Classification of Rock Mineral in Field X based on Spectral Data (SWIR & TIR) using Supervised Machine Learning Methods

S A Pane¹, F M H Sihombing¹

¹Department of Geosciences, Geology Study Program, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia

*Corresponding author: felix.sihombing@ui.ac.id

Abstract. The massive development of science and technology in the industrial era 4.0 includes artificial intelligence, which is purposed to produce research output in the field of geology more accurately and can be completed in a short time using large amounts of data. Machine learning is a part of artificial intelligence that can provide learning processes on computers independently without explicit programming. The process of identifying rocks through classification can be done using machine learning. The study area is in the Manjimup region, Western Australia which consists of Volcanogenic Massive Sulphide (VMS) deposits. This study purposed to determine the classification of rock minerals using accuracy values from the evaluation of models generated using supervised machine learning based on spectral data, namely Short-Wavelength Infrared (SWIR), and Mid or Thermal Infrared (TIR) acquired from electromagnetic spectrum measurements to identify rock mineral features. The spectral data comes from five rock drilling data in the study area. The supervised machine learning method used to determine the best accuracy consists of 5 types of methods, which are K-Nearest Neighbors (K-NN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and Multi-layer Perceptron (MLP). Machine learning is completed by supervised method because the research data contains information about label data, which is the type of rock mineral so that it can produce a classification based on the level of accuracy for each type of rock mineral data. The SVM method produces the best accuracy on SWIR data with 82.5% accuracy and the MLP method produces the best accuracy on TIR data with 82% accuracy for rock mineral classification.

Keywords: Classification, Supervised Learning, Machine Learning, Mineral, Spectral Data, Spectral Geology

1. Introduction

Machine learning is the study of how to make computers capable of learning independently without explicit programming [7]. A machine learning algorithm can be categorized as an optimal algorithm when it can carry out a decision-making process automatically based on the generalization results of sample data whose input and output are known (supervised), and the output is unknown (unsupervised) [5]. Machine learning has been used in earth sciences to classify various geological aspects needed in energy resource estimation activities [2]. This research is based on how to classify the types of minerals in the rock in field X based on SWIR & TIR spectral data using machine learning methods—based on SWIR & TIR spectral data using machine learning methods. The research location is in Western Australia, with coordinates 34°09'28.7"S 115°59'52.9"E, to be precise, in the Manjimup region in Figure 1. The research is using spectral data derived from the measurement results of the Spectroscope instrument from five borehole data.
2. Geological Overview

The study area consists of 5 stratigraphic units in Figure 2, from old to young, namely: Garnet quartzite characterized by granoblastic, equigranular rocks, with equivalent garnet and quartz content [9]. Schist is characterized by a schist containing kyanite. Gneiss quartz-feldspar-biotite (-garnet), characterized by a fine bond, contains blastomylonite. Laterite is characterized by massive, containing layers of pisolithic gravel and minorly laminated sand. The Collovium is characterized by a valley-fill deposit, which has a lateralized variation.

In the study area, base metal mineralization is found, characterized by gneiss quartz-feldspar-biotite and amphibolite amphibole-biotite-feldspar associated with quartz-garnet-biotite-sillimanite-staurolite gneiss in the footwall position of the mineralization zone and volcanic rock. These are formed in layers with sedimentary rocks that undergo alteration associated with the formation of Volcanogenic Massive Sulfide (VMS) deposits [8]. Metamorphism facies found in the study area are granulite facies in the presence of orthogneiss and paragneiss, and amphibolite facies [8].
3. Methods

The research was conducted using spectral data. Spectral data were obtained from borehole data located in the Manjimup region, western Australia. The data is displayed in The Spectral Geologist (TSG) application, which includes information about rock sample codes, rock depths, types of minerals in rocks, and mineral reflectance at specific wavelengths. The wavelengths used are Short-Wavelength Infrared (SWIR) and Thermal Infrared (TIR).

