Accelerated design of chalcogenide glasses through interpretable machine learning for composition–property relationships

Sayam Singla, Sajid Mannan, Mohd Zaki and N M Anoop Krishnan

1 Department of Textile and Fibre Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India
2 Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India
3 Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India
* Author to whom any correspondence should be addressed.
E-mail: krishnan@iitd.ac.in

Abstract
Chalcogenide glasses (ChGs) possess various outstanding properties enabling essential applications, such as optical discs, infrared cameras, and thermal imaging systems. Despite their ubiquitous usage, these materials’ composition–property relationships remain poorly understood, impeding the pace of their discovery. Here, we use a large experimental dataset comprising ∼24 000 glass compositions made of 51 distinct elements from the periodic table to develop machine learning (ML) models for predicting 12 properties, namely, annealing point, bulk modulus, density, Vickers hardness, Littleton point, Young’s modulus, shear modulus, softening point, thermal expansion coefficient, glass transition temperature, liquidus temperature, and refractive index. These models are the largest regarding the compositional space and the number of properties covered for ChGs. Further, we use Shapley additive explanations, a game theory-based algorithm, to explain the properties’ compositional control by quantifying each element’s role toward model predictions. This work provides a powerful tool for interpreting the model’s prediction and designing new ChG compositions with targeted properties. Finally, using the trained ML models, we develop several glass-selection charts that can potentially aid in the rational design of novel ChGs for various applications.

1. Introduction
Chalcogenide glasses (ChGs) are a family of glasses that contains one or more chalcogen elements, namely, sulfur (S), selenium (Se), or tellurium (Te), in combination with other elements from Group IV, V, and VI of the periodic table except for oxygen [1, 2]. The ChGs possess unique properties compared to oxide glasses due to different types of bonds amongst their constituents. The atoms in ChGs are held together by covalent bonds, stronger than ionic bonds in oxide glasses [3], which provides their characteristic small band gap energy ($E_g \sim 1–3$ eV). This feature offers unique properties such as semi-conductivity, photosensitivity, and fast ionic conductivity ($>10^{-3}$ S cm$^{-1}$) and makes them attractive candidates in optics and electronics [4]. Due to these reasons, they are widely used in optical discs, infrared cameras, and thermal imaging systems for automobile navigation, as a waveguide in optical circuits, and as a solid electrolyte in all-solid-state batteries [5, 6], which are active areas of research and development.

Although the glass-forming ability of chalcogens has been known for several decades [7], compared to oxide glasses, they remain under-explored, which is exemplified by the fact that databases like SciGlass and INTERGLAD have more than ∼300 000 oxide glasses compositions compared to ∼30 000 ChG compositions [8, 9]. One of the primary reasons for the limited understanding is that the composition–property relationships in these glasses are still elusive. The Edisonian trial-error method was traditionally used to
develop new glasses based on experience and domain knowledge. Still, this approach is resource-intensive and time-consuming. Recently, researchers have been actively using machine learning (ML) algorithms to develop composition-dependent property models with the available curated databases for predicting properties such as Young’s modulus, glass transition temperature, and refractive index of oxide glasses [10–21]. These approaches, combined with artificial intelligence-based information extraction from the literature, have proven to significantly accelerate materials design and discovery by enabling the automated development of databases and knowledge bases from the literature [19, 20, 22–24]. In addition, data-driven modeling has proved to be quite helpful in understanding the composition–property relationships, thereby enhancing the design of new glasses [25, 26].

However, there have been limited studies on predicting and interpreting the properties of ChGs. For instance, recently, Mastelini et al [26] trained different ML models such as random forest, K-nearest neighbors, neural network, and classification and regression trees for predicting three properties of ChGs, namely, glass transition temperature, refractive index, and thermal expansion coefficient (TEC) using the dataset collected from SciGlass database. They used approximately 450, 930, and 7620 unique ChGs compositions with 1–6 different elements as a feature in each glass composition for predicting refractive index, coefficient of thermal expansion, and glass transition temperature, respectively. To the best of our knowledge, this is the only prior work where ML has been used extensively for predicting the properties of ChGs. However, compositional control of several other significant physical and mechanical properties of ChGs remains unexplored. Further, glasses that exhibit multiple target properties are preferred from a practical perspective. This requirement motivates the development of ML models corresponding to a large set of physical, mechanical, and optical properties covering a broad compositional spectrum to enable the accelerated design of ChGs.

