Deep structural features in prospectivity mapping for hydrothermal-type gold deposits/showings in the Red Lake-Stormy region, Superior Province

Authors: Parham Adiban[1], Lyal B Harris[1], Erwan Gloaguen[1]

Affiliation : Institut national de la recherche scientifique (INRS)[1]

Summary

Orogenic gold deposits are linked to geological structures at different crustal that act as conduits for hydrothermal fluids. The western Superior Province in Canada hosts world-class Au deposits such as the Red Lake camp (Card *et al.*, 1998). In this study, deep regional structures at high angles to general mapped lithological and structural trends identified from enhanced aeromagnetic, pseudogravity and ground gravity data are integrated with geological data. Machine learning algorithms (notably the XGBoost and Random Forest algorithms) were trained using a compilation of the geophysical and geological datasets and engineered datasets from classical spatial analysis techniques (viz. buffer analysis, fry analysis, and fractal analysis) to produce a statistical model that defines the spatial relationship between the regional shear structures and orogenic gold mineralisation. This project was funded through NRCan's TGI mineral exploration program "Increasing Deep Exploration Effectiveness" and NSERC.

Method and Workflow

Analysis comprised structural interpretation using geophysical data, data preparation for prospect mapping, and application of machine learning to develop a predictive model.

Structural interpretation

Deep lineaments without a surface were interpreted from enhanced long wavelength aeromagnetic and ground Bouguer gravity components. Lineaments are classified as linear or curvilinear features with a magnetic and/or gravity response that contrasts their surroundings. Lithological data and existing geological models guide the interpretations to reassure they are reasonable within the geological context. East-West structures are prevalent in the English River, Winnipeg River, and Wabigoon subprovinces and they mark the boundary between the subprovinces. Three major North-South structures extend from the North Caribou Core subprovince through the Uchi, English, River and Winnipeg subprovinces at a high angle the regional East-West trend. The change in trend during the transition from the North Caribou Core to the Uchi subprovince and other smaller structures in the Uchi and English River subprovinces indicate a transpression shear zone where East-West shortening resulted in secondary N-S shear structures.

