gap of composition-property relationship in the field of oxyfluoride glass through machine learning and can also serve as a reference for the study of other glass systems.

Author Contributions: Methodology, Y.X.; software, Y.X.; investigation, Y.X.; data curation, Y.X.; writing—original draft preparation, Y.X.; writing—review and editing, X.W.; methodology, X.W.; funding acquisition, X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the project of Yunnan Key Laboratory of Electromagnetic Materials and Devices, Yunnan University (No. ZZ2024005), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (Grant No. 23KJA510005).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Polishchuk, S.A.; Ignat'eva, L.N.; Marchenko, Y.V.; Bouznik, V.M. Oxyfluoride glasses (a review). *Glass Phys. Chem.* **2011**, *37*, 1–20. [[CrossRef](#)]
2. Fedorov, P.P.; Luginina, A.A.; Popov, A.I. Transparent oxyfluoride glass ceramics. *J. Fluor. Chem.* **2015**, *172*, 22–50. [[CrossRef](#)]
3. Ravinder; Venugopal, V.; Bishnoi, S.; Singh, S.; Zaki, M.; Grover, H.S.; Bauchy, M.; Agarwal, M.; Krishnan, N.M.A. Artificial intelligence and machine learning in glass science and technology: 21 challenges for the 21st century. *Int. J. Appl. Glass Sci.* **2021**, *12*, 277–292. [[CrossRef](#)]
4. Lu, X.; Vienna, J.D.; Du, J. Glass formulation and composition optimization with property models: A review. *J. Am. Ceram. Soc.* **2024**, *107*, 1603–1624. [[CrossRef](#)]
5. Dreyfus, C.; Dreyfus, G. A machine learning approach to the estimation of the liquidus temperature of glass-forming oxide blends. *J. Non-Cryst. Solids* **2003**, *318*, 63–78. [[CrossRef](#)]
6. Brauer, D.S.; Rüssel, C.; Kraft, J. Solubility of glasses in the system P_2O_5 –CaO–MgO–Na₂O–TiO₂: Experimental and modeling using artificial neural networks. *J. Non-Cryst. Solids* **2007**, *353*, 263–270. [[CrossRef](#)]

