Proceedings of CONF-CIAP 2026 Symposium: International Conference on Atomic Magnetometer and Applications
DOI: 10.54254/2753-8818/2026.HZ28754

Application of Machine Learning in Glass Type Classification

Ruiqiao Zhu

Guanghua Cambridge International School, Shanghai, China
leo@huashenglinen.com

Abstract. As an inorganic non-metallic material with a long history and wide application, glass has exerted a profound impact on the production, life and technological development process of human society since its birth. In order to investigate how the type of glass can be classified, this process is here to point out how to solve this kind of problem. This study addresses the inefficiencies of traditional glass classification methods by systematically evaluating four machine learning algorithms—Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM)—for accurate glass type identification. Leveraging glass's chemical and physical properties as features, the research incorporates feature engineering and hyperparameter tuning to optimize model performance. Experiments on a dataset of 214 glass samples show that optimized SVM achieved the highest test accuracy (0.7632), followed by tuned KNN (0.7368), RF (0.7105), and LR (0.6842). Cross-validation confirms model generalization, while RF's feature importance highlights magnesium and aluminum content as key classification drivers. This work provides a cost-effective, automated solution for glass classification, with value for both material science research and industrial applications like quality control and recycling.

Keywords: machine learning, glass type classification, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM)

1. Introduction

Glass plays a crucial role in modern industries, owing to its exceptional properties such as transparency and durability. Nevertheless, the increasing variety of glass types—including architectural, optical, and safety glass, among others—has considerably complicated the process of accurate classification. Traditional methods, which rely on expert judgment or involve laborious tests such as X-ray fluorescence spectroscopy, suffer from inefficiencies, subjectivity, and high costs. In contrast, the rapid advancement of artificial intelligence has established machine learning as a highly promising tool for achieving automated and accurate classification. While recent studies have highlighted its potential in material science, focused research on glass classification remains limited. Importantly, there is a significant lack of studies comparing different algorithms or optimizing feature selection in this area. This study seeks to fill this gap by systematically assessing the performance of Logistic Regression, Random Forest, KNN, and SVM in classifying types of glass. By utilizing a diverse dataset, it incorporates feature engineering and optimization techniques —such as hyperparameter tuning—to improve model performance. This research is of considerable

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.HZ28754

theoretical significance, as it enhances our understanding of how machine learning can be applied to material classification. In practical terms, it will offer cost-effective solutions for the glass industry, enhancing resource management and improving quality control processes.

2. Basic properties and classification system of glass

Glass is an amorphous solid created by cooling molten silica combined with various additives. The properties of glass are influenced by its chemical composition and manufacturing processes—critical factors for the development of machine learning-based classification models.

2.1. Chemical composition of glass

Silica (SiO_2) typically makes up 50–75% of most glass formulations. Additives like sodium oxide (Na_2O, which reduces the melting point), calcium oxide (CaO, which improves durability), and others (e.g., Al_2O_3, K_2O) alter its properties. For example, borosilicate glass, which is distinguished by its high boron oxide (B_2O_3) content, demonstrates outstanding thermal shock resistance, making it an ideal choice for laboratory equipment.

2.2. Physical properties of glass

The key physical characteristics of glass include transparency, varying density (e.g., higher in leaded glass), refractive index (a crucial parameter in optical applications), and thermal properties. Tempered glass, which is strengthened through heat treatment, exhibits superior impact resistance when compared to float glass, manufacturing processes (e.g., float glass vs. others), drawn glass, or its applications, such as architectural glass versus optical glass. However, these classification methods are largely dependent on expert judgment, which leads to inconsistencies.

3. The method of glass classification

3.1. Logistic Regression

Logistic Regression (LR) is a linear model for binary and multiclass classification, extending linear regression by applying the sigmoid function to predict class probabilities [1]. For multiclass tasks, it uses the "one-vs-rest" strategy, training a separate classifier for each class against the others.

Mathematically, the predicted probability of class y given features $\mathbf{x}$ is:$P(y=1|\mathbf{x}) = \frac{1}{1+e^{-\mathbf{w}\cdot\mathbf{x}+b}}$, where $\mathbf{w}$ is the weight vector and b is the bias term.

In the implementation, LR was trained with max_iter=10000 to ensure convergence, achieving an initial test accuracy of 0.6842 and cross-validation accuracy of 0.6543 ± 0.0612 [2]. Its simplicity and interpretability (via feature weights) make it a baseline for classification tasks.

3.2. Random Forest

Random Forest (RF) is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of their predictions [3]. It reduces overfitting by introducing randomness: (1) sampling training data with replacement (bootstrap samples) and (2) selecting a random subset of features for splitting at each node.

RF's strength lies in handling non-linear relationships and high-dimensional data without requiring feature scaling. The RF model achieved an initial test accuracy of 0.7105 and cross-

DOI: 10.54254/2753-8818/2026.HZ28754

validation accuracy of 0.6927 ± 0.0583. Feature importance analysis revealed that magnesium (Mg) and aluminum (Al) content were the most predictive of glass type, aligning with known chemical differences between glass categories [4].

3.3. K-Nearest Neighbors (KNN)

KNN is an instance-based learning algorithm that classifies a sample based on the majority class of its k nearest neighbors in the feature space [5]. Distance metrics (e.g., Euclidean distance) determine "nearness," making feature scaling critical for performance.

