Mineral Hardness and Specific Gravity Estimation Using Tab-Transformer Embeddings for Geological Analysis

Ch. Mounika[1]*, kadem Vamshi Krishna[1], Gajjala Ravinandhana[1], Ponnam Bharath Kumar[1]

[1]Department of Computer Science and Engineering (DS), Kommuri Pratap Reddy Institute of Technology, Ghanpur, Ghatkesar, 501301, Telangana, India.

*Correspondence: Ch. Mounika

To Cite this Article
Ch. Mounika, kadem Vamshi Krishna, Gajjala Ravinandhana, Ponnam Bharath Kumar, "Mineral Hardness and Specific Gravity Estimation Using Tab-Transformer Embeddings for Geological Analysis", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 04, April 2026,pp: 796-805, DOI: http://doi.org/10.64771/jsetms.2026.v03.i04.pp796-805
Submitted: 09-03-2026 Accepted: 16-04-2026 Published: 22-04-2026

ABSTRACT

Mineral hardness and density are important physical properties used in geology, mining, and material science for mineral identification and classification. Traditionally, mineral hardness was determined using the Mohs hardness scale through manual scratch testing, where minerals were compared against standard reference minerals. This traditional approach required laboratory testing, expert knowledge, and significant time to obtain accurate results. In many cases, the process was labor-intensive and sometimes produced inconsistent outcomes due to human errors and variations in experimental conditions. To address these limitations, the proposed system introduces a machine learning based multimodel regression approach for forecasting mineral hardness and specific gravity using chemical and physical features of minerals. The system is implemented as a web-based application using the Django framework. Multiple regression algorithms are applied to improve prediction performance, including CatBoost Regressor (CBR), AdaBoost Regressor (ABR), Random Forest Regressor (RFR), and Greedy Tab Transformer (GTT). These models analyze the relationships between mineral properties and target variables to accurately predict Mohs hardness and specific gravity. Experimental evaluation shows that the GTT model outperforms the other models in terms of prediction performance. The system also provides functionalities such as model training, algorithm comparison, and prediction modules for analyzing mineral properties. By integrating machine learning models with an interactive web platform, the proposed system offers a faster, more efficient, and scalable method for predicting mineral hardness compared to traditional manual testing methods.

Keywords: Mineral hardness, Specific gravity, Mineral identification, Mohs hardness scale, Physical properties, Chemical composition, Mineral classification, Django framework.

1. INTRODUCTION

Determining and characterizing rock types as shown in Fig. 1, is essential for every discipline and industry sector interested in earth sciences. The description of rock types is utilized at every stage, from the design studies for the operation of mine deposits to the fundamental processes applied for production. It is crucial and effective in determining the mining methods and equipment selection. The techniques to characterize the rock type are time-consuming and require unique expertise. In some cases, the determination of rock type can also be inaccurate. However, it is possible no determine the rock type from the rock parameters that have been analysed for the design of mining methods by various statistical and computational methods. Various statistical and computer science methods have been developed to perform deeper statistical analysis in recent years. Data mining is a scientific discipline that provides meaningful and usable information, especially from complex and large data.

Data mining extends beyond classical statistics and uses algorithms to simplify and reveal hidden patterns and relationships within a data set [1]. Researchers have studied extensively in the literature on predicting rock properties, especially those determined due to time-consuming and laborious experiments, using data mining techniques and classical statistical methods. Most studies focus on predicting rocks' elastic modulus and strength parameters. It is reported that data mining methods have been successfully used to solve various problems in rock mechanics and outperform traditional empirical, mathematical, or statistical methods [2-3]. However, despite this, studies are limited because they require a specialization.