Furthermore, the data is carried out to enter the pre-processing stage, where data and mineral types are sorted for use in the data processing. Data sorting is carried out to select data that has important information in machine learning activities in order to reduce errors due to inadequate or missing data information to produce maximum final results. The drill dataset consists of five types of borehole data. Sorting of mineral types is based on geological conditions on SWIR waves, which indicate alteration minerals [1]. Whereas in the TIR wave, sorting of mineral types is based on minerals produced by metamorphism [3].

Continued with machine learning activities using the supervision method, which consists of 5 methods, namely K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and Multi-layer Perceptron (MLP). Machine learning is carried out to determine the mineral classification, which is indicated by the accuracy value for each machine learning method.

After the pre-processing stage, data then used as training for supervised machine learning algorithm. There are five algorithm used in this study, namely the K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and Multi-layer Perceptron (MLP). The objective of machine learning to determine the mineral classification, indicated by their accuracy value.

4. Results

There are five data export results, namely WPD02, WPD03, WPD04, WPD07, and WPD13. Based on this data, mineral types will be considered in determining the data used for training and testing activities in making machine learning models and making predictions. Sorting of mineral types is carried out on each drill to make a selection so that the minerals that will be processed in learning activities are minerals with the dominant amount. In SWIR minerals, the minerals used in the learning activities are alteration minerals, while in the TIR minerals, metamorphic minerals are used.

The Table 1 shows the type of SWIR minerals in WPD03 and WPD13 that, used for training & testing process. Mineral selection is carried out to obtain classification results with good accuracy and contain the required geological information. The geological information needed is the alteration mineral data in the Volcanic Massive Sulphide area.

| No | Mineral   | Total |
|----|-----------|-------|
| 1  | Biotit 954 |
| 2  | Hornblende 917 |
| 3  | Montmorillonite 838 |
| 4  | Phlogopite 570 |
| 5  | Klorit 223 |
| 6  | Kaolinit 90 |
|    | **Total** 3592 |
The Table 2 represents the SWIR mineral types in WPD02, WPD04, and WPD07, used for prediction activities. The types of minerals used in prediction activities are all data from the types of minerals used in training and testing. In mineral prediction, the difference in the amount of data between mineral types is not a problem because all data from mineral types are needed to obtain the predicted rock mineral log using machine learning.

| No | Mineral | Total | No | Mineral | Total | No | Mineral | Total | Total |
|----|---------|-------|----|---------|-------|----|---------|-------|-------|
| 1  | Biotit  | 1816  | 1  | Biotit  | 7724  | 1  | Biotit  | 7365  | 16905 |
| 2  | Hornblende | 7888 | 2  | Hornblende | 7438 | 2  | Hornblende | 9218 | 24544 |
| 3  | Klorit  | 1280  | 3  | Klorit  | 1008  | 3  | Klorit  | 870   | 3158  |
| 4  | Kaolinit | 30    | 4  | Kaolinit | 101   | 4  | Kaolinit | 145   | 276   |
| 5  | Montmorillonite | 279 | 5  | Montmorillonite | 194 | 5  | Montmorillonite | 251 | 724   |
| 6  | Phlogopite | 1091 | 6  | Phlogopite | 1345 | 6  | Phlogopite | 746   | 3182  |
|    | Total     | 12384 |    | Total     | 17810 |    | Total     | 18595 | 48789 |

Table 3 shows the type of TIR minerals in WPD03 and WPD13, which are used for training & testing activities. Mineral selection is carried out to obtain classification results with good accuracy and contain the required geological information. The geological information needed is metamorphic mineral data in the Volcanic Massive Sulphide area.