We optimized k using grid search (1–20), finding $k = 5$ as optimal. The tuned KNN achieved a test accuracy of 0.7368, up from 0.6579 with default parameters. Cross-validation showed stability: 0.6731 ± 0.0725. KNN's simplicity and adaptability to local patterns make it useful for small datasets like ours (214 samples) [6].

3.4. Support Vector Machine (SVM)

SVM finds a hyperplane that maximizes the margin between classes in the feature space [7]. For non-linear data, it uses kernel functions (e.g., RBF) to map features to a higher-dimensional space where linear separation is possible.

Our SVM was optimized by tuning C (regularization) and γ (kernel coefficient). The best parameters ($C = 10$, $\gamma = 0.1$) yielded a test accuracy of 0.7632, the highest among all models, with cross-validation accuracy of 0.7015 ± 0.0642. SVM's strength in handling high-dimensional spaces and avoiding overfitting made it effective for our 9-feature dataset [8].

4. Model performance evaluation metrics

4.1. Accuracy

Accuracy measures the proportion of correctly classified samples: $\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$

It provides a straightforward summary of performance but may be misleading for imbalanced datasets. In our case, the dataset was moderately balanced (e.g., 70 samples of class 2, 51 of class 1), making accuracy a valid primary metric [9]. The optimized SVM achieved the highest accuracy (0.7632), followed by Random Forest (0.7105)

4.2. Confusion matrix

A confusion matrix visualizes classification performance by comparing predicted vs. actual classes [10]. It reveals true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each category.

For the best model (optimized SVM), the confusion matrix showed strong performance in classifying building windows (class 1) and containers (class 5) but struggled with distinguishing vehicle windows (class 2 vs. 3), likely due to their similar chemical compositions [3]. This granular insight is critical for refining models for specific applications (e.g., forensic analysis).

DOI: 10.54254/2753-8818/2026.HZ28754

5. Advantages of machine learning in glass classification

Machine learning (ML) offers significant advantages for glass classification, surpassing traditional manual methods. It facilitates automated and efficient analysis: in contrast to the hours required for manual inspection of glass properties, such as chemical composition or refractive index, trained machine learning models can classify new samples in seconds, making them ideal for high-throughput environments like recycling facilities or forensic laboratories. Additionally, machine learning excels at identifying subtle and complex patterns, as glass types may differ by minute chemical variations (e.g., 1%–3% differences in oxide content) that traditional rule-based methods often overlook. Algorithms such as Random Forest and SVM are able to identify subtle nuances to differentiate between similar categories. Additionally, machine learning (ML) ensures robust generalization; cross-validation demonstrates that models perform reliably on unseen data, which is critical given the often-limited datasets in materials science. It also provides interpretable insights; for instance, the feature importance derived from Random Forest identifies crucial factors, such as magnesium and aluminum content, that drive classification, thus enhancing domain understanding beyond mere predictions.

6. Conclusion

This study systematically explores the application of machine learning in glass classification, addressing gaps in existing research by comparing multiple algorithms and integrating optimization techniques. The results demonstrate that machine learning outperforms traditional manual and rule-based methods, with distinct strengths across models: SVM achieved the highest test accuracy (0.7632) after tuning hyperparameters $C = 10$ and $\gamma = 0.1$, excelling at capturing complex feature relationships in high-dimensional space; Random Forest (0.7105) provided interpretable insights via feature importance, identifying magnesium and aluminum as critical drivers of classification; KNN improved significantly from 0.6579 to 0.7368 with optimal $k = 5$, showing adaptability to small datasets; and Logistic Regression (0.6842) served as a reliable baseline due to its simplicity. Cross-validation results further confirm the robustness of these models, ensuring reliable performance on unseen data—essential for material science's often limited datasets. Practically, this research offers the glass industry automated, cost-effective classification tools to enhance quality control, resource management, and high-throughput processing in recycling or forensic scenarios. The interpretable insights also deepen understanding of how chemical properties influence glass categorization. Future work could expand by incorporating more features (e.g., refractive index) or larger datasets to further improve accuracy, and testing advanced models to address lingering challenges—such as distinguishing chemically similar vehicle window classes (2 vs. 3). Overall, this study validates machine learning as a transformative tool for glass classification, bridging theoretical advancements with industrial utility.

References

[1] Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The Elements of Statistical Learning (2nd ed.). Springer. https: //doi.org/10.1007/978-0-387-84858-7
[2] Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830. https: //jmlr.org/papers/v12/pedregosa11a.html
[3] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https: //doi.org/10.1023/A: 1010933404324
[4] Islam, M. A., et al. (2020). Feature selection and classification of glass types using machine learning. Materials Today: Proceedings, 33, 4788-4793. https: //doi.org/10.1016/j.matpr.2020.07.641

DOI: 10.54254/2753-8818/2026.HZ28754

[5] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21-27. https: //doi.org/10.1109/TIT.1967.1053964
[6] Fix, E., & Hodges, J. L. (1951). Discriminatory analysis: Nonparametric discrimination: Consistency properties. USAF School of Aviation Medicine, Technical Report No. 4, pp. 1-32.
[7] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297. https: //doi.org/10.1007/BF00994018
[8] Schölkopf, B., & Smola, A. J. (2002). Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press.
[9] Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427-437. https: //doi.org/10.1016/j.ipm.2009.03.002
[10] Chicco, D., & Jurman, G. (2020). The confusion matrix: A review. arXiv preprint arXiv: 2005.10497. https: //arxiv.org/abs/2005.10497