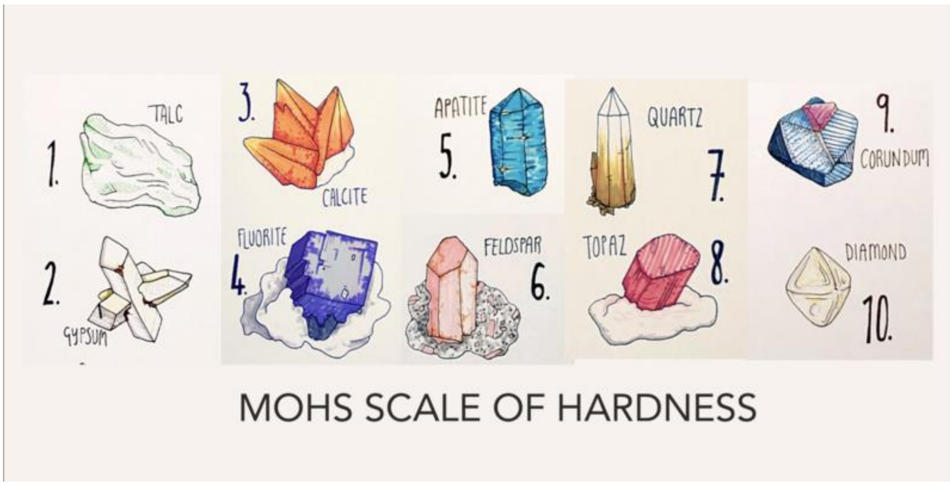


Fig. 1: Minerals of hardness

Martins and Miranda developed models for predicting the rock deformation modulus and the Rock Mass Rating (RMR) with data mining tools. Miranda et al. [4] developed new, simple, and reliable data mining models to predict geo-mechanical parameters such as friction angle, cohesion, and deformability modulus using a geotechnical database of an underground structure constructed in granite rock formation. Martins et al. [5] applied data mining techniques such as multiple regression, artificial neural networks, and support vector machines to predict the uniaxial compressive strength and deformation modulus of some Oporto granite rock characteristics. Ocak and Seker [6] conducted a study on the prediction of modulus of elasticity from some intact rock properties by artificial neural networks. Aboutaleb et al. [7] applied simple and multivariate regression, artificial neural network, and support vector regression to predict uniaxial compressive strength and static Young's modulus of five different intact limestone rock samples. They found that the support vector regression model was preferable and advantageous.

2. LITERATURE SURVEY

Zhang, et al. [8] proposed a data-driven model for hardness predictions of laser powder bed fusion products based on process parameters fused with power spectrum features of melt pool intensity data, which quickly and accurately predicts the microhardness of laser powder bed fusion specimens and can make constructive guidance for closed-loop feedback quality regulation in practical production. The effects of three integrated learning models, Random Forest, XG Boost and Light GBM, are also compared. The results indicate that random forest has the highest prediction accuracy in this dataset; however, it has the limitation of slow training and prediction speeds. The LightGBM algorithm has the fastest training and prediction speeds, about 1.4% and 4.4% of the random forest, respectively; however, the prediction accuracy is lower than that of random forest and XGBoost. XGBoost has the best overall comparative performance with adequate training and prediction speeds, about 23.7% and 37.9% of the random forest, respectively.

Guo, et al. [9] integrated the finite element (FE) and cellular automata (CA) approach to explore the distribution and variation of the grinding temperature of the workpiece surface in a grind-hardening process. Moreover, the simulation of the transformation process of "initial microstructure–austenite–martensite" for the workpiece helps determine the martensite fraction and then predict the hardness of the hardened layer with different grinding parameters. Finally, the effectiveness of the hardness prediction is confirmed by the grind-hardening experiment. Both the theoretical analysis and experiment results show that the variation in the grinding temperature will cause the formation to a certain depth of a hardened layer on the workpiece surface in the grind-hardening process.

Rodríguez-Rosales, et al. [10] evaluated the effect of temperature and time austempering on microstructural characteristics and hardness of ductile iron, validating the results by means of a statistical method for hardness prediction. Ductile iron was subjected to austenitization at 950 °C for 120 min and then to austempering heat treatment in a salt bath at temperatures of 290, 320, 350 and 380 °C for 30, 60, 90 and 120 min. By increasing austempering temperature, a higher content of carbon-rich austenite was obtained, and the morphology of the thin acicular ferrite needles produced at 290 °C turned completely feathery at 350 and 380 °C. A thickening of acicular ferrite needles was also observed as austempering time increased. An inversely proportional behavior of hardness values was thus obtained, which was validated through data analysis, statistical tools and a regression model taking temperature and time austempering as input variables and hardness as the output variable, which achieved a correlation among variables of about 97%.