In this work, we develop ML models for predicting 12 properties of ChGs such as annealing point (AP), bulk modulus (K), density (ρ), Vickers hardness (HV), Littleton point (LP), Young’s modulus (E), shear modulus (G), softening point (SP), TEC, glass transition temperature (Tg), liquidus temperature (Tl), and refractive index (nρ) by considering up to 51 distinct elements from the periodic table. These are the most extensive set of models developed so far for ChGs regarding the number of input features and compositional space compared to the literature [26]. To understand the complexity of these trained ML models, we employ the Shapley additive explanations (SHAP) algorithm based on game theory [27]. To this end, we plot the interaction value plot and selection chart using the developed ML models for each property to see the contribution of different features toward the predicted property values. Altogether, this study will help accelerate the design and discovery of new ChGs with tailored properties.

2. Methodology

2.1. Data collection and data processing

Here, we collected the data from SciGlass and INTERGLAD [8, 9], well-known glass databases for composition–property value. These two datasets contain about 350,000 glass compositions from research papers, patents, and handbooks. We extracted compositions and properties of ChGs from these databases to train ML models for predicting their properties. We follow the data-cleaning methodology mentioned by Mastelini et al (2021) [28]. First, we removed all the glasses which contained any amount of oxygen, nitrogen, fluorine, chlorine, bromine, iodine, gold, silver, platinum, and palladium and considered only those glasses which contained at least one of the non-zero amount of sulfur, selenium, or tellurium. The extracted datasets were converted into atomic mol% using stoichiometric analysis. For example, If the glass has 20% of As2S3 and 80% of As2Te3, then the total molar concentration of each of the components are calculated as shown below

\[
\text{As} = \frac{2}{5} \times 20\% + \frac{2}{5} \times 80\% = 40\%; \quad \text{S} = \frac{3}{5} \times 20\% = 12\%; \quad \text{Te} = \frac{3}{5} \times 80\% = 48\%.
\]

The input feature is a vector comprising the mol% of all the components associated with a property. Note that the values of features that are not present in glass are assigned a value of zero (see supplementary section 9 Stoichiometric calculation of elements for detailed calculations). After converting compositions from compounds to elemental form, we removed all the glasses whose composition percentage did not add up to 100%. Duplicate entries were combined by calculating the mean of all property values within a range of ±2.5% of the mean of all associated property values, while dropping the outliers with Z-scores beyond the interval of ±3 of the corresponding property values [10]. Also, only those elements present in at least ten glasses were taken.
In this work, we report ML models to predict the 12 properties of ChGs: AP, K, ρ, Hr, LP, E, G, SP, TEC, Tg, T1, and nd. The cleaned dataset for each property was further divided into training and test sets in the ratio of 80:20, respectively. The training set was then subjected to four-fold cross-validation for optimizing the hyperparameters. The model that performed best on the validation set was chosen to evaluate the performance on the test set. This method was employed for all 12 properties. All the codes used for training the models are provided at https://github.com/M3RG-IITD/chgs-ai.

2.2. Model development

We trained several ML models: support vector machine, random forest, neural network, and XGBoost [29]. The results of the best model, XGBoost, in this case, for each property are considered (Details of each model is provided in the supplementary table S4). The elemental composition of glass in mol% is used as the input features for all 12 properties, and the output of the model is the accompanying property values in their respective units. Using the squared error metric as the loss function for the optimizer, we used two types of boosters, namely, ‘gbtree’ and ‘dart’.

In addition, hyperparameter tuning was performed with the Optuna package [30]. Optuna provides a quick sampling and pruning approach to optimize the objective function based on the validation score for a given set of hyperparameters. Independent (tree-structured parzen estimator [31]) and relational (covariance matrix adaptation evolution strategy [32], generalized pseudo bayesian optimization [33]) sampling techniques were used to study new trials. It uses pruning techniques like asynchronous successive halving [34] to eliminate futile trials. This library also provides the functionality for adding custom sampling and pruning methods for obtaining an ML model with optimal hyperparameters.

2.3. Shapley additive explanations (SHAP)

SHAP is a unified and interpretable game theory approach for explaining the output of any ML model. It explains model predictions by quantifying the contribution of each feature towards the predicted values by respecting the constraints required for an explanation algorithm [10, 27] SHAP determines the feature contribution by calculating the prediction error while perturbing the input feature values. The SHAP values are unique for each prediction and sum up the difference between the model prediction and the expected value of the model’s prediction for that example. There are specific model explanation methods for different ML models, such as DeepSHAP for neural networks, Linear SHAP for linear models, TreeSHAP for tree-based models, and kernel SHAP which is model agnostic, to evaluate the SHAP values. Also, different plots convey the feature’s importance, such as bee swarm, violin, bar, and river flow plots. We used bar and violin plots to explain the features’ importance here. The bar plots tell us the overall impact of individual features on the model prediction. The features are arranged from top to bottom based on the highest absolute SHAP values. To get more insights into the directionality of control, we show the violin plots for all properties that inform about the increment and decrement of property value from their respective mean SHAP values. The mean SHAP value can be defined as the value predicted by the model when none of the features are present. Hence the positive impact of an element for a given property implies that the presence of the element increases the model predictions from the mean for a given glass composition. It is equal to the mean of the property values used to obtain the SHAP values. Here, we plot the top five features governing the property value. In the context of SHAP, the top components are features with the highest absolute SHAP values, meaning they have the most considerable impact (positive or negative) on the prediction.

