



Predicting glass properties by using physics- and chemistry-informed machine learning models

Yueh-Ting Shih, Yunfeng Shi, Liping Huang^{*}

Department of Materials Science and Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, United States

ARTICLE INFO

Keywords:

Machine learning
Oxide glass
Young's modulus
Shear modulus
Electrical resistivity

ABSTRACT

Physics- and chemistry-informed machine learning (ML) models were trained by using descriptors in the element physical and chemical properties domain, which include stoichiometric, elemental-property-based, valance orbital occupation, and ionicity features. Young's modulus, shear modulus and electrical resistivity (ρ) data for a group of oxide glasses were used to train artificial neural network (ANN), support vector machine (SVM), and random forest (RF) models. In comparison with experimental values, the ANN performs the best in predicting elastic moduli, whereas the RF is the best in predicting the temperature dependence of ρ in terms of the coefficient of determination (R^2) value. The benefits of the ML models using descriptors in the element physical and chemical properties domain were demonstrated by revealing the relationships between the predicted glass properties and their first and second important features through a grid search.

1. Introduction

The combination of machine learning (ML) with the domain knowledge of materials science can explore the hidden relationships among material properties that cannot be explained by physical or chemical theories, and accelerate the prediction of material properties and the development of new materials [1–5]. Recently, ML technique has been successfully applied in the field of glass. After training a suitable ML algorithm with a reliable dataset, it can be used to predict different glass properties [6–22]. For instance, by combining molecular dynamics (MD) simulation with ML, Hu et al. performed a rapid and low-cost prediction of density and elastic properties of a multicomponent glass system [6]. By training models with a large dataset collected by Corning Incorporated, Deng showed good performance in the prediction of density and elastic moduli using different ML algorithms [7]. Other glass properties, such as Vickers hardness (H_v) [10], glass transition temperature (T_g) [10–13], coefficient of thermal expansion (CTE) [10], viscosity [14,15], and chemical durability [16–18] were well predicted by ML models as well.

Unlike the conventional statistical model, e.g., linear regression, a ML model often has a rather complex structure with multiple hyperparameters [23,24]. The complexity of ML model makes it work like a black box, and it is hard to obtain the physical or chemical insights from the predicted results directly. To obtain ML models that are highly

informative, one strategy is to use descriptors that include physical and/or chemical information instead of just using chemical composition [25]. To this end, several alternative descriptors have been proposed. For instance, Hu et al. used the descriptors based on MD potential parameters [6]. However, the performance of this method depends on the availability and reliability of the force field parameters. Liu et al. incorporated the topological constrain theory to compensate the lack of structural information in ML, but it is still limited to systems with several elements such that the number of constraints can be calculated and used as one of the model descriptors [18]. In contrast to these descriptors, element-based descriptors used by Hwang [14] and Cassar [15] are more versatile. Given that elements are the basic components of materials, element-based descriptors convert glass chemical composition into corresponding element-based properties, and predict well the temperature dependence of viscosity [14,15]. However, the dominating descriptors and the relationships between them and the glass properties have not been revealed.

In this work, to fully explore the capability of element-based descriptors and to reveal the dominating descriptors for glass properties, we investigated Young's modulus (E), shear modulus (G), and temperature dependence of electrical resistivity (ρ), which are important for electronic applications such as amorphous solid electrolytes in lithium ion batteries. After the feature extraction and feature selection process, the importance of each descriptor was revealed. After training artificial

^{*} Corresponding author.

E-mail address: huangl5@rpi.edu (L. Huang).

<https://doi.org/10.1016/j.jnoncrysol.2022.121511>

Received 17 October 2021; Received in revised form 15 February 2022; Accepted 16 February 2022

Available online 23 February 2022

0022-3093/© 2022 Elsevier B.V. All rights reserved.