Liu, et al. [11] proposed a machine-learning approach to predict extraterrestrial rock hardness using morphological features. A custom dataset of 1496 rock images, including granite, limestone, basalt, and sandstone, was created. Ten features, such as roundness, elongation, convexity, and Lab color values, were extracted for prediction. A foundational model combining Random Forest (RF) and Support Vector Regression (SVR) was trained through cross-validation. The output of this model was used as the input for a meta-model, undergoing linear fitting to predict Mohs hardness, forming the Meta-Random Forest and Support Vector Regression (MRFSVR) model. The model achieved an R^2 of 0.8219, an MSE of 0.2514, and a mean absolute error of 0.2431 during validation. Meteorite samples were used to validate the MRFSVR model's predictions. The model is used to predict the hardness distribution of extraterrestrial rocks using images from the Tianwen-1 Mars Rover Navigation and Terrain Camera (NaTeCam) and a simulated lunar rock dataset from an open-source website.

Bermanec, et al. [12] described the survey of the average Mohs hardness of minerals throughout Earth's history reveals a significant and systematic decrease from >6 in presolar grains to ~5 for Archean lithologies to <4 for Phanerozoic minerals. Two primary factors contribute to this temporal decrease in the average Mohs hardness. First, selective losses of softer minerals throughout billions of years of near-surface processing lead to preservational biases in the mineral record. Second, changes in the processes of mineral formation play a significant role because more ancient refractory stellar phases and primary igneous minerals of the Hadean/Archean Eon are intrinsically harder than more recently weathered products, especially following the Paleoproterozoic Great Oxidation Event and the production of Phanerozoic biominerals.

Xu, et al. [13] proposed a novel modelling approach that integrates mutual information (MI)-based parameter screening with random forest (RF) modelling to achieve an accurate quantitative prediction of surface hardness and impact energy in two martensitic stainless steels (1Cr13 and 2Cr13).

Preliminary analyses indicated that the magnetic parameters derived from Barkhausen noise (MBN), and the incremental permeability (IP) measurements showed limited linear correlations with the target properties (surface hardness and impact energy). To address this challenge, an MI feature screening method has been developed to identify both the linear and non-linear parameter dependencies that are critical for predicting target mechanical properties. The selected features were then fed into an RF model, which outperformed traditional multiple linear regression in handling the complex, non-monotonic relationships between magnetic signatures and mechanical performance. A key advantage of the proposed MI-RF framework lies in its robustness to small sample sizes, where it achieved high prediction accuracy (e.g., $R^2 > 0.97$ for hardness, and $R^2 > 0.86$ for impact energy) using limited experimental data. By leveraging MI's ability to capture multivariate dependencies and RF's ensemble learning power, it effectively mitigates overfitting and improves generalisation.

Ghadernejad, et al. [14] investigated sensor-based technologies, hyperspectral imaging, and portable X-ray fluorescence (pXRF) integrated with machine learning (ML) algorithms for characterizing rock hardness in open-pit gold mining contexts. A total of 159 rock samples from two Canadian open-pit gold mines were analysed through Leeb rebound hardness (LRH), short-wave infrared (SWIR) hyperspectral imaging, and a pXRF analyser for chemical characterization. The most critical spectral features of SWIR images were extracted using a novel and automated feature extraction approach and further refined by applying a recursive feature elimination (RFE) algorithm to reduce the dimensionality of the spectral feature space. Three ML algorithms, including Random Forest Regressor (RFR), Adaptive Boosting (AdaBoost), and Multivariate Linear Regression (MLR), were applied to develop predictive hardness models considering three scenarios: using chemical features, using refined spectral features, and their combination.