3. Results and discussion

3.1. Data visualizations

To visualize all the compositions used in this work, we first use an unsupervised ML algorithm, k-means, to cluster the compositions into 30 clusters. We chose the value of clusters to be 30 by performing a parametric analysis and identifying the value when the square sum of the distance between cluster centers and respective points is minimum. Further only those clusters having at least 3% of the total number of data points are separately colored. The remaining families of glasses were combined to form the ‘Miscellaneous’ class. These compositions are then projected in two dimensions for visualization using t-distributed Stochastic Neighbor Embedding (t-SNE) embeddings. Figure 1(a) shows a two-dimensional t-SNE plot of all the glass compositions color-coded into clusters according to the outcomes of k-means clustering. The names of at least three most prevalent glass components appear next to the 13 most prominent clusters. For all 12 properties separate t-SNE plots are included in the supplementary material (see figure S11).

Figure 1(b) shows the number of glasses and elements used for all the properties discussed in this work (i.e. the dimensionality of the dataset corresponding to each property). For most properties, the number of
Figure 1. (a) All the glass compositions used in this study are embedded in a two-dimensional t-SNE plot. The influential families are marked according to the color in legends. (b) Bar chart showing the total number of glass compositions and components for given properties.

3.2. Model predictions
After training the different models using the steps described in the methodology section, we evaluated them on the test dataset hidden during the training and validation phase. The hyperparameters associated with the training are included in the supplementary material (tables S1 and S2). Figures 2 (physical properties) and 3 (mechanical and optical properties) show the comparisons between the measured and predicted properties of ChGs. The points belonging to training and test sets are shown with blue and pink colors, respectively. The performance of trained models is shown as $R^2$ scores within the respective subfigures. Further, the straight line at 45° gives a qualitative idea of how many points are closer to this line, implying the goodness of the fit. Note that several studies have attempted to discover materials with extreme property values by analyzing the outliers in the data [35, 36]. Accordingly, we have developed ML models by including the outliers. We observe that the inclusion of the outliers leads to poorer performance of the model suggesting that the outlier values may not be reliable and representative of the true property values. Detailed analysis of dataset and model performance are included in supplementary (see figures S1–S3, S5–S9, and tables S5 and S6).

Figures 2(a)–(f) shows predicted vs measured values for six physical properties, namely, AP, LP, SP, $T_L$, TEC, and $T_g$, based on the optimized XGBoost models. For most properties, we observe that the $R^2$ values are high for AP, LP, $T_L$, and $T_g$, suggesting that (a) the models can predict the properties reasonably and (b) elemental compositions mainly govern these properties and is sufficient to predict them. However, some properties, such as the TEC and SP, have a slightly lower value of $R^2$ score, suggesting that these properties might be influenced by other parameters like measurement techniques or noise in the measurement of these properties, which might be higher than the other properties discussed.

Figures 3(a)–(f) shows measured vs predicted values for five mechanical properties and one optical property obtained using the trained XGBoost model. It has been observed that $R^2$ scores are reasonably high for the $K$, $H_v$, $E$, and $\rho$, which implies these properties are predominantly composition dependent. For some
properties, such as $n_d$ and $G$, ML gives poorer predictions despite having larger data points than properties like AP. To analyze this further, we plotted the distribution of the output data and the number of glass compositions associated with a given input feature for all three properties (see supplementary material figures S1 and S3). We observe that although AP has only 60 data points, it has only four features. In contrast,
and $G$ have a large number of input features, with many of those present in only a few glass compositions. This could lead to poor training of the model as the number of data points associated with a given input feature is low for these properties. Further, especially in the case of $G$, which is generally measured indirectly, we observe that the data is highly clustered around two values, one lower and one higher. Such non-uniform distribution of the data could also cause the ML model's poor performance.

It is clear from figures 2 and 3 that both training and test sets have some data points for which the experimental and predicted values differ, leading to reduced performance of the trained models. To investigate the reason behind such cases, we take all the glasses with a given element and compute the residual between the measured and predicted values, as shown in figures S10 (a)–(l) in supplementary material. It is observed that such data points contain elements in fewer glasses; thus, learning their relationship with property values became challenging.

