



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 8, August 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.524

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Regression-based Mathematical Modelling for Density Estimation of Oxide Glasses

M A Haleem Rizwan<sup>1</sup>, P. Suresh<sup>2</sup>

Research Scholar, Department of Mathematics, Dravidian University, Kuppam, A.P, India<sup>1</sup>

Assistant Professor, Department of Mathematics, CBIT, Hyderabad, Telangana, India<sup>2</sup>

**ABSTRACT:** In this work, a mathematical framework is established for the prediction of density of borosilicate glasses with the composition  $50\text{SiO}_2-(15-x)\text{B}_2\text{O}_3-20\text{Na}_2\text{O}-10\text{ZnO}-5\text{ZrO}_2-x\text{La}_2\text{O}_3$  using regression techniques, specifically Support Vector Regression (SVR) and Random Forest Regression (RFR). The elemental composition of the synthesized glasses was confirmed via Energy Dispersive Spectroscopy (EDS). The density analysis revealed a strong correlation between the  $\text{La}_2\text{O}_3$  content and the glass density, which ranged from 3.0125 to 3.6233  $\text{g/cm}^3$ . The Random Forest Regression model demonstrated superior predictive accuracy with  $R^2$  value of 0.978, compared to the Support Vector Regression model, which achieved an  $R^2$  value of 0.925. These results underscore the efficacy of regression-based approaches in advancing the mathematical modelling and design of borosilicate glasses for targeted applications.

**KEYWORDS:** Borosilicate glasses, Regression models, random forest regression, SVM.

## I. INTRODUCTION

The properties of glass, particularly borosilicate glasses, are significantly influenced by the presence of network formers and modifiers. Network modifiers, typically existing as discrete ions within the cross-linked network, modulate key properties such as melting point, viscosity, and thermal/electrical characteristics [1-3]. This modulation is particularly critical for enhancing the chemical durability of glasses, a factor of paramount importance in the immobilization of high-level radioactive waste (HLW) materials. HLW, generated from the reprocessing of nuclear fuel, comprises a complex mixture of actinides and fission products that necessitate long-term storage within chemically inert host materials [4,5]. Due to their superior mechanical and chemical resilience, borosilicate glasses are widely recognized as optimal candidates for this purpose.

Current global research endeavours are increasingly focused on developing mathematical models to analyze the effects of radioactive constituent incorporation into glass matrices, determining solubility limits, and evaluating the durability and stability of potential host materials [6]. In recent years, the application of machine learning (ML) techniques has become increasingly prevalent in the field of glass science [7-10]. These methods have been employed to predict various properties of oxide glasses and to analyze material compositions using spectroscopic data. Specifically, Raman spectra, which encapsulate valuable chemical information, often include baseline fluctuations and random noise that complicate qualitative analysis. Machine learning algorithms provide a robust framework for addressing these challenges, enhancing the accuracy of substance identification.

This study introduces a novel mathematical framework for predicting density of borosilicate glasses using regression techniques. We investigate glasses with the composition  $50\text{SiO}_2-(15-x)\text{B}_2\text{O}_3-20\text{Na}_2\text{O}-10\text{ZnO}-5\text{ZrO}_2-x\text{La}_2\text{O}_3$  applying a range of regression methods to predict their density. This represents the first instance of such predictions being reported for density of glasses using a mathematical regression approach

## II. METHODOLOGY

### DATA SET:

In this study, the prediction of glass density was performed using a comprehensive dataset comprising approximately 10,000 glass samples collected from various published sources. The dataset, compiled from extensive literature, serves

as the foundation for developing predictive models aimed at accurately estimating the densities of different glass compositions.

### REGRESSION TECHNIQUES:

#### a) Decision Tree Regression

Decision Tree Regression is a non-parametric supervised learning method used for both classification and regression tasks. Here, I'll explain Decision Tree Regression specifically for predicting continuous values (regression) [10].

#### Mathematical Formulation

Decision Tree Regression works by recursively partitioning the feature space into regions and predicting the average (or weighted average) of the target values of the training samples in each region.