Cheng, et al. [15] developed, an artificial neural network (ANN) hardness model is based on a thermal–metallurgical model for mild steel. The objective is to establish non-linear relationships between the input process parameters and the desired output, i.e., hardness. The thermal–metallurgical model utilizes a well-distributed heat source model, a death-and-birth algorithm, and a metallurgical model to simulate the temperature field and to calculate the microstructure phase fraction. The temperature prediction errors at four thermocouple positions are mostly below 20%. Because of the limited experimental data, twenty-five simulation experiments are performed using the L25 orthogonal array based on the Taguchi method. The analysis of variance (ANOVA) reveals that the travel speed has the greatest impact on hardness. With the dataset from the thermal–metallurgical model, an ANN model to predict hardness is developed. A comparison to experimental data shows excellent performance and accuracy, with the Mean Absolute Percentage Error (MAPE) of ANN predictions within 10% of the targeted hardness.

3. PROPOSED SYSTEM

The proposed system focuses on developing a machine learning–based web application to forecast mineral hardness and specific gravity using chemical and physical features of minerals. The system integrates data analysis, model training, and prediction modules to provide accurate estimation of mineral properties as described in Fig. 2. By using multiple regression algorithms, the system analyzes relationships between mineral features and target variables to generate reliable predictions.

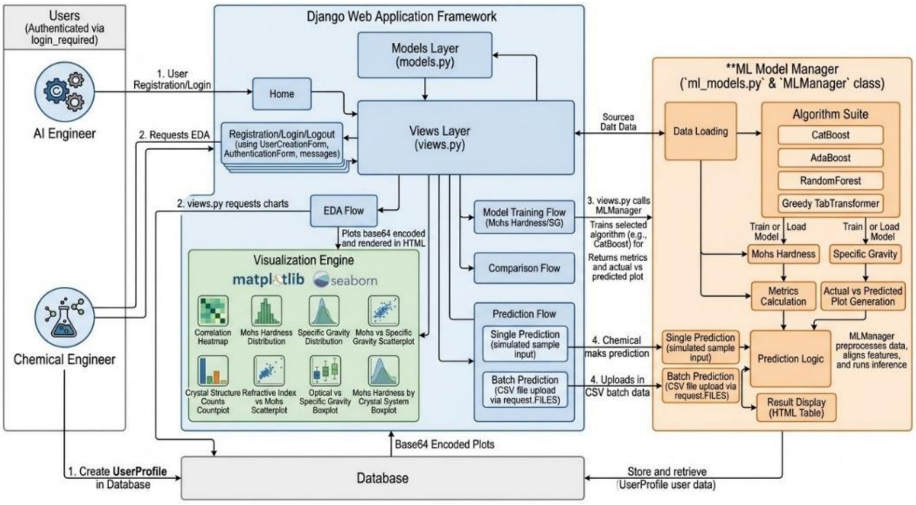


Fig. 2: Proposed system architecture.

Dataset Collection: The first step in the proposed system is collecting the mineral dataset containing various chemical and physical properties. The dataset includes important attributes such as refractive index, crystal structure, optical properties, iron content, specific gravity, and Mohs hardness. These features provide useful information for understanding mineral characteristics. The collected dataset serves as the primary source for training and evaluating the machine learning models.

Data Preprocessing: After collecting the dataset, preprocessing is performed to prepare the data for analysis and model training. This step involves organizing the dataset and selecting relevant features required for prediction. Data preprocessing helps ensure that the dataset is structured properly and suitable for machine learning algorithms. Proper preprocessing improves model performance and ensures reliable prediction results.

Exploratory Data Analysis (EDA): EDA is performed to understand the relationships between mineral features and their properties. Various visualization techniques are used to examine feature distributions and correlations within the dataset. These visualizations help identify patterns and important relationships between variables such as hardness and specific gravity. EDA plays a key role in gaining insights into the dataset before training machine learning models.

Feature Selection and Target Identification: In this step, the input features used for prediction are selected from the dataset. These features represent chemical and physical properties that influence mineral hardness and density. The target variables for prediction are defined as Mohs Hardness and Specific Gravity. Selecting appropriate features and targets is important for building accurate prediction models.