3.3. Elucidating the composition–property relationship

Now, we apply the Shapley additive explanations (SHAP), a game-theoretic technique, to explain the compositional control of the properties. SHAP measures the contribution of the glass components toward the final property value from the mean value for each composition. Thus, it enables both qualitative and quantitative interpretations of the role played by the glass components in controlling a particular property for each of the 12 properties. Top five glass components for each property, in terms of their mean SHAP values, are shown in figures 4 and 5 (see figure S4 in supplementary for the complete list of components). The components are arranged from top to bottom, with the most critical component at the top and the element having the least effect on model predictions at the bottom based on their absolute SHAP value. Note that the SHAP violin plots essentially reveal the influence of a component toward increasing or decreasing property values. The SHAP values on the x-axis provide the quantitative change that a given element will induce to the predicted property values. The results of SHAP analysis can be used to determine the following:

First, elements that affect property values positively or negatively; second, elements that have conflicting effects depending on the other components in the glasses; and lastly, the precise change in the predicted property values for if the value of any element changes in each glass composition. We have divided 12 properties into physical, mechanical, and optical groups based on their application for better understanding.

3.4. Physical properties

Figure 4(a) shows that germanium (Ge) enhances the AP, whereas selenium (Se) impacts negatively and thus reduces the property values. Figure 4(b) shows elements like Ge, Se, boron (B), and arsenic (As) show a strong influence on LP. While Ge enhances the LP values, Se has a negative impact; others have mixed effects. Figure 4(c) shows tellurium (Te) and Se as the main elements negatively controlling SP; Ge and As have mixed effects. Figure 4(d) shows elements like zirconium (Zr), Se, Te, and iron (Fe), which show a high impact on $T_L$. Se and Te show a negative impact, Fe shows a positive impact, whereas other elements have mixed effects. This behavior observed in SHAP analysis could be interpreted based on the glass network structure. Adding elements such as Se, Te, and S generally reduces the degree of connectivity in the glass matrix by modifying the network usually formed by Ge and gallium (Ga) in the glass structure. The rings formed by the glasses with the addition of these elements are easier to break, resulting in lower values of physical properties such as AP, LP, SP, and $T_L$.

Figure 4(e) shows that Se and sulfur (S) positively impact the TEC value, Ga has a negative impact, whereas Ge and As show mixed effects. Figure 4(f) depicts the elements like Ge, Se, Te, and Ga, which show higher percentages of value in determining the $T_g$ value. Like other physical properties, Se and Te negatively impact $T_g$, while Ge and Ga positively impact $T_g$. However, if the content of these elements is increased beyond saturation, it can lead to homopolar bonds between the same type of element. In the case of Ge, this could lead to the formation of Ge–Ge bonds. The formation of these homopolar bonds can eventually lead to a reduction in the $T_g$ [37]. The experimentalists have also noticed that as more Ge is added to the glass system, it develops a two-dimensional character held by van der Waals forces, increasing $T_g$ [38]. It is interesting to observe from figure 4(e) that the effect of Ge and Se on TEC is opposite compared to their effect on other properties. This observation is consistent with the fact that highly polymerized glasses will result in lower TEC while that with lower polymerization exhibit larger values [39, 40].

3.5. Mechanical and optical properties

Figure 5(a) shows the top five elements that govern the bulk modulus of ChGs. Se reduces the model predictions of $K$, whereas Ni, B, Cu, and Ge have a primarily increasing effect on the predicted values. Figure 5(b) shows that high percentages of copper (Cu), Se, and As increase the model prediction of $H_v$.
Higher Cu concentration positively impacts $H_v$, and Cu shows a linear dependence with $H_v$ \cite{41}. Se also shows a positive impact, whereas As shows mixed effects on $H_v$. Figure 5(c) shows the prominent elements governing the density of ChGs. It is observed that Pb and Tl impact the model prediction of $\rho$ positively. This could be attributed to the relatively heavier atomic mass of the network formers, which dominates the density \cite{42}. It also shows that the higher S content negatively impacts the $\rho$, whereas Te has a mixed effect on $\rho$. It is experimentally proven in the literature that higher S content negatively impacts and shows linear dependence with $\rho$ \cite{43}. Figure 5(d) shows the compositional control of the $E$ value where Se and S show the model predictions of $E$. In figure 5(e), Cu and Ni positively influence $G$, whereas Se and Ge show mixed effects; this could be attributed to the increase in tetrahedral structures of [GeSe$_4$] in the glass matrix that leads to the formation of a rigid network structure of glass leading to enhancement of elastic properties \cite{44}.