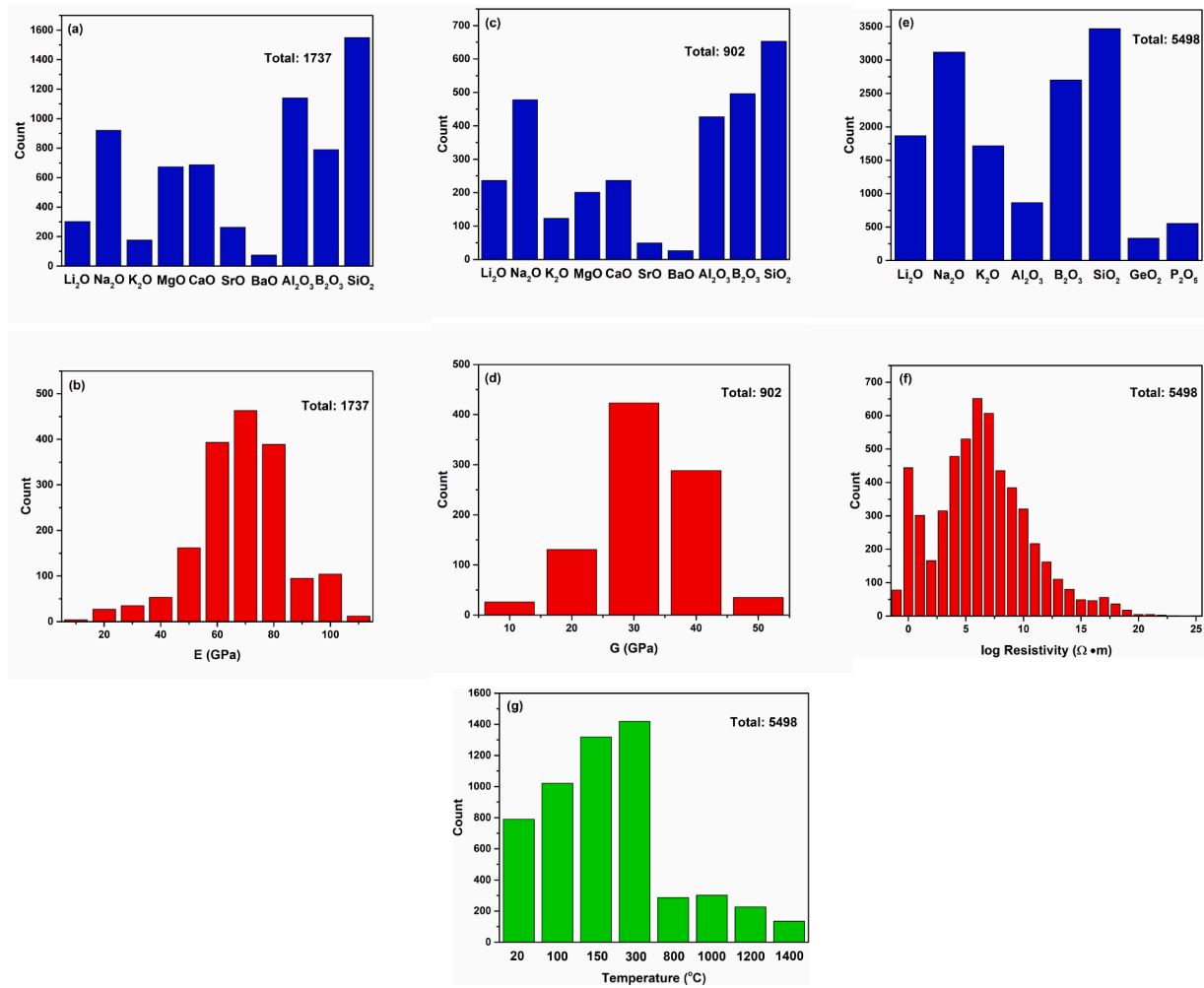


Fig. 1. (1) Young's modulus dataset distribution [(a) Composition and (b) value], (2) shear modulus dataset distribution [(c) composition and (d) value], and (3) electrical resistivity (in log-scale) dataset distribution [(e) composition, (f) value and (g) temperature].

Table 1

Elemental-property-based features included in this study.

Atomic number	Dipole polarizability [35]
Atomic volume	Density
Atomic weight [32]	Period
Boiling point	Group
Melting point	First ionization energy
Covalent radius [33]	Second ionization energy
Van der Waals radius [34]	Third ionization energy
Heat of formation [34]	

neural network (ANN), support vector machine (SVM), and random forest (RF) using element-based descriptors, the performance of each model was evaluated by the root mean square error (RMSE) and the coefficient of determination (R^2) value, and compared with models using the chemical composition descriptors. Furthermore, a grid search was carried out to illustrate the relationship between glass property and its dominating descriptors, which will facilitate the design of new glass composition with desired properties.

2. Methodology

ML models were trained to predict the Young's modulus, shear modulus and the temperature dependence of electrical resistivity by using the Scikit-Learn package [26]. The procedure includes data pre-processing in Section 2.1, feature extraction and selection in Section

2.2, and model training and evaluation in Section 2.3.

2.1. Data collection and cleaning

Raw data in this work was collected from the SciGlass database [27]. In order to obtain more reliable data, glass compositions were limited to those that had been widely studied. A glass system with a maximum of six components was investigated in this work, which can be noted as $\{M_i\}_{iMax=3}-\{F_j\}_{jMax=3}$, where M is the glass modifier and F is the glass former or intermediate. For elastic properties at ambient conditions, M includes Li, Na, K, Mg, Ca, Sr, Ba, and F includes Al, B, Si. For the electrical resistivity (in a log scale), M includes Li, Na, K, and F includes Al, B, Si, Ge, P. Elastic moduli at room temperature and electrical resistivity data at 20, 100, 150, 300, 800, 1000, 1200, and 1400 °C were used.

After data were extracted and rearranged from the SciGlass, the local outlier factor (LOF) algorithm was used to detect and remove the outliers in the raw dataset [28]. The algorithm estimates the local density of a given data point by the distance of k-nearest neighbors. If a point is an outlier, its local reachability density (LRD) is less than the average LRD of its neighbors, then the LOF value will be high. The criterion for removing an outlier is defined as $LOF > 1.1$ in this study. The distribution of data after removing the outliers is shown in Fig. 1.