Model Training: Multiple machine learning regression algorithms are applied to train prediction models using the prepared dataset. The algorithms used in this research include CBR, ABR, RFR, and GTT. Each algorithm learns the relationships between mineral features and target variables during the training process. This step enables the models to generate predictions for mineral hardness and specific gravity.

Model Evaluation and Comparison: After training the models, their performance is evaluated to determine how accurately they predict the target variables. Each algorithm is tested and its prediction results are compared with actual values. The evaluation process helps identify the most effective model for mineral property prediction. Comparing multiple algorithms ensures that the best-performing model is selected for the prediction module.

Prediction Module: The prediction module allows users to estimate mineral hardness and specific gravity based on input features. The system supports both single prediction and batch prediction for multiple mineral samples. The trained model processes the input data and generates predicted values for the target variables. This module enables users to quickly analyze mineral properties using the developed system.

Result Visualization: The final step involves presenting the prediction results and model performance in a clear and understandable format. Graphical visualizations and tables are used to display the predicted values and comparison results. These visual outputs help users easily interpret the model predictions and analysis outcomes. Visualization improves the overall usability of the system.

4. RESULTS ANALYSIS

This section presents the visual outputs and interface screens generated during the execution of the mineral property prediction system. The developed application provides a structured web interface that enables users to interact with the system for dataset analysis and prediction tasks. The interface supports user authentication, data processing, and model-based prediction of Mohs hardness and specific gravity using the implemented ML models CBR, ABR, RFR, and GTT.

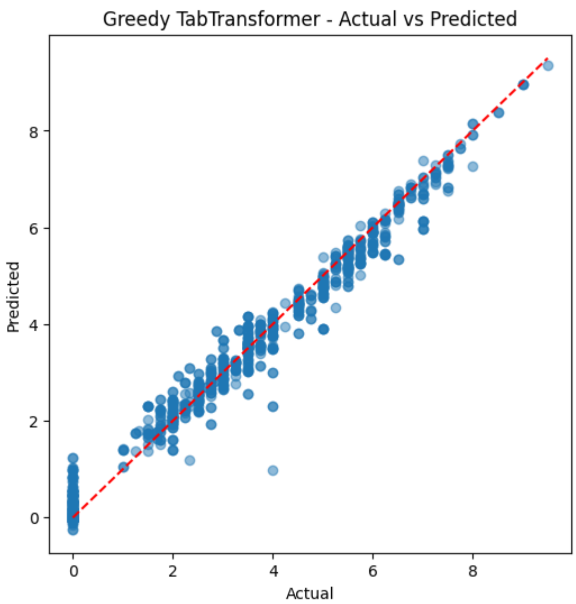


Fig. 3: Scatter plots of mohs hardness target attribute of various regressor models.

Fig. 3 GTT**:** The scatter plot displays the performance of the GTT. The points closely follow the diagonal line, demonstrating that GTT achieves the highest alignment between actual and predicted hardness values, reflecting its advanced feature representation and model efficiency.