Figure 4. Interpretation of composition–property relation for physical properties using SHAP bar and violin plots. (a) Annealing point (AP), (b) Littleton point (LP), (c) softening point (SP), (d) liquidus temperature ($T_L$), (e) thermal expansion coefficient (TEC), and (f) glass transition temperature ($T_g$).
Figure 5. Interpretation of composition–property relations for mechanical and optical properties using SHAP bar and violin plots. (a) Bulk modulus ($K$), (b) Vickers hardness ($H_v$), (c) density ($\rho$), (d) Young’s modulus ($E$), (e) shear modulus ($G$), (f) refractive index ($n_d$).

Figure 5(f) shows that As and Te positively impact the $n_d$, whereas S reduces the $n_d$ value, and Se shows a mixed effect. $n_d$ is mainly governed by the network structure’s electron shell polarizability and packing density. As network formers, As and Te form a chain with other elements and enhance the $n_d$ [45]. Altogether, the SHAP plots provide the interpretation of the individual contribution of elements towards the property predictions.
3.6. Interaction value plot
Understanding the composition–property relationships and the interactions among components in governing the properties of ChGs is a crucial aspect of glass science. Examples of such interactions include the mixed-modifier effects and the boron anomaly [14]. Knowing how two elements interact with each other in a multi-component glass only tells the partial governance of properties. Understanding these correlations requires high-precision experimentation, such as magic angle spinning nuclear magnetic resonance studies, or advanced computational simulations, such as density functional theory modeling of the glass structure [10]. Here, using the SHAP interaction values (figures 6 and 7), we seek to ascertain the link between each input component and the attributes. The interaction plots show how different elements interact with each other to govern the properties of ChGs (see supplementary section 10 for detailed calculations).

3.7. Physical properties
Figures 6(a)–(f) show the SHAP interaction values for different physical properties. It may be inferred that the more populated the interaction value plot, the more complex the interactions in the system. It is also interesting to note that the interaction values associated with the same pairs of elements for different properties are different. Specifically, figure 6(a) shows that the pair of network formers Se–As govern the AP of glasses. Figure 6(b) indicates pairs like Ge–B, Ge–Se, and Ge–As have high interaction values for LP. Figure 6(c) indicates that Ge has the highest interaction value with other elements. Pairs like Ge–Te and Ge–Se show higher interactions among all pairs for SP. Figure 6(d) indicates elements like Zr, Se, Te, and As having higher interaction values with other elements. Major pairs with high interaction values for $T_L$ are Se–Te, Zr–Se, Ge–Se, Se–Pb, and S–Se. Figure 6(e) indicates Ge, S, Se, and As, with $S$–Se, Ge–As, and Ge–S being the three combinations of elements with the highest interaction values for TEC. Figure 6(f) indicates elements like Ge, Se, As, and Te having higher interaction values with other elements. Pairs like Ge–Se, Ge–As, Se–As, and Se–S show higher interaction for $T_g$ pairs.

3.8. Mechanical and optical properties
Figure 7(a) indicates Se, B, Ni, and Ge, with Ni–B, Se–As, and Se–Ge being the three combinations of elements with the highest interaction values. Figure 7(b) shows elements like Cu, As, and Se having higher interaction values with other elements. Pairs like Cu–Se, Cu–As, and Se–As are the most influencing
Figure 7. Interpretation of interdependency among input features for mechanical properties using a heat map. (a) Bulk modulus ($K$), (b) Vickers hardness ($H_v$), (c) density ($\rho$), (d) Young’s modulus ($E$), (e) shear modulus ($G$), (f) refractive index ($n_d$).

elements in controlling $H_v$. As shown in figure 3(b), most of the ChGs generally have lower hardness values, which is in agreement with [46][44], which indicates that Ge–Se-based glasses have lower $H_v$ values. Figure 7(c) shows elements like S, Se, and Tl interacting with most elements, establishing their significant contribution to the $\rho$. Figure 7(d) indicates that Se, Fe, and As are the elements with higher interaction values with other elements. Se–Fe and S–As, are the two vital combinations of elements governing the $E$ values. Figure 7(e) indicates Cu, Ni, Se, and Ge with Cu–Ni, and Ge–Se are the common pairs with the highest interaction values for $G$. Figure 7(f) indicates that S, Se, Te, and As interact with most of the elements and has the highest interaction value for $n_d$.

