Enhancing Lithology Classification through Machine Learning Models: Assessing the Generalizability of Single Boreholes in Northwestern Bowen Basin, Australia

Yu, Z[1] Si, G[2] Tang, K[3] Salamakha, V[4] Oh, J[5] Wu, X[6]

1. PhD Candidate, School of Minerals and Energy Resources Engineering, University of New South Wales, Sydney, NSW 2052, Australia. Email: zexin.yu@student.unsw.edu.au
2.A/Professor, School of Minerals and Energy Resources Engineering, University of New South Wales, Sydney, NSW 2052, Australia. Email: g.si@unsw.edu.au
3.Postdoctoral Research Fellow, School of Minerals and Energy Resources Engineering, University of New South Wales, Sydney, NSW 2052, Australia. Email: kunning.tang@unsw.edu.au
4. Undergraduate, School of Minerals and Energy Resources Engineering, University of New South Wales, Sydney, NSW 2052, Australia. Email: v.salamakha@student.unsw.edu.au
5. A/Professor, School of Minerals and Energy Resources Engineering, University of New South Wales, Sydney, NSW 2052, Australia. Email: joung.oh@unsw.edu.au
6. Mphil Candidate, School of Minerals and Energy Resources Engineering, University of New South Wales, Sydney, NSW 2052, Australia. Email: xuebin.wu1@student.unsw.edu.au
Keywords: Machine Learning, generalizability, Bowen Basin, lithology classification, geophysical logs

ABSTRACT

The development of machine learning (ML) algorithms has led to promising advances in lithology classification from geophysical logs, which is indispensable in various underground engineering applications. Due to the resolution and availability of logging data, ML models can recognize lithology change in a definitive form at a fine scale. While the results from previous models indicated high performance in the classification, limited literature reported whether these models have a robust capability of generalizing lithology prediction from one borehole to the other, which forms a fundamental step for correlation among boreholes. To this end, this paper aims to investigate the ability of 4 ML models trained on single reference boreholes to test the other in the same region, which includes Support Vector Classifier (SVC), Random Forest (RF), eXtreme Gradient Boosting (XGBoost) and ResNet10. The dataset involves 11 boreholes from a coal mine in northwestern Bowen Basin (Queensland, Australia), where four logs were applied as input: DENB, GRDE, SSN, MC6F. Additionally, a histogram matching method was introduced to transform the other boreholes' data for better classification results. The results show that there is no significant difference among the models trained by the algorithms and different reference boreholes, despite the various performance on their validation data. The histogram matching can align log readings in the reference boreholes with those of the other, contributing an average increase of 0.10 in Macro-F1 score. Meanwhile, the classification results can be more regulated than the direct generalization results, which strengthen the models' correlations with boreholes' spatial distribution and overall geological trend. The matching method can serve as a data preprocessing step in the ML workflow, facilitating regional lithology identification, geological or geophysical modelling and stratigraphic analysis.