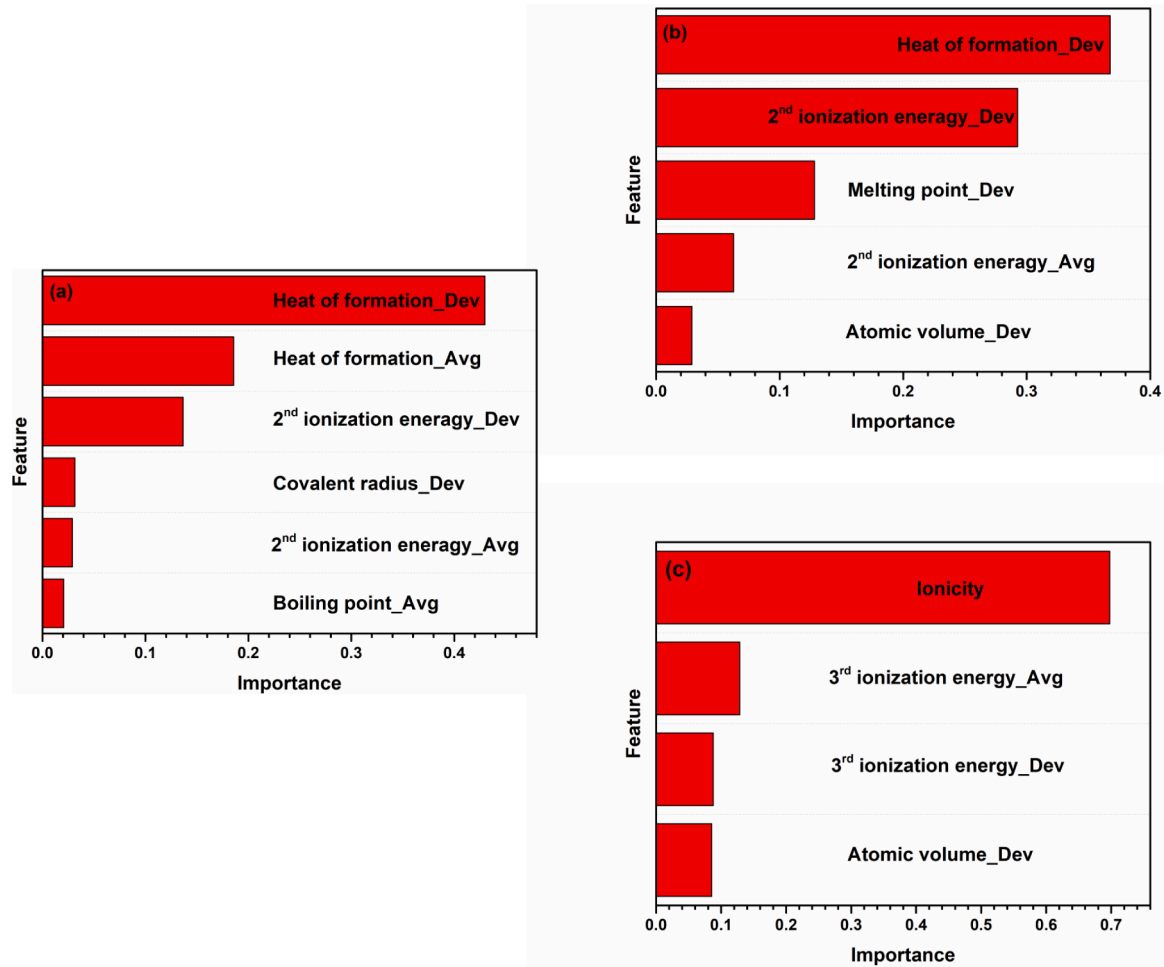


Fig. 2. Selected features as a function of importance for (a) Young's modulus, (b) shear modulus, and (c) electrical resistivity.

Table 2

Hyperparameters for each ML model using element chemical and physical properties descriptors (numbers in the bracket represent the start, end, and increment value of a grid search, respectively).

ML Models	Hyperparameters
ANN	Activation function: Logistic sigmoid function; Hyperbolic tan function; Identical function; Rectified linear unit function Hidden layer size: [1, 3, 1] Neurons in the hidden layer: [10, 200, 10]
SVM	Radial basis function kernel coefficient: [0.1, 1, 0.1] Regularization parameter: [100, 1000, 100] Tolerance parameter: [0.1, 1, 0.1]
RF	Number of estimators: [100, 1000, 100]

2.2. Feature extraction and selection

In the original data from the SciGlass, a glass is represented by the mole fraction of its constituents, i.e., by using features in the chemical composition domain. In order to develop physics- and chemistry-informed ML models, we extracted features in the element physical and chemical properties domain as proposed by Ward et al. [1,2]. These features can be classified into four categories:

2.2.1. Stoichiometric features

These features capture the stoichiometry of a glass. They are based on L^p norms of a vector representing the mole fraction (x_i) of each

Table 3

Optimized hyperparameters for each ML model using element chemical and physical properties descriptors and glass property.

	Optimized hyperparameters	Young's modulus	Shear modulus	Electrical resistivity
ANN	Activation function	Logistic sigmoid	Hyperbolic tan	Hyperbolic tan
	Hidden layer structure	(50)	(30, 30)	(100, 100)
SVM	Radial basis function kernel coefficient	0.5	0.5	0.5
	Regularization parameter	500	500	500
	Tolerance parameter	1	0.5	0.5
RF	Number of estimators	100	1000	500

Table 4

Training set performance of ML models with dominating descriptors for the prediction of the Young's modulus, shear modulus and electrical resistivity evaluated by R^2 (RMSE).

ML models	<i>Element chemical and physical properties descriptors</i>		
	Young's modulus	Shear modulus	Electrical resistivity
ANN	0.965 (3.100)	0.966 (1.382)	0.926 (1.119)
SVM	0.973 (2.738)	0.978 (1.118)	0.946 (0.984)
RF	0.989 (1.762)	0.822 (0.988)	0.987 (0.474)
	<i>Chemical composition descriptors</i>		
	Young's modulus	Shear modulus	Electrical resistivity
ANN	0.974 (2.663)	0.689 (4.189)	0.953 (0.893)
SVM	0.973 (2.721)	0.666 (4.343)	0.964 (0.782)
RF	0.990 (1.665)	0.989 (0.796)	0.988 (0.460)

Table 5

Testing set performance of ML models with dominating descriptors for the prediction of the Young's modulus, shear modulus and electrical resistivity evaluated by R^2 (RMSE).

<i>Element chemical and physical properties descriptors</i>			
ML models	Young's modulus	Shear modulus	Electrical resistivity
ANN	0.944 (3.845)	0.861 (2.917)	0.938 (1.040)
SVM	0.938 (4.070)	0.862 (2.911)	0.943 (0.997)
RF	0.914 (4.788)	0.698 (4.306)	0.948 (0.949)
<i>Chemical composition descriptors</i>			
ANN	0.955 (3.460)	0.472 (5.693)	0.954 (0.897)
SVM	0.953 (3.529)	0.477 (5.667)	0.961 (0.823)
RF	0.925 (4.458)	0.856 (2.978)	0.954 (0.900)

element i in the glass as shown in Eq. (1).

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p} \quad (1)$$

where n is the number of elements in the glass. In this work, we used $p = 2, 3, 5$, and 7 norms [11], which were selected by the criterion that the relative difference between $\|x\|_p$ and $\|x\|_{p+1}$ is higher than 1%.

2.2.2. Elemental-property-based features

The elemental-property-based features included in this study are

listed in Table 1, which are available in the Python modules mendelev (<https://github.com/lmmentel/mendelev>) and matminer [29]. Fraction-weighted mean ($\bar{f}$, Avg) and average deviation ($\hat{f}$, Dev) of each feature (f_i) are calculated by using Eqs. (2) and (3), respectively.

$$\bar{f} = \sum x_i f_i \quad (2)$$

$$\hat{f} = \sum x_i |f_i - \bar{f}| \quad (3)$$

2.2.3. Valance orbital occupation features

These features are the fraction-weighted average of the number of valance electrons in each orbital divided by the fraction-weighted average of the total number of valance electrons. Taking $0.2\text{Na}_2\text{O}-0.8\text{SiO}_2$ as an example, the fraction-weighted average of the number of valance electrons in the s-orbital is calculated by Eq. (4), where $2/15$, $4/15$, and $3/5$ are the mole fraction of Na, Si, and O, respectively.

$$2/15 \times 1 + 4/15 \times 2 + 3/5 \times 2 = 1.87 \quad (4)$$

Similarly, the fraction-weighted average of the number of valance electrons in the p-orbital can be calculated by Eq. (5).

$$2/15 \times 0 + 4/15 \times 2 + 3/5 \times 4 = 2.93 \quad (5)$$

Thus, the valance orbital occupation feature of s-orbital and p-orbital are $\frac{1.87}{(1.87+2.93)}$ and $\frac{2.93}{(1.87+2.93)}$, respectively.