3.9. Glass selection chart (GSC)

Due to functional requirements, multiple properties are required in a single glass composition. However, due to the contrasting effect of elements on desired properties, e.g. the higher concentration of Se increases the TEC but reduces the $T_g$, glass designers need to select chemical constituents to achieve targeted properties. To this extent, the GSCs (also known as Ashby plots [47, 48] for glasses) can be developed using the ML models. Note that preparation of such GSCs based on experimental data is challenging since all the properties need to be characterized on a wide family of glass compositions. Thus, the ML models can be used as surrogates develop a wide range of GSCs which can in turn, accelerate glass discovery. We only consider two-dimensional GSCs due to their ease of visualization. However, multidimensional GSCs can be prepared to screen ChGs for tailored properties.

For designing optical fibers made of ChGs, components that provide the desired TEC and $H_v$ should be selected to facilitate the formation of core and clad, simultaneously reaching desired strength [49]; the GSC presented in figure 8(a) can be used for such applications. Figure 8(b) shows the glass families having high $n_d$ and $H_v$ up to 4.5 GPa. These glasses are used for temperature sensing in nuclear reactors as their $n_d$ changes with phase-change from amorphous to crystalline and require thermal and mechanical stability. Such phase change occurs in nuclear power plants associated with temperature rise due to radiations in the reactors [50]. Hence figures 8(b) and (c) can be used together to have glasses with desired $n_d$, Vickers hardness, and TEC [51].

To manufacture glasses for high-speed communication fibers and non-linear optical devices, compositions with a wide range of glass transition temperatures dictating the ease of manufacturing of glass and $n_d$ [51, 52] can be obtained from the GSC (see figure 8(d)). Glass compositions for nuclear waste immobilization with enhanced $\rho$ and preferred $T_g$ can be selected using figure 8(e). ChGs for solid-state electrolytes in Li and Na ion batteries require good mechanical properties and $T_g$ depending upon the range of operational temperatures; hence, figure 8(f) can guide the composition selection [53]. The $T_g$ of the
material governs its formability, and the $E$ of the materials indicates its strength. The glasses belonging to As–Ge–S and As–Ge–Se families feature in both figures 8(g) and (h). Therefore, glasses for optical devices with required TEC and $E$ can be obtained using the GSCs presented in figures 8(g) and (h).
4. Conclusions

Altogether, we show that ML models can successfully predict the physical, mechanical, and optical properties of ChGs. Further, we show that the interpretable ML method can explain the properties of ChGs in terms of their constituents. Specifically, SHAP bar plots provide the mean absolute effect of each element. In contrast, the violin plots explain the effect of the elements with respect to their actual concentration present in the glass. Further, SHAP interaction value plots provide insights into how different glass constituents interact together and control the properties of the given ChGs. Therefore, the SHAP plots will help the researchers to accelerate the design ChGs with targeted properties by providing insights into the compositional control of the properties. Further, this work also reports the GSCs, which enable researchers to select glass families with desired properties. Overall, the insights gained from SHAP plots and GSCs coupled with predictive ML models will help researchers to design tailored ChGs.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

N M Anoop Krishnan acknowledges the financial support for this research provided by the Department of Science and Technology, India, under the INSPIRE faculty scheme (DST/INSPIRE/04/2016/002774) BRNS YSRA (53/20/01/2021–BRNS) and DST SERB Early Career Award (ECR/2018/002228) award by the Government of India. Sayam Singla acknowledge the funding received from undergraduate research award by IRD unit of IIT Delhi.

ORCID iDs

Sayam Singla https://orcid.org/0009-0000-5759-8611
Sajid Mannan https://orcid.org/0000-0002-7887-2250
Mohd Zaki https://orcid.org/0000-0002-4551-3470
N M Anoop Krishnan https://orcid.org/0000-0003-1500-4947

References

[1] Lezal D 2003 Chalcogenide glasses—survey and progress J. Optoelectron. Adv. Mater. 5 23–34
[2] Varshneya A K 2013 Fundamentals of Inorganic Glasses (Amsterdam: Elsevier)
[3] Kokorina V F 1996 Glasses for Infrared Optics vol 13 (Boca Raton, FL: CRC press)
[4] Wang R and Luther-Davies B 2014 Structural and physical properties of Ge_xAs_xSe_1−x glasses Amorphous Chalcogenides: Advances and Applications (New York: CRC Press) p 97
[5] Toupin P, Brillard L, Renversez G and Troles J 2013 All-solid all-chalcogenide microstructured optical fiber Opt. Express 21 14643–8
Hayashi A and Tatsumisago M 2012 Invited paper: recent development of bulk-type solid-state rechargeable lithium batteries with sulfide glass-ceramic electrolytes Electron. Mater. Lett. 8 199–207

Tichý I and Tichá H 2000 Remark on the glass-forming ability in Ge$_x$Se$_{1-x}$ and As$_x$Se$_{1-x}$ systems J. Non-Cryst. Solids 261 277–81