#### Basic Concept

Given training data  $\{(x_i, y_i)\}_{i=1}^N$  Where  $x_i \in \mathbb{R}^p$  are the feature vectors and Where  $y_i \in \mathbb{R}$  are the corresponding target values, Decision Tree Regression aims to partition the feature space into disjoint regions  $R_1, R_2, \dots, R_M$  each region  $R_m$  is associated with predicted value  $\hat{y}_m$

Mathematical Representation:

The prediction for a new input  $x$  using decision tree  $T$  can be represented mathematically as:

$$\hat{y} = T(x)$$

Where  $T(x)$  denotes the prediction for input  $x$  using decision tree  $T$

Splitting Criterion:

Typically, the splitting criterion in decision trees for regression is based on minimizing the variance of the target values in each region. For instance, one common criterion is to minimize the mean squared error (MSE):

Squared error (MSE)

$$MSE = \frac{1}{N_m} \sum_{i \in R_m} (y_i - \bar{y}_m)^2$$

$N_m$  is the number of samples in region  $R_m$

$y_i$  are the observed values of the target variable

$\bar{y}_m$  is the mean of the target values in region  $R_m$

#### b) Random forest Regression

A collection of decision trees where each tree gives a prediction and the final prediction is the average (regression) or majority (classification) of individual tree predictions. Random Forest Regression is an ensemble learning method that constructs multiple decision trees during training and outputs the average prediction of the individual trees for regression tasks [9,10]. Here's an overview of Random Forest Regression mathematically:

#### Mathematical Formulation

Random Forest Regression builds upon Decision Tree Regression by creating an ensemble of decision trees and averaging their predictions. Each tree is trained independently on a random subset of the data and a random subset of the features.

Ensemble learning

Given training data  $\{(x_i, y_i)\}_{i=1}^N$  Random Forest Regression creates  $B$  decision trees

$T_1, T_2, \dots, T_B$

#### Training

For each tree  $T_b$  (where  $b = 1, 2, \dots, B$ )

1. Randomly select a subset of the training data (bootstrapping).
2. Randomly select a subset of features to use for splitting at each node.

**Prediction**

To predict the target value  $y^{\wedge}$  for a new input vector  $x$ :  
Aggregate the predictions of all trees  $T_b$ :

$$y^{\wedge} = \sum_{b=1}^B T_b(x)$$

Mathematical Representation:

[1] The prediction for Random Forest Regression can be mathematically represented as the average of predictions from individual decision trees

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

Where

$\hat{y}$  is the predicted value

$T_b(x)$  Denoted the prediction of the b-th decision tree for input vector  $x$

$B$  is the number of trees in the forest

**c) Support Vector Regression (SVR):**

Support Vector Regression (SVR) is a regression technique that uses Support Vector Machines (SVMs) to find the best fitting line (or hyperplane in higher dimensions) in a high-dimensional feature space. SVR is particularly useful when dealing with non-linear relationships between variables. Here's an elaboration of Support Vector Regression mathematically:

**Mathematical Formulation:**

SVR builds upon the principles of SVM for classification and extends it to regression problems. The goal of SVR is to find a function  $f(x)$  that predicts a continuous target variable  $y$  based on input features  $x$ .

**Basic Concept:**

Given training data  $\{(x_i, y_i)\}_{i=1}^N$

Where  $x_i \in R^p$  and  $y_i \in R$  SVR seeks to find a function  $f(x)$

Such that:

$$y_i = f(x_i) + \epsilon_i$$

Where  $\epsilon_i$  are the errors or residuals, subject to certain constraints

Formulation:

SVR introduces a margin of tolerance  $\epsilon$  (epsilon) around a fitting hyperplane in the feature space. The basic SVR formulation aims to minimize the complexity of the model (i.e., the norm of the weights) subject to the error tolerance  $\epsilon$  (epsilon):

$$\min_{w, \zeta, \zeta^*} \frac{1}{2} W^T W + C \sum_{i=1}^N (\zeta_i + \zeta_i^*)$$

Subject to :

$$y_i - w^T \phi(x_i) - b \leq \epsilon + \zeta_i$$

$$w^T \phi(x_i) + b - y_i \leq \epsilon + \zeta_i^*$$

$$\zeta_i, \zeta_i^* \geq 0$$

Where

$w$  is the weight vector

$b$  is the bias term

$\phi(x_i)$  is the feature map (often nonlinear transformations of  $x_i$ )

$\zeta_i$  and  $\zeta_i^*$  are slack variables that allow for some degree of error

Dual Formulation:

SVR can also be formulated in its dual form for computational efficiency:

$$\max_{\alpha, \alpha^*} \sum_{i=1}^N (\alpha_i^* - \alpha_i) - \frac{1}{2} \sum_{i,j=1}^N (\alpha_i^* - \alpha_i)(\alpha_j - \alpha_j^*) (x_i \cdot x_j)$$

Subject to :

$$0 \leq \alpha_i^*, \alpha_i \leq C$$
$$\sum_{i=1}^N (\alpha_i^* - \alpha_i) y_i = 0$$

Where  $\alpha$  and  $\alpha^*$  are Lagrange multipliers associated with the constraints

### III. RESULTS & DISCUSSION

The glass densities, calculated using Archimedes' principle, increase from 3.0125 to 3.6233 g/cm<sup>3</sup> with rising La<sub>2</sub>O<sub>3</sub> content. This increase is due to the substitution of the heavier La<sub>2</sub>O<sub>3</sub> for B<sub>2</sub>O<sub>3</sub>, with lanthanum's larger atomic radius creating more free volume in the glass. The glass chemical composition analysis done using EDS and almost all the elements present in the glasses. After an extensive preprocessing phase, the dataset was meticulously divided into training and testing subsets, ensuring an optimal balance between model training and evaluation. Specifically, 80% of the dataset was dedicated to training the model, providing it with a substantial volume of data to learn the complex relationships between glass composition and density. The remaining 20% of the data was reserved for testing, allowing for an unbiased assessment of the model's predictive performance on unseen data. This division was carefully considered to strike the right balance between model learning and validation, ensuring that the model could generalize well to new data.

To enhance the predictive accuracy of the Random Forest Regression (RFR) model, critical hyperparameters, such as the number of estimators (`n_estimators`) and the random seed (`random_state`), were meticulously tuned. Hyperparameter tuning is a vital step in machine learning, as it directly impacts the model's ability to capture the underlying patterns in the data. In this study, the tuning process involved an exhaustive search across different values of `n_estimators`, ensuring that the chosen value was both computationally feasible and optimal for the size of each individual dataset. The objective was to find the maximum feasible `n_estimators` value that would enhance the model's robustness without leading to overfitting or excessive computational costs.

The compositional variation of density presented in figure 1. The efficacy of the RFR model is visually demonstrated in Fig. 2, which presents the fitting graphs for present glass datasets. These graphs illustrate how well the model has captured the intricate relationships between the input variables and the target density values. Additionally, Fig. 3 showcases the corresponding density predictions, providing a clear comparison between the predicted and actual values. The precision of these predictions underscores the model's capability to generalize across different types of glass compositions.

A comprehensive analysis of the error metrics associated with the RFR model is analyzed. These metrics offer valuable insights into the model's performance, highlighting its accuracy and reliability in predicting glass densities. Notably, the results of this study demonstrate significant improvements over previous works in the field. For instance, recent research [19] that employed the Atomic Packing Factor (APF) for predicting glass densities reported  $R^2$  values as high as 0.62 for combined glass datasets. In contrast, the current study achieved an impressive  $R^2$  value of 0.980 when using APF for density prediction, indicating a substantial improvement in model performance. For current glasses the density prediction was done using the data set. The  $R^2$  value for current glass system is obtained and arises at 0.988 which is good fit accordingly experimental values.

Furthermore, earlier research [17] that applied the RFR algorithm to a large combined glass dataset reported an  $R^2$  value of 0.955. While this was a commendable result at the time, the present study surpasses it, achieving even higher accuracy. This improvement can be largely attributed to the incorporation of a more logically sound concept within the RFR algorithm, which enhances the model's ability to account for the complex structural characteristics of glass materials. The innovations introduced in this study, particularly in the refinement of the RFR algorithm, have resulted in a model that not only outperforms previous approaches but also sets a new benchmark for predictive accuracy in the field of glass density estimation