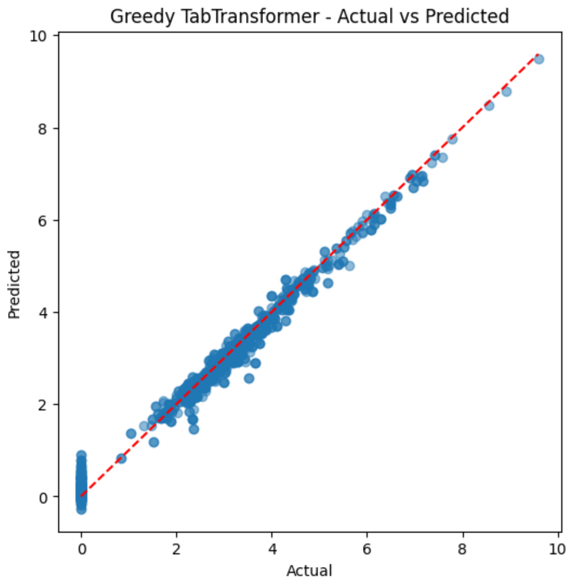
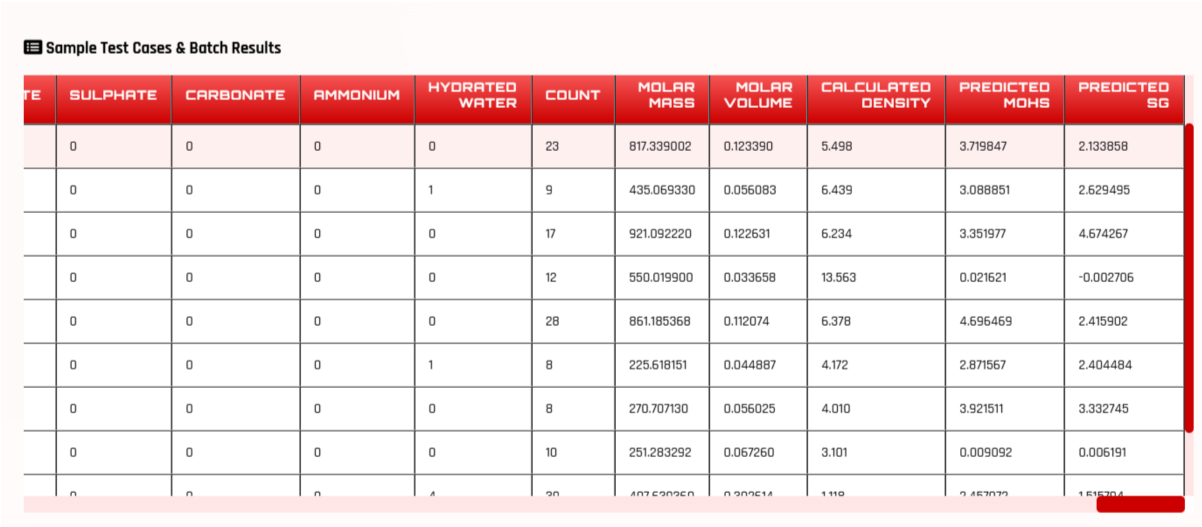


Fig. 4: Scatter plots of specific gravity target attribute of various regressor models

Fig. 4 GTT: The scatter plot presents predictions from the GTT The points are closely aligned along the diagonal line, demonstrating that GTT provides the highest accuracy and reliability in predicting specific gravity among all tested models.

Fig. 5 shows the Prediction Results interface, where the system displays the predicted outputs for the uploaded mineral dataset. The page presents a tabular view of input features such as Name, Chemical Composition (Iron, Silicon, etc.), Crystal Structure, Optical Properties, Refractive Index, along with the model-generated predictions for Mohs hardness and specific gravity appended to the dataset. Users can verify individual predictions directly and download the complete results in CSV format for further analysis. This screen allows seamless evaluation of batch prediction results and supports validation of the models' performance on unseen mineral samples



Sample Test Cases & Batch Results

SULPHATE	CARBONATE	AMMONIUM	HYDRATED WATER	COUNT	MOLAR MASS	MOLAR VOLUME	CALCULATED DENSITY	PREDICTED MOHS	PREDICTED SG
0	0	0	0	23	817.339002	0.123390	5.498	3.719847	2.133858
0	0	0	1	9	435.069330	0.056083	6.439	3.088851	2.629495
0	0	0	0	17	921.092220	0.122631	6.234	3.351977	4.674267
0	0	0	0	12	550.019900	0.033658	13.563	0.021621	-0.002706
0	0	0	0	28	861.185368	0.112074	6.378	4.696469	2.415902
0	0	0	1	8	225.618151	0.044887	4.172	2.871567	2.404484
0	0	0	0	8	270.707130	0.056025	4.010	3.921511	3.332745
0	0	0	0	10	251.283292	0.067260	3.101	0.009092	0.006191

Fig.5 Prediction results from test csv file.