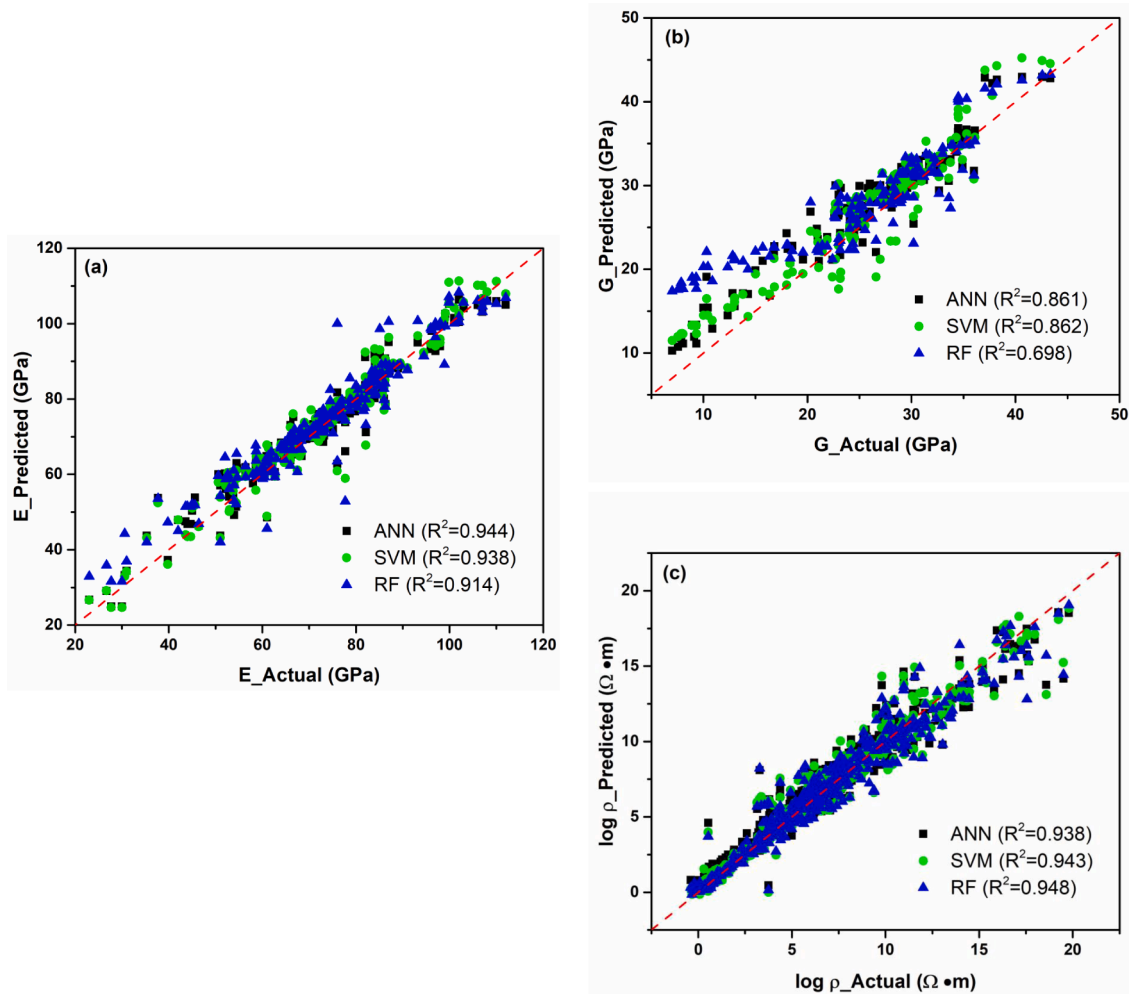


Fig. 3. Experimental values of (a) Young's modulus, (b) shear modulus, and (c) electrical resistivity in comparison with those predicted by artificial neural network (ANN), support vector machine (SVM), and random forest (RF) models in testing dataset by using descriptors in the element physical and chemical properties domain. The red dashed line represents the trend that the predicted values are equal to the experimental values ($R^2=1$).

