**Random Forests**

Random Forest (RF) is a machine learning algorithm that has shown great potential in various fields, including mineral analysis. This feasibility study aims to evaluate the suitability of RF for future mineral analysis by examining its advantages, limitations, and performance metrics compared to other models.

RF is an ensemble of decisions tree’s that divides the feature space into distinct regions based on the input attributes. At each node in the tree, a decision is made by selecting the best split that optimises a certain criterion, such as impurity or information gain. In a RF, multiple decision trees are trained on different subsets of the data using the bagging method and the predictions of each tree (classifier) is aggregated and the class that gets the most votes is the prediction. This ensemble method leverages the diversity among the trees to capture more variability in the data, ultimately reducing both bias and variance, and often resulting in a more robust and accurate prediction.

Rf has been successfully applied in various fields, including the classification of mineral grains using energy-dispersive X-ray spectroscopy (EDS) data. One study “Classification of Minerals and the Assessment of Lithium and Beryllium Content in Granitoid Rocks by Laser-Induced Breakdown Spectroscopy” focused on using laser-induced breakdown spectroscopy (LIBS) to classify mineral grains and assess the content of lithium and beryllium in granitoid rocks. The researchers employed a RF algorithm to classify mineral grains based on their major components, such as aluminium, iron, potassium, and calcium. The RF achieved high accuracy in classifying mineral grains, with an accuracy of 95.4% for lithium and 97% for beryllium. The RF algorithm effectively utilized the diverse training subsets generated through bootstrap aggregating, resulting in robust classification performance.

A second study "Data-Driven Predictive Modelling of Mineral Prospectivity Using Machine Learning and Deep Learning Methods" aimed to predict mineral prospectivity in the southern Jiangxi Province of China using machine learning. The researchers employed a RF algorithm as an ensemble learning algorithm to aggregate multiple decision trees for repeated predictions. The RF algorithm demonstrated high accuracy in predicting mineral prospectivity, showcasing its ability to handle diverse training subsets and generate reliable predictions.

Another study "Mineral Grains Recognition Using Computer Vision and Machine Learning" introduced an original computational approach to automate the recognition of mineral grains from numerical images obtained with a simple optical microscope. Random Forest was one of the machine learning algorithms used for classification. The study split the dataset into training and testing sets, with 70% used for training and 30% for testing. They achieved a classification accuracy of at least 92.6% for all three minerals (quartz, feldspar, biotite) using RF. This outperformed other algorithms, such as CART and k-NN, achieving a global accuracy that was 12% and 6% better, respectively.

Advantages of Random Forest for Mineral Analysis:

1. High Accuracy: Research by Zhang et al. Demonstrates that RF outperforms other machine learning models, such as Support Vector Machine (SVM) and Artificial Neural Network (ANN), in mineral prospectivity mapping. The RF model achieved an accuracy of 92.38%, while the SVM and ANN models achieved 87.62% and lower accuracies, respectively.
2. Robustness to Complex Data: RF has been proven to handle complexly distributed and nonlinearly associated input features effectively. This is particularly advantageous in mineral analysis, where the relationships between geological features and mineralization can be intricate and non-linear.
3. Variable Importance: RF provides a measure of variable importance (feature importance), allowing researchers to identify the most significant features contributing to mineral predictions. This information can aid in understanding the geological processes and indicators of mineralisation.

Limitations of Random Forest for Mineral Analysis:

1. Interpretability: Random forests are often considered as "black-box" models, making it challenging to interpret the underlying decision-making process. Understanding the specific reasons behind the predictions can be difficult.
2. Overfitting: Although random forests are less prone to overfitting compared to individual decision trees, using too many trees can still lead to overfitting. It is crucial to find the right balance between the number of trees and the complexity of the model.
3. Computational Complexity: Random forests can be computationally expensive, especially when dealing with large datasets or a high number of features. Training and evaluating a large number of decision trees can require significant computational resources.
4. Sensitivity to Noisy Data: Random forests can be sensitive to noisy or irrelevant features in the dataset. Including irrelevant features may lead to decreased performance and increased computational requirements.

Conclusion:

Random forests have proven to be a successful and feasible machine learning algorithm in various domains. Their ability to handle large datasets, generate accurate predictions, and capture complex relationships makes them a valuable tool for classification and regression tasks. However, it is important to consider the limitations of RF, such as their interpretability challenges, potential for overfitting, computational complexity, and sensitivity to noisy data. By understanding these limitations, informed decisions can be made when applying RF for mineral analysis.

*References*

*Huizhen Hao, Ronghua Guo, Qing Gu, & Xiumian Hu. (2019). Machine learning application to automatically classify heavy minerals in river sand by using SEM/EDS data. Minerals Engineering.* [*https://doi.org/https://doi.org/10.1016/j.mineng.2019.105899*](https://doi.org/https://doi.org/10.1016/j.mineng.2019.105899)

*Li, C., Wang, D., & Kong, L. (2021). Application of Machine Learning Techniques in Mineral Classification for Scanning Electron Microscopy - Energy Dispersive X-Ray Spectroscopy (SEM-EDS) Images. Journal of Petroleum Science and Engineering, 200, 108178.* [*https://doi.org/10.1016/j.petrol.2020.108178*](https://doi.org/10.1016/j.petrol.2020.108178)

*Maitre, J., Bouchard, K., & Bédard, L. P. (2019). Mineral grains recognition using computer vision and machine learning. Computers & Geosciences, 130, 84–93.* [*https://doi.org/10.1016/j.cageo.2019.05.009*](https://doi.org/10.1016/j.cageo.2019.05.009)

*Biau, G., & Scornet, E. (2016). A Random Forest guided tour. TEST, 25(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7*