Integrating Geochemical Data and Machine Learning: A Case Study from the Mohmand District, Pakistan

Muhammad Shahab[1*], Muhammad Atif Bilal[2], Liaqat Ali[3], Fahad Alshehri[1], Irfan Ullah[1], Manzoor Ilahi[2]

[1] *Abdullah Al Rushaid Earth Sciences Remote Sensing Research Chair, King Saud University, Saudi Arabia.*

[2] *College of Geo-Exploration Science & Technology, Jilin University, Changchun 130061, China*

[3] *GIS and Space Application in Geosciences Lab, National Centre of Excellence in Geology, University of Peshawar, Peshawar 25130, Khyber Pakhtunkhwa, Pakistan*

**Email: shahabgeo07@gmail.com*

This study investigates the geochemical properties of rocks and minerals from the Mohmand District, Pakistan, focusing on the prediction of copper (Cu) concentrations. Geochemical data were obtained through atomic absorption spectroscopy, ensuring precise measurements of trace and major elements. The geochemical dataset, was used to develop predictive models employing Random Forest (RF) and Support Vector Machines (SVM). The RF model outperformed SVM, achieving a coefficient of determination (R^2) of 0.81 and a root mean square error (RMSE) of 17.42. These results indicate a moderate level of predictive accuracy for Cu concentration. Significant variables influencing the prediction included Cr2O3 and other trace elements, highlighting the complex interplay of oxides and elemental compositions in Cu geochemistry. The inclusion of such variables demonstrated their critical role in enhancing model performance. While the RF model provided robust predictions within the constraints of the dataset. This research establishes a solid baseline for using machine learning models in geochemical exploration. The findings underscore the potential of RF as a reliable tool for geochemical prediction, particularly in regions like Pakistan, where mineral exploration remains underdeveloped. Future work will focus on integrating additional geochemical features and larger, more representative datasets to enhance the precision of predictions and contribute to the broader field of geoscience research.