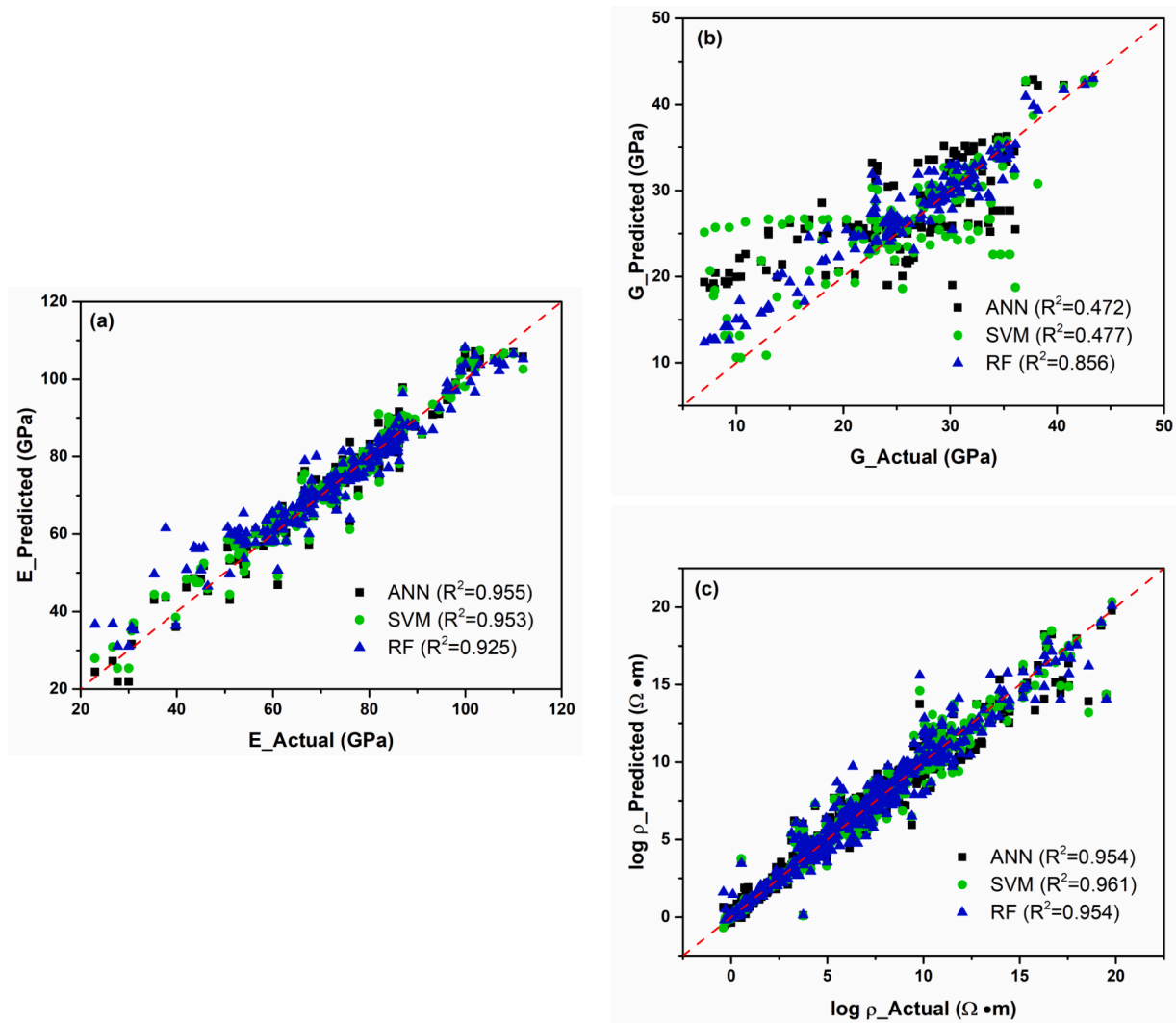


Fig. 4. Experimental values of (a) Young's modulus, (b) shear modulus, and (c) electrical resistivity in comparison with those predicted by artificial neural network (ANN), support vector machine (SVM), and random forest (RF) models in testing dataset by using descriptors in the chemical composition domain. The red dashed line represents the trend that the predicted values are equal to the experimental values ($R^2=1$).

2.2.4. Ionicity feature [30]

The ionic character (I) is shown in Eq. (6), where X_i and X_O represent the electronegativity of a constituent element and oxygen, respectively.

$$I(X_i, X_O) = 1 - \exp(-0.25(X_i - X_O)^2) \quad (6)$$

The ionicity feature computes the mean ionic character, which is calculated by using Eq. (7).

$$\bar{I} = \sum x_i x_O \times I(X_i, X_O) \quad (7)$$

Once again, using $0.2\text{Na}_2\text{O}-0.8\text{SiO}_2$ as an example, the electronegativity of Na, Si, and O are 0.93, 1.9, and 3.44, respectively, the ionic character between Na and O is 0.79, and between Si and O is 0.45, the ionicity feature is $2/15 \times 3/5 \times 0.79 + 4/15 \times 3/5 \times 0.45 = 0.135$.

After the feature extraction, the resulting features were selected by the importance factor calculated by the random forest algorithm [31], which was trained using the same procedure as in Section 2.3. It is important to note that since we studied the electrical resistivity at different temperatures, temperature was considered as one of the features in the ML models. Only features in the element physical and chemical properties domain with importance higher than 2% will be considered in the following training process. The selected features in the

element physical and chemical properties domain and the corresponding importance are shown in Fig. 2.

2.3. Model training and evaluation

Data after pre-processing were randomly separated into training and testing sets with a ratio of 80:20. Each dataset was standardized separately by centering to the mean and scaling to unit variance before training and testing. Artificial neural network (ANN) [36], support vector machine (SVM) [37] and random forest (RF) [38] were trained separately. For ANN, Adma was used as an optimizer with a learning rate of 0.001 [39]. For RF, the minimum number of samples required to split an internal node was 2. Besides training ANN and SVM, we also re-trained RF with the dataset by using the selected features with importance higher than 2%.

A 5-fold cross-validation (CV) was conducted to optimize the hyperparameters of each model by using the training set. Hyperparameters considered in the CV were grid searched in a certain range as listed in Table 2. The optimized hyperparameters for each ML model and glass property are listed in Table 3. The performance of each model was evaluated by the root mean square error (RMSE) and the coefficient of determination (R^2) value, where $R^2=1$ indicates a perfect prediction; while the smaller the R^2 value, the worse a model's performance [40].

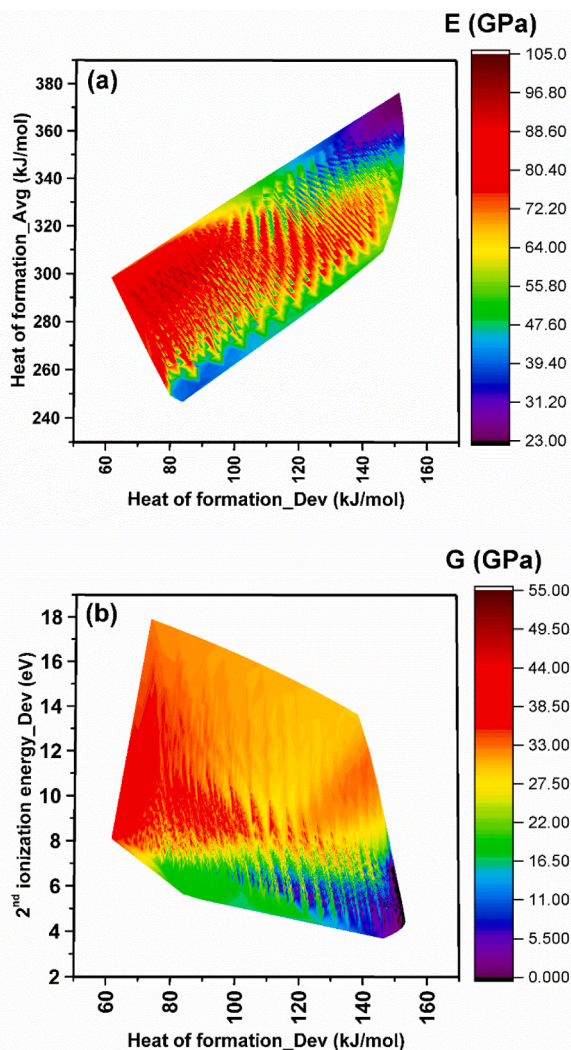


Fig. 5. Projection of (a) the Young's modulus and (b) shear modulus on the plane of the first (x-axis) and the second (y-axis) important feature. The trained ANN model is employed to predict the elastic moduli.

Meanwhile, ML models using chemical composition descriptors in terms of mole fraction of constituent oxides were also trained and evaluated by the same procedure for comparison. R^2 and RMSE of ML models (trained with dominating descriptors for each property) for the training and testing dataset are summarized in Table 4 and Table 5, respectively.