| No | Mineral   | Total |
|----|-----------|-------|
| 1  | Hornblende | 1961  |
| 2  | Andesine  | 1939  |
| 3  | Anorthite | 1898  |
| 4  | Kuarsa    | 1807  |
| 5  | Labradorit | 1479  |
| 6  | Oligoklas | 411   |
| 7  | Biotit    | 175   |
| 8  | Microcline | 100   |
| 9  | Augite    | 48    |
| 10 | Almandine | 38    |
|    | Total     | 9856  |
Table 4 shows the type of TIR minerals in WPD02, WPD04, and WPD07, which are used for prediction activities. The types of minerals used in prediction activities are all data from the types of minerals used in training and testing. In mineral prediction, the difference in the amount of data between mineral types is not a problem because all data from mineral types are needed to obtain the predicted rock mineral log using machine learning.

| Table 4. Distribution of TIR mineral data for prediction |
|---------------------------------------------------------|
| **Distribution of TIR mineral data (prediction)**        |
|                                                        |
| ![Table 4](image)                                       |
|                                                        |
| The supervised method is a machine learning method used to classify by labeling the data that will be carried out by training & testing. Supervised methods consist of KNN, SVM, DT, RF, and MLP. The most optimal method with the highest accuracy value is the SVM method for SWIR data and the MLP method for TIR data. SWIR data that has been used in machine learning with the SVM method consists of hornblende, biotite, chlorite, kaolinite, montmorillonite, and phlogopite minerals. SVM is one of the supervised learning methods used to classify and regress a mathematical model and possible extension for complex models that cannot be defined with a hyperplane in the input space [5]. **Figure 3a** is a confusion matrix of SWIR training & testing data using the SVM method. The results show that hornblende mineral has the largest TP value, namely 180 out of 180 data, then biotite with 177 data from 194 data, and montmorillonite with 173 out of 184 data. Meanwhile, **Figure 3b** is a confusion matrix of WPD02 SWIR prediction data using the SVM method. The Confusion Matrix is a table that displays information about predictions and actual conditions of machine learning results [4]. Confusion matrix is used in evaluating the performance of a classification model using machine learning. The results show that hornblende mineral has the highest TP value, namely 19972 from 24544 data, then biotite with 15819 from 16905 data, and chlorite with 1942 from 3158 data. |
Figure 3. The SWIR data configuration matrix uses the SVM method (a: training & testing, b: prediction)

The TIR data that has been used in machine learning using the MLP method consists of minerals almandine, andesine, anorthite, augite, biotite, hornblende, labradorite, microcline, oligoclasts, and quartz. MLP is one of the supervised learning methods in artificial neural networks and the parameters used in the MLP method are the number of hidden layers that represent the number of neurons in the hidden layer, the activation parameters used to activate the hidden layer function, and the settlement or solver parameters used for data weight optimization [6]. Figure 4a is a confusion matrix of TIR training & testing data using the MLP method. The results showed that quartz mineral has the highest TP value, namely 372 from 386 data, then hornblende with 359 from 401 data and anorthite with 313 from 369 data. Meanwhile, Figure 4b is a confusion matrix of the WPD247 TIR prediction data using the MLP method. The results show that quartz mineral has the highest TP value, namely 34705 from 39226 data, then labradorite with 2736 from 3400 data and andesine with 1727 from 3329 data.

Figure 4. The TIR data configuration matrix uses the MLP method (a: training & testing, b: prediction)

Evaluation of the supervised machine learning model is carried out to determine the most optimal classification method characterized by the method with the highest accuracy value on the prediction data. Table 5 is a recap of each classification method's accuracy value, namely KNN, SVM, RF, DT, and MLP, using SWIR data. In SWIR data, the method with the highest accuracy value for training & testing data is the SVM method with a value of 91.9%. The highest accuracy value in predictive data is owned by the SVM method, with a value of 82.5%. The SVM method is classified by paying attention to the kernel parameter and the C or Cost parameter, which functions as an SVM optimization. The kernel that managed to achieve the highest accuracy value was the linear kernel, and the C parameter value was 10000 based on repeated experiments. The second highest accuracy value is owned by the MLP method with a training & testing accuracy value of 77.2% and a predictive accuracy of 78%.
Table 5. Recap accuracy values using SWIR data

|          | SWIR KNN (%) | SVM (%) | RF (%) | DT (%) | MLP (%) |
|----------|--------------|---------|--------|--------|---------|
| Training & Testing Data | WPD313 | 75.5 | 91.9 | 77 | 70.5 | 77.2 |
| Prediction Data | WPD247 | 57 | 82.5 | 56 | 44 | 78 |