INTERGLAD Ver. 8 (available at: www.newglass.jp/interglad_n/gaiyo/info_e.html) (Accessed 14 October 2022)

SciGlass 2022 EPAM Systems (available at: https://github.com/epam/SciGlass) (Accessed 7 November 2022)

Bhattoo R, Bishnoi S, Zaki M and Krishnan N M A 2022 Understanding the compositional control on electrical, mechanical, optical, and physical properties of inorganic glasses with interpretable machine learning Acta Mater. 242 118439

Zaki M et al 2022 Interpreting the optical properties of oxide glasses with machine learning and Shapely additive explanations J. Am. Ceram. Soc. 105 1834–47

Bishnoi S, Singh S, Ravinder R, Bauchy M, Gosvami N N, Kodamana H and Krishnan N M A 2019 Predicting Young's modulus of oxide glasses with sparse datasets using machine learning J. Non-Cryst. Solids 524 119643

Bishnoi S, Ravinder R, Singh Grover H, Kodamana H and Anoop Krishnan N M 2021 Scalable Gaussian processes for predicting the optical, physical, thermal, and mechanical properties of inorganic glasses with large datasets Adv. Mater. 2 477–87

Ravinder R, Srividhara R K, Bishnoi S, Grover H S, Bauchy M, Kodamana H and Krishnan N M A 2020 Deep learning aided rational design of oxide glasses Mater. Horiz. 7 1819–27

Alkoçbaça E, Mastelini S M, Botari T, Pinheiro B A, Cassar D R, de Carvalho A C P L F and Zanotto E D 2020 Explainable machine learning algorithms for predicting glass transition temperatures Acta Mater. 188 92–100

Cassar D R 2021 ViscNet: neural network for predicting the fragility index and the temperature-dependency of viscosity Acta Mater. 206 116602

Cassar D R, Santos G G and Zanotto E D 2021 Designing optical glasses by machine learning coupled with a genetic algorithm Ceram. Int. 47 10555–64

Cassar D R, de Carvalho A C P L F and Zanotto E D 2018 Predicting glass transition temperatures using neural networks Acta Mater. 159 249–56

Zaki M, Jayadeva and Krishnan N M A 2021 Extracting processing and testing parameters from materials science literature for improved property prediction of glasses Chem. Eng. Process 180 106067

Ravinder, Venugopal V, Bishnoi S, Singh S, Zaki M, Grover H S, Bauchy M, Agarwal M and Krishnan N M A 2021 Artificial intelligence and machine learning in glass science and technology: 21 challenges for the 21st century Int. J. Appl. Glass Sci. 12 277–92

Dreyfus C and Dreyfus G 2003 A machine learning approach to the estimation of the liquidus temperature of glass-forming oxide blends J. Non-Cryst. Solids 318 63–78

Gupta T, Zaki M, Krishnan N M A and Mausam 2022 MatSciBERT: a materials domain language model for text mining and information extraction npj Comput. Mater. 8 102

Gupta T, Zaki M, Krishnan N M A and Mausam 2022 DISCoMaT: distantly supervised composition extraction from tables in materials science articles (arXiv:2007.01079)

Zaki M, Namireddy S R, Pittie T, Bihani V, Keshri S R, Venugopal V, Gosvami N N and Krishnan N M A 2022 Natural language processing-guided meta-analysis and structure factor database extraction from glass literature J. Non-Cryst. Solids 15 100103

Mauro J C, Tandia A, Vargheese K D, Mauro Y Z and Smedskjaer M M 2016 Accelerating the design of functional glasses through modeling Chem. Mater. 28 4267–77

Mastelini S M, Cassar D R, Alkoçbaça E, Botari T, de Carvalho A C P L F and Zanotto E D 2022 Machine learning unveils composition–property relationships in chalcogenide glasses Acta Mater. 240 118302

Landberg S M and Lee S-I 2017 A unified approach to interpreting model predictions J. Neural Inf. Process. (California, USA, 4–9 December) 30 4765–74

Mastelini S M, Cassar D R, Alkoçbaça E, Botari T, de Carvalho A C P L F and Zanotto E D 2021 Machine learning unveils composition–property relationships in chalcogenide glasses (arXiv:2106.07749)

Ramraj S, Uzir N, Sunil R and Banerjee S 2016 Experimenting XGBoost algorithm for prediction and classification of different datasets Int. J. Control. Theory Appl. 9 651–62

Akiba T, Sano S, Yanase T, Ohta T and Koyama M 2019 Optuna: a next-generation hyperparameter optimization framework (arXiv:1907.10902)