3. Results and discussions

Fig. 3 shows the performance of the trained ML models by using the descriptors in the element physical and chemical properties domain. The red dashed line represents the trend that the predicted values are equal to the experimental values ($R^2=1$), whereas a larger deviation from the line represents a worse prediction.

Among the three models in Fig. 3(a), ANN has the best performance for predicting the Young's modulus (E), followed by SVM and RF. Moreover, the best predicted Young's modulus lies in the range of 60–80 GPa. As seen in Fig. 1(b), the highest distribution of the Young's modulus is in the range of 60–80 GPa, which suggests that the performance of ML model becomes worse as the training data becomes sparse.

Among the three glass properties we studied in this work, the shear modulus has the sparsest dataset, which results in the worst performance of ML models as seen in Fig. 3(b). Among the three models, the performance of ANN is similar to that of SVM; however, the RF has a quite

weak performance in predicting the shear modulus. Fig. 3(c) shows the comparison of predicted and actual electrical resistivity (ρ) on a log-scale, where the accuracy of the prediction deteriorates with increasing resistivity. Once again, this may be due to the fact that the sparsity of data increases with increasing resistivity as shown in Fig. 1(f). The three models studied here well predict the temperature dependence of the electrical resistivity and RF exhibits a slightly better performance than the other two.

For comparison, Fig. 4 shows the performance of the trained ML models using the descriptors in the chemical composition domain. In general, the accuracy of the ML models using the descriptors in the chemical composition domain deteriorates with increasing sparsity of data, similar to the observations in Fig. 3. Comparing the performance of the models using different types of descriptors in terms of the RMSE and R^2 in Table 5, it can be found that these two types of descriptors exhibit insignificant differences in predicting the Young's modulus and electrical resistivity. However, for shear modulus, ANN and SVM perform better than RF when descriptors in the element physical and chemical properties domain are used, the opposite is true when the descriptors in the chemical composition domain are used.

To reveal the relationship between the glass property and the dominating physical and/or chemical features of constituent elements, a grid search was conducted using the ML model with the best performance, namely the ANN for elastic moduli and the RF for electrical resistivity. The composition interval of the grid search was set as 5 mol% in terms of constituent oxides. In addition, to follow the glass formation theory, the total amount of modifier and intermediate is ≤ 40 mol% [41]. The resulting composition grid was then converted into features in the element physical and chemical properties domain as described in Section 2.2, and the trained ML model was used to predict properties. Afterwards, the relationship between the glass property and its dominating physical and/or chemical features of constituent elements was visualized by projecting the glass property on the plane of the first (x-axis) and the second (y-axis) important features.

Fig. 5(a) shows the projection of the predicted Young's modulus on the feature plane of the first important feature, the average deviation of heat of formation (H_f), and the second important feature, the fraction-weighted mean of H_f . At a given average H_f , the Young's modulus increases (more reddish) with decreasing average deviation of H_f . Moreover, the highest Young's modulus is observed in the region of 260–320 KJ/mol of average H_f and decreases while the average H_f moves away from this region. The projection of the predicted shear modulus on the feature plane of the average deviation of H_f and the average deviation of the second ionization energy (2^{nd} IE) is shown in Fig. 5(b). In the region where the 2^{nd} IE deviation that is higher than 7 eV, the shear modulus exhibits insignificant change with the features. Below this region, the shear modulus decreases with increasing average deviation of H_f and with decreasing the 2^{nd} IE deviation. The different second important feature in the Young's modulus and shear modulus may be because they represent the resistance of glass to different deformations (bond stretching vs. bond bending).

From the results of the grid search, it was found that low elastic moduli generally occur in glasses with high B_2O_3 and high alkali modifier contents, whereas the large additions of SiO_2 result in an enhancement in both the Young's modulus and shear modulus, which are consistent with experimental observations [42]. Comparing with the widely used Makishima–Mackenzie (MM) model in which the Young's modulus is proportional to the dissociation energy of the constituent oxide and the atomic packing fraction [43,44], the average H_f in the ML model corresponds to the dissociation energy in the MM model. Meanwhile, due to the large difference between the H_f of the modifier and the network former or intermediate [34], the glass with a higher H_f deviation implies higher modifier content, which generally leads to lower connectivity of glass network and thus lower elastic moduli. The H_f deviation learned from the ML model as the first important feature improves the prediction of the Young's modulus in the MM model [45,46].