7. Echezarreta-López, M.M.; Landin, M. Using machine learning for improving knowledge on antibacterial effect of bioactive glass. *Int. J. Pharmaceut.* **2013**, *453*, 641–647. [\[CrossRef\]](#)
8. Mauro, J.C.; Tandia, A.; Vargheese, K.D.; Mauro, Y.Z.; Smedskjaer, M.M. Accelerating the design of functional glasses through modeling. *Chem. Mater.* **2016**, *28*, 4267–4277. [\[CrossRef\]](#)
9. Cassar, D.R.; de Carvalho, A.C.; Zanotto, E.D. Predicting glass transition temperatures using neural networks. *Acta Mater.* **2018**, *159*, 249–256. [\[CrossRef\]](#)
10. Krishnan, N.M.A.; Mangalathu, S.; Smedskjaer, M.M.; Tandia, A.; Burton, H.; Bauchy, M. Predicting the dissolution kinetics of silicate glasses using machine learning. *J. Non-Cryst. Solids* **2018**, *487*, 37–45. [\[CrossRef\]](#)
11. Bishnoi, S.; Singh, S.; Ravinder, R.; Bauchy, M.; Gosvami, N.N.; Kodamana, H.; Krishnan, N.M.A. Predicting Young's modulus of oxide glasses with sparse datasets using machine learning. *J. Non-Cryst. Solids* **2019**, *524*, 119643. [\[CrossRef\]](#)
12. Ravinder, R.; Sridhara, K.H.; Bishnoi, S.; Grover, H.S.; Bauchy, M.; Kodamana, H.; Krishnan, N.M.A. Deep learning aided rational design of oxide glasses. *Mater. Horiz.* **2020**, *7*, 1819–1827. [\[CrossRef\]](#)
13. Deng, B. Machine learning on density and elastic property of oxide glasses driven by large dataset. *J. Non-Cryst. Solids* **2020**, *529*, 119768. [\[CrossRef\]](#)
14. Cassar, D.R.; Mastelini, S.M.; Botari, T.; Alcobaca, E.; de Carvalho, A.C.; Zanotto, E.D. Predicting and interpreting oxide glass properties by machine learning using large datasets. *Ceram. Int.* **2021**, *47*, 23958–23972. [\[CrossRef\]](#)
15. Cassar, D.R. ViscNet: Neural network for predicting the fragility index and the temperature-dependency of viscosity. *Acta Mater.* **2021**, *206*, 116602. [\[CrossRef\]](#)
16. Bishnoi, S.; Ravinder, R.; Grover, H.S.; Kodamana, H.; Krishnan, N.M.A. Scalable Gaussian processes for predicting the optical, physical, thermal, and mechanical properties of inorganic glasses with large datasets. *Mater. Adv.* **2021**, *2*, 477–487. [\[CrossRef\]](#)
17. Zaki, M.; Venugopal, V.; Bhattoo, R.; Bishnoi, S.; Singh, S.K.; Allu, A.R.; Jayadeva; Krishnan, N.M.A. Interpreting the optical properties of oxide glasses with machine learning and Shapely additive explanations. *J. Am. Ceram. Soc.* **2022**, *105*, 4046–4057. [\[CrossRef\]](#)
18. Bishnoi, S.; Badge, S.; Krishnan, N.M.A. Predicting oxide glass properties with low complexity neural network and physical and chemical descriptors. *J. Non-Cryst. Solids* **2023**, *616*, 122488. [\[CrossRef\]](#)
19. Cassar, D.R. GlassNet: A multitask deep neural network for predicting many glass properties. *Ceram. Int.* **2023**, *49*, 36013–36024. [\[CrossRef\]](#)
20. Liu, C.; Su, H. Prediction of glass transition temperature of oxide glasses based on interpretable machine learning and sparse data sets. *Mater. Today Commun.* **2024**, *40*, 109691. [\[CrossRef\]](#)
21. Tian, J.; Zhao, Y.; Huang, Y.; Li, Y.; Zhang, C.; Peng, S.; Han, G.; Liu, Y. Theoretical prediction of Vickers hardness for oxide glasses: Machine learning model, interpretability analysis, and experimental validation. *Materialia* **2024**, *33*, 102006. [\[CrossRef\]](#)
22. Kang, Z.; Wang, L.; Li, X.; Gao, W.; Dong, X.; Li, J.; Cao, Y.; Yue, Y.; Kang, J. Interpretable machine learning accelerates development of high-specific modulus glass. *Comp. Mater. Sci.* **2025**, *246*, 113482. [\[CrossRef\]](#)
23. Mastelini, S.M.; Cassar, D.R.; Alcobaça, E.; Botari, T.; de Carvalho, A.C.; Zanotto, E.D. Machine learning unveils composition–property relationships in chalcogenide glasses. *Acta Mater.* **2022**, *240*, 118302. [\[CrossRef\]](#)
24. Singla, S.; Mannan, S.; Zaki, M.; Krishnan, N.M.A. Accelerated design of chalcogenide glasses through interpretable machine learning for composition–property relationships. *J. Phys-Mater.* **2023**, *6*, 024003. [\[CrossRef\]](#)
25. Ahmmad, S.K.; Jabeen, N.; Ahmed, S.T.U.; Hussainy, S.F.; Ahmed, B. Density of fluoride glasses through artificial intelligence techniques. *Ceram. Int.* **2021**, *47*, 30172–30177. [\[CrossRef\]](#)
26. SciGlass. 2025. Available online: <https://github.com/epam/SciGlass> (accessed on 15 March 2025).
27. Glasspy. 2025. Available online: <https://glasspy.readthedocs.io/en/latest/index.html> (accessed on 15 March 2025).
28. Peterson, L.E. K-nearest neighbor. *Scholarpedia* **2009**, *4*, 1883. [\[CrossRef\]](#)
29. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
30. López, O.A.M.; López, A.M.; Crossa, J. Support vector machines and support vector regression. In *Multivariate Statistical Machine Learning Methods for Genomic Prediction*; Springer International Publishing: Cham, Switzerland, 2022; pp. 337–378. [\[CrossRef\]](#)
31. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13 August 2016; pp. 785–794. [\[CrossRef\]](#)
32. Agrawa, T. *Hyperparameter Optimization in Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2021. [\[CrossRef\]](#)
33. Bartz, E.; Bartz-Beielstein, T.; Zaefferer, M.; Mersmann, O. *Hyperparameter Tuning for Machine and Deep Learning with R: A Practical Guide*; Springer Nature: Berlin/Heidelberg, Germany, 2023; p. 323. [\[CrossRef\]](#)
34. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [\[CrossRef\]](#)
35. Reback, J.; McKinney, W.; Van Den Bossche, J.; Augspurger, T.; Cloud, P.; Klein, A.; Hawkins, S.; Roeschke, M.; Tratner, J.; She, C.; et al. *Pandas-Dev/Pandas: Pandas 1.0.5*; Zenodo: Genève, Switzerland, 2020.
36. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [\[CrossRef\]](#)

37. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
38. Liu, L.; Chen, C.; Wang, B. Predicting financial crises with machine learning methods. *J. Forecast.* **2022**, *41*, 871–910. [[CrossRef](#)]
39. Mangalathu, S.; Hwang, S.H.; Jeon, J.S. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Eng. Struct.* **2020**, *219*, 110927. [[CrossRef](#)]
40. Lyngdoh, G.A.; Li, H.; Zaki, M.; Krishnan, N.M.A.; Das, S. Elucidating the constitutive relationship of calcium–silicate–hydrate gel using high throughput reactive molecular simulations and machine learning. *Sci. Rep.* **2020**, *10*, 21336. [[CrossRef](#)] [[PubMed](#)]
41. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [[CrossRef](#)]
42. Wang, R.; Zhou, J.; Li, B.; Li, L. Study on the properties of $\text{AlF}_3\text{-MgF}_2\text{-SiO}_2$ system low temperature co-fired oxyfluoride glass-ceramics. *Rare Metal. Mat. Eng* **2009**, *38*, 1117–1119.
43. Duan, R.; Liang, K.; Gu, S. The Structure of $\text{CaF}_2\text{-Al}_2\text{O}_3\text{-SiO}_2$ System Glass. *J. Inorg. Mater.* **1998**, *13*, 593–598.
44. Peng, K.; Liu, K.; Han, Q.; Wang, Y.; Ma, H. Study on properties of $\text{Al}_2\text{O}_3\text{-CaO-CaF}_2$ SLAG. *Ferro-Alloys* **2010**, *6*, 15–17.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.