Table 6 is a recap of each classification method's accuracy value, namely KNN, SVM, RF, DT, and MLP, using TIR data. In the TIR data, the method with the highest accuracy value for training & testing data is the MLP method with a value of 74.3%. The highest accuracy value in predictive data is owned by the MLP method, with a value of 82%. The MLP method parameters used to achieve the highest accuracy value are the number of hidden layers of 100 layers, the activation parameter of the relu or rectified linear unit, which is used to create the limit at zero, and the settlement parameter, namely adam.

Table 6. Recap accuracy values using TIR data

|          | TIR KNN (%) | SVM (%) | RF (%) | DT (%) | MLP (%) |
|----------|-------------|---------|--------|--------|---------|
| Training & Testing Data | WPD313 | 70.6 | 73.9 | 72.4 | 59.4 | 74.3 |
| Prediction Data | WPD247 | 77 | 81 | 80 | 69 | 82 |

5. Conclusion

The classification of rock minerals in the SWIR data was carried out by machine learning using supervised methods, namely the KNN, SVM, DT, RF, and MLP methods. The best method marked with the highest accuracy value compared to other methods is the SVM method with the training & testing accuracy value of 91.1%, and the predictive accuracy value of WPD247 is 82.5%. Meanwhile, the classification of rock minerals in the TIR data was carried out by machine learning using supervised methods, namely the KNN, SVM, DT, RF, and MLP methods. The best method marked with the highest accuracy value compared to other methods is the MLP method with training & testing accuracy of 74.3%, and the predictive accuracy value of WPD247 is 82%.

6. Acknowledgements

This research and publication is funded by PUTI Grant No. NKB-969/UN2.RST/HKP.05.00/2020. We acknowledge Universitas Indonesia that provide the grant.

References

[1] Bonnet, A.L., and Corriveau, L., (2007). *Alteration vectors to metamorphosed hydrothermal systems in gneissic terranes*, in *Mineral deposits of Canada: A synthesis of major deposit-types, district metallogeny, the evolution of geological provinces, and exploration methods*: Geological Association of Canada, Mineral Deposits Division, Special Publication No. 5, p. 1035–1049.

[2] Cate, et al (2017). *Machine Learning as a Tool for Geologists*. Institut National de la Recherche Scientifique.
[3] Kornprobst, Jacques, (2002). *Metamorphic Rocks and Their Geodynamic Significance: a Petrological Handbook.*

[4] Liu, Bing. (2011). Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Springer-Verlag, Berlin, Heidelberg.

[5] Muller, Andreas C & Guido, Sarah., (2016). *Introduction to Machine Learning with Python: A Guide for Data Scientists.* O'Reilly Media, Inc. United States of America.

[6] Pedregosa et al., (2011). Scikit-learn: Machine Learning in Python. JMLR 12, pp. 2825-2830.

[7] Samuel, L. Arthur. (1959). *Some studies in machine learning using the game of checkers, IBM J. Res. Dev.*, vol. 44, no. 1–2, pp. 207–219.

[8] Savage, B.C. (2007). Wheatley project, Annual report E70/2258 & E70/2339; Hampton Hill Mining; Geological Survey of Western Australia, Statutory mineral exploration report, A75300, 13p.

[9] Wilde, S. A & Walker, I. W. (Ian William) & Geological Survey of Western Australia (1984). *Pemberton-Irwin Inlet, Western Australia: sheet SI/50-10, SI/50-14 international index.* Geological Survey of Western Australia, Perth, W.A.