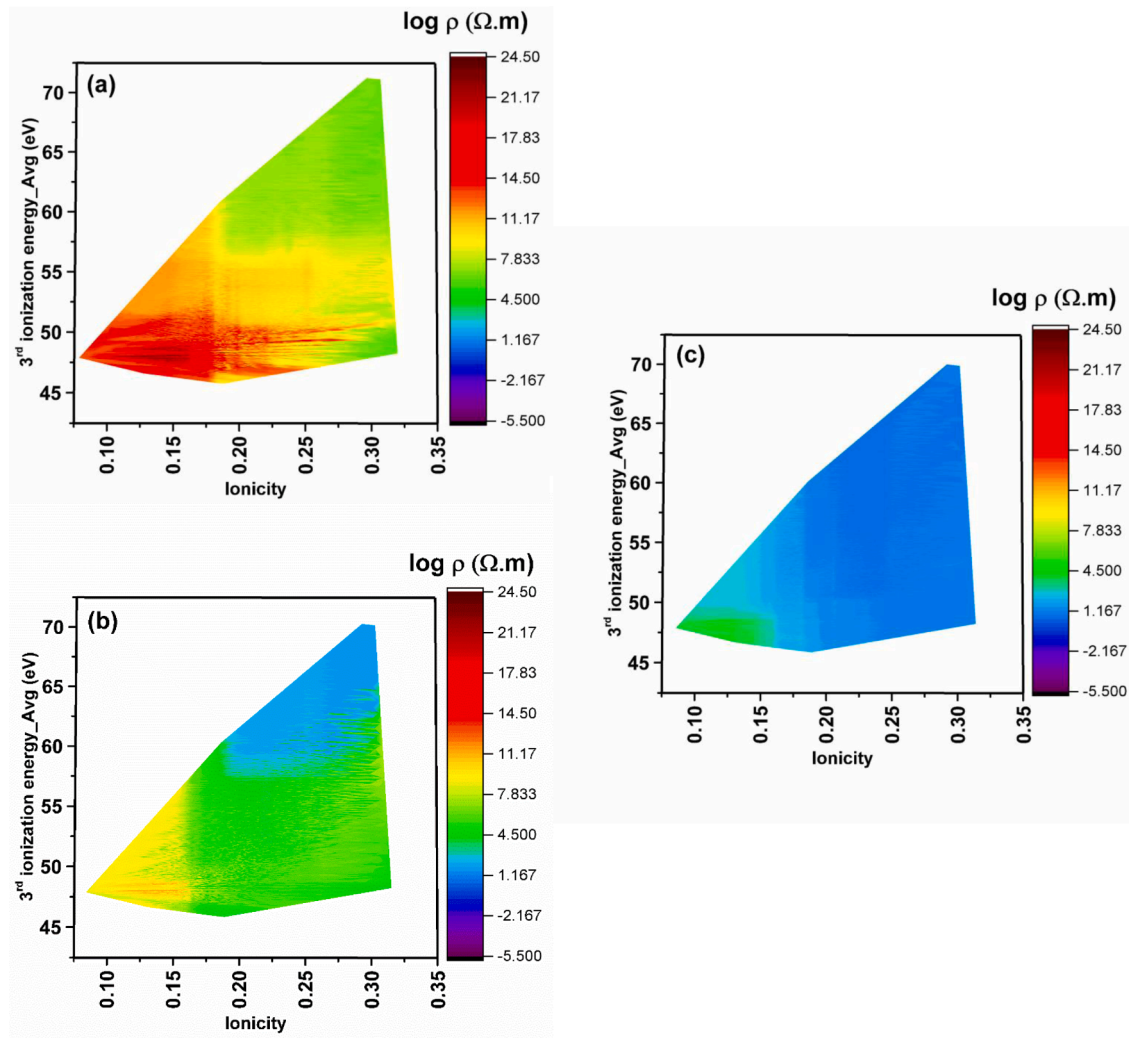


Fig. 6. Projection of electrical resistivity at (a) 20, (b) 300, and (c) 800 °C on the plane of the first (x-axis) and the second (y-axis) important feature. The trained RF model is employed to predict the properties.

Fig. 6 shows the projection of the predicted electrical resistivity on a log-scale at different temperatures on the feature plane of the average ionicity and the average third ionization energy (3rd IE). The importance of the 3rd order IE here may be due to elements with more valence electrons to ionize, such as phosphorus. For these elements, the first distinct change of IE occurs at the higher order and enhances the importance of the feature. The electrical resistivity decreases significantly with increasing temperature as the ion mobility increases. From the perspective of element physical and chemical features, the electrical resistivity decreases with both increasing the average ionicity and the 3rd IE.

From the results of the grid search, it was found that low electrical resistivity generally occurs in glasses with high Li₂O content, whereas the decreasing of modifier contents results in a significant increase in electrical resistivity, which are consistent with experimental observations [47]. Based on experimental results, the electrical resistivity of glass increases with the increasing bond strength of ions in the network and their size [48,49]. In the ML model, the covalency of the glass increases with decreasing ionicity, suggesting a stronger bond strength and a higher electrical resistivity. Moreover, the ionization energy increases with decreasing atomic size according to the periodic trends, which indicates that the smaller size of the alkali modifier ion, the higher ionization energy, hence a lower resistivity as predicted by the ML model.

In summary, descriptors in the element physical and chemical properties domain provide a new perspective for the ML models. Even though we did not include any glass structural information, such as coordination number, bond length and bond angle in our models directly due to the difficulty in obtaining reliable structural data in multi-component glasses, the models can still capture some structural information implicitly, such as the connectivity of the glass network and the ionic radius. Moreover, the important feature of the covalent radius in the Young's modulus, and the atomic volume in the shear modulus and the electrical resistivity identified in Fig. 2 also help compensate the lack of structural information in ML models. Comparing with models using descriptors in the chemical composition domain, ML models using features in the element physical and chemical properties domain can not only predict glass properties, but also help interpret them from the physical and chemical point of view, which provides key insights for the design of new glass composition with desired properties.

4. Conclusions

Physics- and chemistry-informed machine learning models were trained by using descriptors in the element physical and chemical properties domain instead of the chemical composition domain. The ANN exhibits the best performance in the prediction of elastic moduli, whereas the RF is the best in predicting the temperature dependence of

electrical resistivity (ρ) among the ML algorithms studied in this work. Through a grid search, the projection of predicted glass property on the plane of the first and the second important features reveals the relationship among them. At a given average heat of formation (H_f), the Young's modulus increases with decreasing average deviation of H_f . Moreover, the highest Young's modulus is observed in the region where the average H_f is of 260–320 KJ/mol, and decreases while the average H_f moves away from this region. The shear modulus exhibits insignificant change with the features in the region where the second ionization energy (2^{nd} IE) deviation is higher than 7 eV, and decreases with increasing average deviation of H_f and decreasing 2^{nd} IE deviation below the region. The electrical resistivity is strongly dependent on the temperature and exhibits a decreasing trend with both increasing ionicity and the third ionization energy (3^{rd} IE).

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. An input dataset of the Young's modulus and its trained model for Scikit-Learn can be accessed at <https://github.com/compmatricipi/PCIML-model>.

CRedit authorship contribution statement

Yueh-Ting Shih: Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **Yunfeng Shi:** Conceptualization, Formal analysis, Writing – review & editing. **Liping Huang:** Supervision, Funding acquisition, Conceptualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Science Foundation under Grant No. DMR-1508410 and DMR-1936368.