Table. 1: Performance comparison of mohs hardness using various models

Algorithm	MAE	MSE	RMSE	R^2 Score

CBR Model	0.1895	0.3427	0.5854	0.8458
ABR Model	0.2722	0.5084	0.7130	0.7713
RFR Model	0.1773	0.3340	0.5780	0.8497
GTT Model	0.0459	0.0205	0.1432	0.9908

The table 1 presents a comparative analysis of four regression models CBR, ABR, RFR, and GTT for predicting Mohs hardness. Among these models, the GTT demonstrates superior performance, achieving the lowest error values with an MAE of 0.0459, MSE of 0.0205, and RMSE of 0.1432. It also records the highest R^2 score of 0.9908, indicating an excellent fit and very high predictive accuracy. RFR and CBR show moderate performance, with RMSE values of 0.5780 and 0.5854, respectively, and R^2 scores around 0.85, reflecting reasonable prediction accuracy. ABR exhibits the weakest performance, with the highest MAE (0.2722), MSE (0.5084), and RMSE (0.7130), along with the lowest R^2 (0.7713). These results confirm that GTT provides the most precise and reliable predictions for Mohs hardness among all evaluated models, highlighting its effectiveness in capturing complex relationships between chemical and physical mineral features.

Table. 2: Performance comparison of specific gravity using various models

Algorithm	**MAE**	**MSE**	**RMSE**	**R^2 Score**
CBR Model	0.1320	0.2069	0.4549	0.8555
ABR Model	0.1764	0.3144	0.5607	0.7805
RFR Model	0.1136	0.1700	0.4123	0.8813
GTT Model	0.0290	0.0075	0.0869	0.9947

The table 2 presents a performance comparison of four regression models CBR, ABR, RFR, and GTT for predicting Specific Gravity. Among these models, the GTT achieves the best performance, with the lowest errors: MAE of 0.0290, MSE of 0.0075, and RMSE of 0.0869. It also records the highest R^2 score of 0.9947, reflecting excellent prediction accuracy and model fit. RFR and CBR demonstrate moderate performance, with RMSE values of 0.4123 and 0.4549, and R^2 scores of 0.8813 and 0.8555, respectively, indicating reasonable predictive capability. ABR performs the weakest, showing the highest error values and the lowest R^2 score of 0.7805. Overall, these results confirm that GTT provides the most precise and reliable predictions for Specific Gravity, effectively capturing complex relationships between the chemical and physical properties of minerals.

5. CONCLUSION

This research successfully demonstrates a multimodel regression system capable of accurately predicting Mohs hardness and specific gravity of minerals using a combination of chemical and physical features. The system integrates multiple machine learning models, including CBR, ABR, RFR, and GTT, with advanced feature representation through the TT. Among these, the GTT achieves the highest predictive accuracy, producing the lowest MAE, MSE, and RMSE values while maintaining the highest R^2 scores. The TT embeddings capture complex relationships among mineral properties, and the GT component introduces clear decision boundaries, improving model interpretability and stability. By combining these models, the system provides consistent and reliable predictions that outperform traditional approaches based on manual Mohs hardness testing. The final results confirm that this web-based multimodel regression framework is robust, scalable, and well-suited for real-world applications in mineral classification, quality assessment, and geological analysis. The research delivers an efficient, accurate, and practical solution for predicting mineral properties, reducing the need for labor-intensive laboratory testing and enabling faster decision-making in scientific and industrial contexts.

REFERNCES

[1] Martins, F. F. & Miranda, T. F. S. Estimation of the rock deformation Modulus and RMR based on data mining techniques. Geotech. Geol. Eng. 30, 787–801. https://doi.org/10.1007/s10706-012-9498-1 (2012).

[2] Lawal, A. I. & Kwon, S. Application of artificial intelligence to rock mechanics: an overview. J. Rock. Mech. Geotech. Eng. 13, 248–266. https://doi.org/10.1016/j.jrmge.2020.05.010 (2021).

[3] Shao, W. et al. The application of machine learning techniques in geotechnical engineering: A review and comparison. Mathematics 11, 3976. https://doi.org/10.3390/math11183976 (2023).