Bergstra J, Bardenet R, Bengio Y and Kégl B 2011 Algorithms for hyper-parameter optimization J. Neural Inf. Process. (Granada, Spain, 12-14 December) 24 2546–54

Hansen N and Ostermeier A 2001 Completely derandomized self-adaptation in evolution strategies Evol. Comput. 9 159–95

Shahriari B, Swersky K, Wang Z, Adams R P and De Freitas N 2015 Taking the human out of the loop: a review of Bayesian optimization Proc. IEEE 104 148–75

Jamieson K and Talwalkar A 2016 Non-stochastic best arm identification and hyperparameter optimization Artificial intelligence and statistics pp 240–8

Meredig B et al 2018 Can machine learning be the next high-temperature superconductor? Examining extrapolation performance for materials discovery Mol. Syst. Des. Eng. 3 819–25

Kong Z, Cui Y, Liu Z, Zhao Y, Hu M and Hu J 2020 Evaluating explorative prediction power of machine learning algorithms for materials discovery with k-fold forward cross-validation Comput. Mater. Sci. 171 102903

Shiryaev V S, Filatov A I, Karakina E V and Nezhdanov A V 2021 Structural peculiarities of Ge-rich Ga-Ge-Sb-Se chalcogenide glasses J. Solid State Chem. 303 122454

Thiruvikraman P K 2006 Rings, chains and planes: variation of T$_g$ with composition in chalcogenide glasses Bull. Mater. Sci. 29 371–4

Senapati U and Varshneya A K 1995 Configurational arrangements in chalcogenide glasses: a new perspective on Phillips’ constraint theory J. Non-Cryst. Solids 185 289–96

Voronova A E, Ananievh V A and Blinov L N 2001 Thermal expansion of melts and glasses in the As-Se system Glass Phys. Chem. 27 267–73

Samudrala K and Devarasetty S B 2018 Effect of the copper on thermo-mechanical and optical properties of S-Se-Cu chalcogenide glasses IOP Sci. Conf. Ser.: Mater. Sci. Eng. 330 012042

Cui J, Zhang H, Liu L, Xu Y, Xiao X, Li M, Ma W and Guo H 2022 The influence of different antimony content in Ga-As-Sb-Se chalcogenide glass system: modification of physical & spectroscopic properties and fiber forming ability Ceram. Int. 48 25756–63

Yang Y, Yang Z, Lucas P, Wang Y, Yang Z, Yang A, Zhang B and Tao H 2016 Composition dependence of physical and optical properties in Ge-As-S chalcogenide glasses J. Non-Cryst. Solids 440 36–42
[44] Cao Z, Dai S, Ding S, Wang M, Xu L, Liu C and Lin C 2021 Chalcogenide glass ceramics: a high-performing innovative infrared acousto-optic material J. Eur. Ceram. Soc. 41 7215–21

[45] Aio L G, Efimov A M and Kokorina V F 1978 Refractive index of chalcogenide glasses over a wide range of compositions J. Non-Cryst. Solids 27 299–307

[46] Guin J-P, Rouxel T, Sangleboeuf J-C, Melscoet I and Lucas J 2002 Hardness, toughness, and scratchability of germanium–selenium chalcogenide glasses J. Am. Ceram. Soc. 85 1545–52

[47] Ashby M F, Bréchet Y J M, Cebon D and Salvo I. 2004 Selection strategies for materials and processes Mater. Des. 25 51–67

[48] Ashby M F and Johnson K 2013 Materials and Design: The Art and Science of Material Selection in Product Design (Netherlands: Butterworth-Heinemann)

[49] Sanghera J S and Aggarwal I D 1999 Active and passive chalcogenide glass optical fibers for IR applications: a review J. Non-Cryst. Solids 256–257 6–16

[50] Simon A-A-A, Kadrager K, Badamchi B, Subbaraman H and Mitkova M 2019 Temperature sensing in nuclear facilities: application of the phase change effect of chalcogenide glasses Proc. Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies (Orlando, FL, USA) pp 9–14

[51] Zhu L, Yang D, Wang L, Zeng J, Zhang Q, Xie M, Zhang P and Dai S 2018 Optical and thermal stability of Ge–As–Se chalcogenide glasses for femtosecond laser writing Opt. Mater. 85 220–5

[52] Kumar P, Kaur J, Tripathi S K and Sharma I 2017 Effect of antimony (Sb) addition on the linear and non-linear optical properties of amorphous Ge–Te–Sb thin films Indian J. Phys. 91 1503–11

[53] Zhang Y, Jiao Q, Ma B, Lin C, Liu X and Dai S 2019 Structure and ionic conductivity of new Ga2S3–Sb2S3–NaI chalcogenide glass system Physica B 570 53–57