Advancing Continental-Scale MVT Prospectivity Mapping: A Comparative Analysis of Bayesian Neural Networks and Data-Driven Negative Sampling in North America

Roberto De La Rosa[1], Michael Steffen[1], Andreas Knobloch[1] and Joshua Rosera[2]

[1] *Beak Consultants GmbH, 09599 Freiberg, Germany*
[2] *U.S. Geological Survey, Geology, Energy & Minerals Science Center, Reston, VA, USA*

Abstract

Mississippi Valley-type (MVT) Zn-Pb deposits are a critical source of base metals worldwide. Recent continental-scale assessments across the United States and Canada have successfully applied machine learning for prospectivity mapping. However, mineral occurrence data are inherently positive-unlabeled. While known deposits provide positive labels for training, the vast majority of the study area remains unlabeled. Because defining distinctly barren zones requires expert knowledge, researchers traditionally extract negative training data randomly from these unlabeled areas. This random sampling introduces stochastic bias and creates false negatives by misclassifying prospective environments as barren.

Standard deterministic models lack a formal mechanism to quantify epistemic uncertainty, often mistaking a lack of data with low geological potential. To address this, we present an integrated deep learning workflow targeting 1,823 positive labels (derived from 185 MVT deposits and 7,176 occurrences). We implement a Bayesian neural network (BNN) coupled with a self-organizing map (SOM) sampling strategy. The application of BNNs allows for the decomposition of predictive uncertainty by treating internal parameters as probability distributions updated via variational inference. Simultaneously, our data-driven SOM approach utilizes unsupervised clustering to characterize the multivariate geological space, effectively filtering "prospective-like" signatures from the negative training set.

We evaluate performance using a comprehensive suite of metrics, including accuracy, balanced accuracy, precision, recall, F1 Score, the Matthews correlation coefficient, the receiver operating characteristic – area under curve (ROC - AUC) score, and the average precision score. Results demonstrate that the SOM-guided methodology outperforms unconstrained random sampling. We visualize these findings using a bivariate triangular color scale that simultaneously maps prospectivity and uncertainty. This reveals that while standard models often indicate low prospectivity in under-explored regions, BNN-derived uncertainty layers expose high epistemic risk. By distinguishing between areas of low geologic potential and high epistemic risk, this workflow improves the reliability of continental-scale prospectivity mapping.