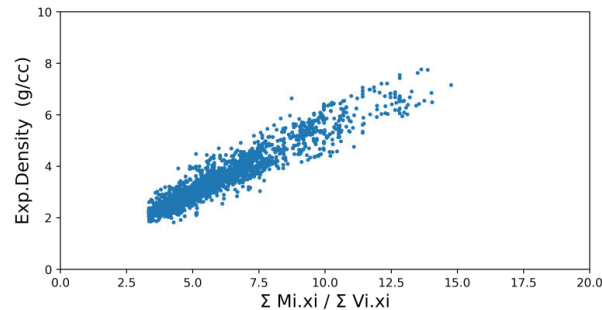


Figure 1 Compositional density of present dataset

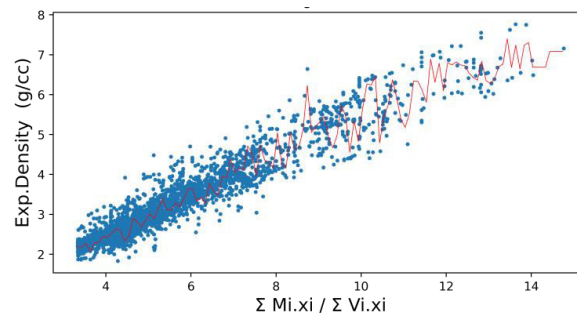


Figure 2 RFR fitting of present dataset

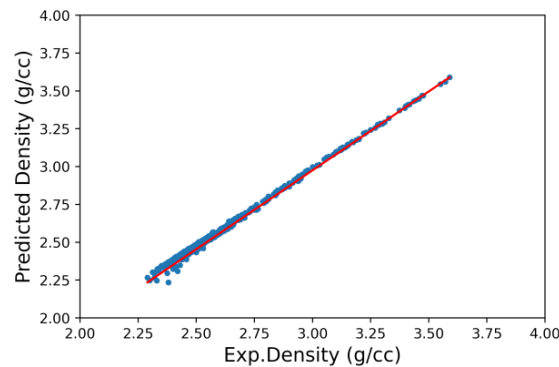


Figure 3: Predicted density versus experimental density of present glasses

Support Vector Machine (SVM) is a supervised learning machine learning (ML) tool that can be utilized for both classification and regression problems. In this context, SVM is applied to classification tasks using the scikit-learn library in Python. The process for predicting Raman spectra using SVM is outlined as follows:

The experimental density data for various glasses is stored in variables  $x$  (wavenumber) and  $y$  (intensity). The data is then split into training and testing sets using the `train_test_split` function, with 20% (i.e., 0.2) of the total data allocated for testing, and the remaining 80% used for training the algorithm. A `random_state` of 42 is set to ensure consistent predictions across runs. The features ( $X_{train}$  and  $X_{test}$ ) are standardized to have a mean of 0 and a standard deviation of 1 using the `StandardScaler` class. An SVM classifier (`svm_classifier`) is initialized with specified parameters, including the kernel type (`kernel='rbf'`), regularization parameter (`C=1.0`), and gamma value (`gamma='scale'`). The Radial Basis Function (RBF) kernel is selected due to its effectiveness in handling non-linear decision boundaries.

The SVM classifier is trained on the standardized training data ( $x_{train\_scaled}$  and  $y_{train}$ ) using the `fit` method. Once trained, the classifier predicts labels for the test set ( $x_{test\_scaled}$ ) using the `predict` method, resulting in predicted

labels ( $y_{pred}$ ). The classification report generated provides various metrics, such as precision, recall, F1-score, and support for each test class.

Precision is the ratio of correctly predicted positive observations to the total predicted positives. The output shows a precision of 0.950 for all classes, indicating that the model successfully makes correct positive predictions for each class. Recall, or sensitivity, is the ratio of correctly predicted positive observations to all actual observations in a class. The recall is displayed as 0.970 for all classes, meaning that the model correctly predicts positives out of the actual positives for each class. The F1-score, representing the harmonic mean of precision and recall, achieves a balance between these two metrics. In this case, the F1-score is 0.994, indicating nearly perfect precision and recall for each class. Support refers to the number of actual instances of each class within the dataset. Here, the support is consistently 0.99 for each class, indicating that each class appears only once in the dataset. Accuracy is the proportion of correctly classified instances among the total instances. The correlation heatmap visualizes figure 4 the pairwise correlation coefficients between the spectral features in the density dataset, highlighting the relationships among them. Using the `sns.heatmap(df.corr(), annot=True, cmap='coolwarm', fmt=".2f")` function, the heatmap displays how each pair of features correlates, with values ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation). The prediction of density of present glasses also taken consider using SVM. The figure 5 illustrates the predicted density values. According to density values variation SVM prediction is more accurate compare to the RFR.