[4] Miranda, T., Correia, A. G., Santos, M., Ribeiro e Sousa, L. & Cortez, P. New models for strength and deformability parameter calculation in rock masses using data-mining techniques. Int. J. Geomech. 11 (1), 44–58. https://doi.org/10.1061/ASCEGM.1943-5622.0000071 (2011).

[5] Martins, F. F. & Begonha, A. Amália Sequeira braga, M. Prediction of the mechanical behavior of the Oporto granite using data mining techniques. Expert Syst. Appl. 39, 8778–8783. https://doi.org/10.1016/j.eswa.2012.02.003 (2012).

[6] Ocak, I. & Seker, S. E. Estimation of elastic modulus of intact rocks by artificial neural network. Rock. Mech. Rock. Eng. 45, 1047–1054. https://doi.org/10.1007/s00603-012-0236-z (2012).

[7] Aboutaleb, S., Behnia, M., Bagherpour, R. & Bluekian, B. Using non-destructive tests for estimating uniaxial compressive strength and static young's modulus of carbonate rocks via some modeling techniques. Bull. Eng. Geol. Environ. 77, 1717–1728. https://doi.org/10.1007/s10064-017-1043-2 (2018).

[8] Zhang, T.; Zhou, X.; Zhang, P.; Duan, Y.; Cheng, X.; Wang, X.; Ding, G. Hardness Prediction of Laser Powder Bed Fusion Product Based on Melt Pool Radiation Intensity. Materials 2022, 15, 4674. https://doi.org/10.3390/ma15134674

[9] Guo, Y.; Liu, M.; Yan, Y. Hardness Prediction of Grind-Hardening Layer Based on Integrated Approach of Finite Element and Cellular Automata. Materials 2021, 14, 5651. https://doi.org/10.3390/ma14195651

[10] Rodríguez-Rosales, N.A.; Montes-González, F.A.; Gómez-Casas, O.; Gómez-Casas, J.; Galindo-Valdés, J.S.; Ortiz-Cuellar, J.C.; Martínez-Villafañe, J.F.; García-Navarro, D.; Muñiz-Valdez, C.R. Statistical Data-Driven Model for Hardness Prediction in Austempered Ductile Irons. Metals 2022, 12, 676. https://doi.org/10.3390/met12040676

[11] Liu, S.; Zhao, H.; Yuan, Z.; Xiao, L.; Shen, C.; Wan, X.; Tang, X.; Zhang, L. A Machine Learning Approach for the Autonomous Identification of Hardness in Extraterrestrial Rocks from Digital Images. Aerospace 2025, 12, 26. https://doi.org/10.3390/aerospace12010026

[12] Bermanec, M.; Eleish, A.M.; Morrison, S.M.; Prabhu, A.; Wong, M.L.; Hazen, R.M. The Evolution of Mineral Hardness Reveals Both Changing Parageneses and Preservational Bias in the Mineralogical Record. Minerals 2023, 13, 1089. https://doi.org/10.3390/min13081089

[13] Xu, C.; Dong, H.; Yan, Z.; Wang, L.; Ning, M.; Liu, X.; He, C. Micromagnetic and Quantitative Prediction of Hardness and Impact Energy in Martensitic Stainless Steels Using Mutual Information Parameter Screening and Random Forest Modeling Methods. Materials 2025, 18, 1685. https://doi.org/10.3390/ma18071685

[14] Ghadernejad, S.; Esmaeili, K.; Consens, M.P. Sensor-Based Rock Hardness Characterization in a Gold Mine Using Hyperspectral Imaging and Portable X-Ray Fluorescence Technologies. Remote Sens. 2025, 17, 2062. https://doi.org/10.3390/rs17122062

[15] Cheng, J.; Ling, Y.; De Waele, W. An ANN Hardness Prediction Tool Based on a Finite Element Implementation of a Thermal–Metallurgical Model for Mild Steel Produced by WAAM. Metals 2024, 14, 556. https://doi.org/10.3390/met14050556