References

- [1] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, *NPJ Comput. Mater.* 2 (1) (2016) 16028.
- [2] B. Meredig, A. Agrawal, S. Kirklin, J.E. Saal, J.W. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, *Phys. Rev. B* 89 (9) (2014), 094104.
- [3] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, *NPJ Comput. Mater.* 3 (1) (2017) 54.
- [4] A. Agrawal, A. Choudhary, *APL Mater.* 4 (5) (2016), 053208.
- [5] J. Schmidt, M.R.G. Marques, S. Botti, M.A.L. Marques, *NPJ Comput. Mater.* 5 (1) (2019) 83.
- [6] Y.J. Hu, G. Zhao, M. Zhang, B. Bin, T. Del Rose, Q. Zhao, Q. Zu, Y. Chen, X. Sun, M. de Jong, L. Qi, *NPJ Comput. Mater.* 6 (1) (2020) 25.
- [7] B. Deng, *J. Non. Cryst. Solids* 529 (2020), 119768.
- [8] S. Bishnoi, S. Singh, R. Ravinder, M. Bauchy, N.N. Gosvami, H. Kodamana, N.M. A. Krishnan, *J. Non Cryst. Solids* 524 (2019), 119643.
- [9] K. Yang, X. Xu, B. Yang, B. Cook, H. Ramos, N.M.A. Krishnan, M.M. Smedskjaer, C. Hoover, M. Bauchy, *Sci. Rep.* 9 (1) (2019) 8739.
- [10] R. Ravinder, K.H. Sridhara, S. Bishnoi, H.S. Grover, M. Bauchy, H. Kodamana Jayadeva, N.M.A. Krishnan, *Mater. Horiz.* (2020).
- [11] E. Alcobaça, S.M. Mastelini, T. Botari, B.A. Pimentel, D.R. Cassar, A.C.P.d.L.F. de Carvalho, E.D. Zanotto, *Acta Mater.* 188 (2020) 92–100.
- [12] D.R. Cassar, A.C.P.L.F. de Carvalho, E.D. Zanotto, *Acta Mater.* 159 (2018) 249–256.
- [13] D.R. Cassar, G.G. Santos, E.D. Zanotto, *Ceram Int.* (2020).
- [14] J. Hwang, Y. Tanaka, S. Ishino, S. Watanabe, *Sci Technol. Adv. Mater.* 21 (1) (2020) 492–504.
- [15] D.R. Cassar, *Acta Mater.* 206 (2021), 116602.
- [16] Y. Zhang, A. Li, B. Deng, K.K. Hughes, *NPJ Mater. Degrad.* 4 (1) (2020) 14.
- [17] T. Han, N. Stone-Weiss, J. Huang, A. Goel, A. Kumar, *Acta Biomater.* 107 (2020) 286–298.
- [18] H. Liu, T. Zhang, N.M. Anoop Krishnan, M.M. Smedskjaer, J.V. Ryan, S. Gin, M. Bauchy, *NPJ Mater. Degrad.* 3 (1) (2019) 32.
- [19] J.D. Musgraves, J. Hu, L. Calvez, *Springer Handbook of Glass*, Springer International Publishing, 2019.
- [20] D.R. Cassar, S.M. Mastelini, T. Botari, E. Alcobaça, A.C.P.L.F. de Carvalho, E. D. Zanotto, *Ceram Int.* (2021).
- [21] H. Liu, Z. Fu, K. Yang, X. Xu, M. Bauchy, *J Non Cryst. Solids* 557 (2021), 119419.
- [22] V.V. Ravinder, S. Bishnoi, S. Singh, M. Zaki, H.S. Grover, M. Bauchy, M. Agarwal, N.M.A. Krishnan, *Int. J. Appl. Glass Sci.* 12 (3) (2021) 277–292.
- [23] D.W.T.H.R.T. Gareth James, *An Introduction to Statistical learning: With Applications in R*, Springer, New York, 2013 [2013]©2013.
- [24] T. Hastie, R. Tibshirani, J.H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2009.
- [25] I. Tanaka, *Nanoinformatics*, Springer, Singapore, 2018.
- [26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, G. Louppe, *J. Mach. Learn. Res.* 12 (2012).
- [27] A.I. Priven, O.V. Mazurin, *Adv. Mater. Res.* 39–40 (2008) 147–152.
- [28] M.M. Breunig, H.-P. Kriegel, R.T. Ng, J. Sander, *LOF: identifying density-based local outliers*, *SIGMOD Rec.* 29 (2000) 93–104.
- [29] L. Ward, A. Dunn, A. Faghaninia, N.E.R. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K.A. Persson, G. J. Snyder, I. Foster, A. Jain, *Comput. Mater. Sci.* 152 (2018) 60–69.
- [30] L. Pauling, *J. Am. Chem. Soc.* 54 (9) (1932) 3570–3582.
- [31] G. Louppe, L. Wehenkel, A. Suter, P. Geurts, in: *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1*, Lake Tahoe, Nevada, Curran Associates Inc., 2013, pp. 431–439.
- [32] J. Meija, T.B. Coplen, M. Berglund, W.A. Brand, P.D. Bièvre, M. Gröning, N. E. Holden, J. Irrgeher, R.D. Loss, T. Walczyk, T. Prohaska, *Pure Appl. Chem.* 88 (3) (2016) 265–291.
- [33] B. Cordero, V. Gómez, A.E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán, S. Alvarez, *Dalton. Trans.* (21) (2008) 2832–2838.
- [34] W.M. Haynes, *CRC Handbook of Chemistry and Physics*, CRC Press, 2014.
- [35] P. Schwerdtfeger, J.K. Nagle, *Mol Phys* 117 (9–12) (2019) 1200–1225.
- [36] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, The MIT Press, 2016.
- [37] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, B. Scholkopf, *IEEE Intell. Syst. Their Appl.* 13 (4) (1998) 18–28.
- [38] L. Breiman, *Mach. Learn.* 45 (1) (2001) 5–32.
- [39] D.P. Kingma, J. Ba, *CoRR* (2015) abs/1412.6980.
- [40] D.J. Ozer, *Psychol. Bull.* 97 (2) (1985) 307–315.
- [41] A.K. Varshneya, J.C. Mauro, *Fundamentals of Inorganic Glasses*, Elsevier Science, 2019.
- [42] K. Januchta, T. To, M.S. Bødker, T. Rouxel, M.M. Smedskjaer, *J. Am. Ceram. Soc.* 102 (8) (2019) 4520–4537.
- [43] A. Makishima, J.D. Mackenzie, *J. Non Cryst. Solids* 12 (1) (1973) 35–45.
- [44] A. Makishima, J.D. Mackenzie, *J. Non Cryst. Solids* 17 (2) (1975) 147–157.
- [45] M. Plucinski, J.W. Zwanziger, *J. Non Cryst. Solids* 429 (2015) 20–23.
- [46] Y. Shi, A. Tandia, B. Deng, S.R. Elliott, M. Bauchy, *Acta Mater.* 195 (2020) 252–262.
- [47] M.J. Lakin, H. Scholze, *Glass: Nature, Structure, and Properties*, Springer, New York, 2012.
- [48] W. Wang, R. Christensen, B. Curtis, S.W. Martin, J. Kieffer, *Phys. Chem. Chem. Phys.* 20 (3) (2018) 1629–1641.
- [49] F.M.E. Eldin, N.A. El Alaily, *Mater. Chem. Phys.* 52 (2) (1998) 175–179.