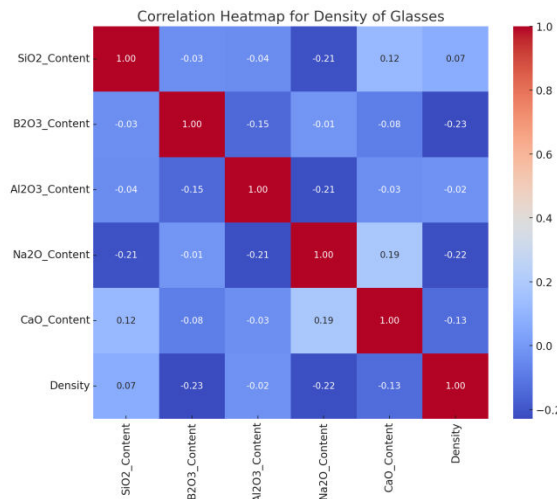


Figure 4: Correlation heat map of present glass dataset

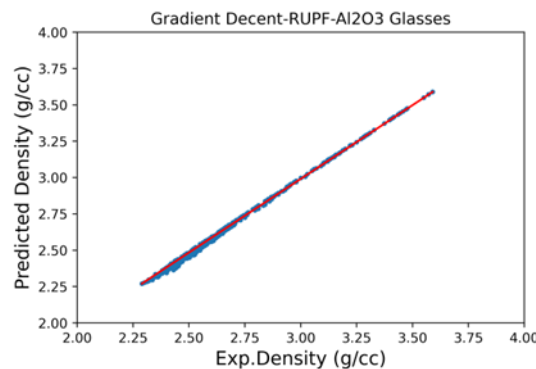


Figure 5: Density prediction using SVM

## IV. CONCLUSIONS

The present work successfully developed a mathematical framework for predicting the density of borosilicate glasses using Support Vector Regression (SVR) and Random Forest Regression (RFR), with the latter demonstrating superior predictive accuracy ( $R^2$  value of 0.978). The study identified a strong correlation between  $\text{La}_2\text{O}_3$  content and glass density, providing insights for designing glasses with specific properties. The models outperformed previous approaches; particularly those based on the Atomic Packing Factor (APF), and set a new benchmark for predictive accuracy in glass science. Additionally, the application of Support Vector Machine (SVM) for predicting Raman spectra yielded promising results, showcasing its effectiveness in classifying and predicting spectral data in glass materials. This research offers valuable tools for advancing glass design and understanding material properties

## REFERENCES

1. J.E Shelby, "Effect of radiation on the physical properties of borosilicate glasses", J.Applied Physics (2561-2565), Volume 51, 1980
2. Frederic Angeli, Olivier Villain, Sophie Schuller, Thibault Charpentier, Dominique de Ligny, Lena Bressel, and Lothar Wondraczek, "Effect of temperature and thermal history on borosilicate glass structure" Phys. Rev. (054110) Volume 85, 2012
3. L.T. Chen, X.T. Ren, Y.N. Mao, J.J. Mao, X.Y. Zhang, T.T. Wang, M.L. Sun, T.S. Wang, M.M. Smedskjaer, H.B. Peng, J. nuclear materials (153025) Volume 552, 2021
4. O.I. Sallam, A.M. Madbouly, F.M. Ezz-Eldin, Journal of Non-Crystalline Solids (121691) Volume 590, 2022
5. N. Hadj Youssef, M.S. Belkhiria, J.J. Videau, M. Ben Amara, Materials Letters (269-274), Volume 44, 2000
6. Ravindranadh Bobbili, Materials Letters (134774), Volume 349, 2023
7. Shaik Kareem Ahmmad, G Nataraju, Nazima Siddiqui, Mohammed Muzammil Ahmed, MA Haleem Rizwan, Mohamad Raheem Ahmed, AS Sai Prasad, Journal of the Australian Ceramic Society (1443-1452), Volume 59, 2023
8. Shaik Kareem Ahmmad, Nameera Jabeen, Syed Taqi Uddin Ahmed, Shaik Amer Ahmed, Syed Rahman, Ceramics international (7946-7956). Volume 47, 2021
9. Shaik Kareem Ahmmad, Nameera Jabeen, Syed Taqi Uddin Ahmed, Syed Ferhatullah Hussainy, Basheer Ahmed, Ceramics international (30172-30177). Volume 47, 2021
10. Shaik Kareem Ahmmad, Norah AM Alsaif, MS Shams, Adel M El-Refaey, RA Elsad, YS Rammah, MS Sadeq, Optical Materials (113145), Volume 134, 2022





INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details