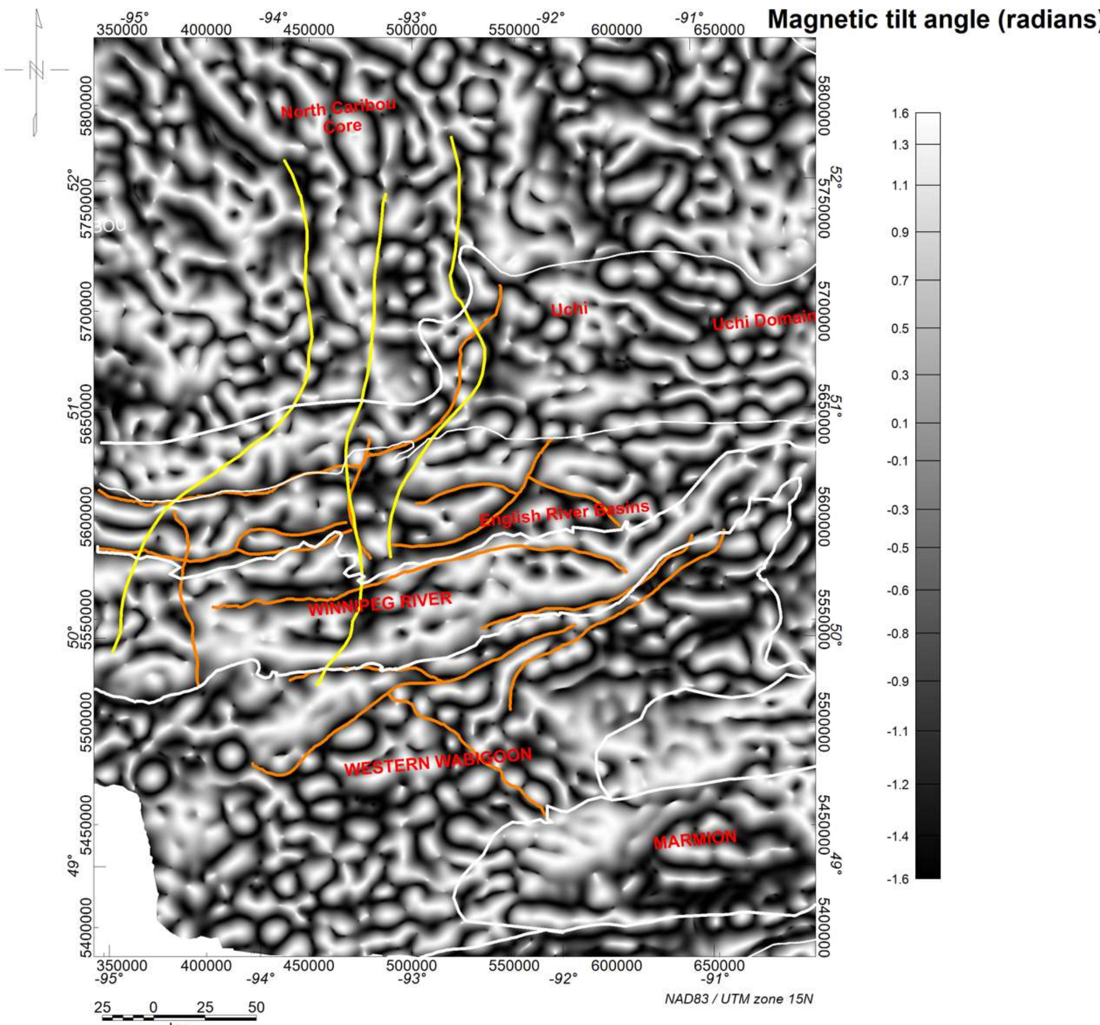

Figure 1 Tilt angle map of the reduced to pole total magnetic data from the Red lake – Stormy region. The white lines are subprovince boundaries, the yellow lines are North-South trending structures, and red lines are East-West trending structures. The tilt angle θ is calculated by the arctangent of the vertical derivative divided by the horizontal derivative of the magnetic anomaly (Blakely et al., 2016). Source: Ontario Geological Survey.

Data preparation

The data set prepared for developing a mineral prospectivity map consisted of: the location of all mapped faults; EW striking lineaments; NS striking deep lineaments (perpendicular to structural trend), the reduced to pole total field magnetic field; the pseudo-gravity (transforming the magnetic data from a dipolar field to a monopolar field); the tilt angle of the reduced to pole image; the tilt angle of the pseudo-gravity image; the bouguer gravity anomaly; and the analytical signal. Statistical properties mean, minimum, maximum, median, variance, standard deviation, skewness, and kurtosis of each map were calculated over 20 px × 20 px rolling window.

The area of study is a region approximately 400 km × 300 km.

Machine learning

Training a machine learning model involves finding the parameters that can fit the input “x” and make predictions based on labels “y”. The machine learning algorithms implemented are extreme gradient boosting (XGBoost) and random forest (RF); both are decision tree ensembles. Decision trees rely on a chain metrics that separate the data and the end of each chain, the labels of the input data are evaluated and assigned a prediction. Single decision trees are prone to overfitting. Ensemble methods combine several decision trees to reduce overfitting. Random forests use bagging (or bootstrap aggregation). Bagging trains decision trees on random subsets of the data, reduces variance and makes the model less prone to overfitting. RF also bags the features, with a random subset of features considered at each split of the decision tree. This technique reduces the impact of strong predictors and reduces correlation of trees. “Boosting” samples the data sequentially and adds weight to instances that were incorrectly predicted. Therefore, the model focuses on instances that are harder to learn, and they are sampled more often resulting in a lower probability that the learner will make the same error again. Moreover, the gradient of the error (prediction vs. correct outcome) determines the direction the parameters change to reduce error in later samples. In extreme gradient boosting: second order gradients of the loss function are calculated, and they allow minimizing the loss function; and L1 and L2 regularization reduce overfitting.

The XGBoost and RF algorithms models were trained on three datasets: an oversampled dataset, an undersampled dataset, and the original imbalanced dataset (positive and negative labels are not equal).

Results and Conclusions

The oversampled models (Figure 2) overestimate locations of gold showings/deposits, whereas the undersampled models underestimate them. Although a very low threshold was used to determine what is considered a prospective pixel in my data, the unsampled dataset remains highly imbalanced (positive labels make up 5240 data points of the overall 39424). Another method of undersampling is removing regions manually; by removing areas further than 15km from faults and lineaments, however, there is a risk of overtraining the data around structures, losing positive results outside of the buffer regions, which risk losing information for negative results (i.e. areas where there are not any gold prospects).

Overall, the oversampled dataset is preferred as it has the highest recall rate, but prone to overfitting as the minority class is randomly replicated to balance the data. The prediction model developed using the original dataset produced very high accuracy values, however the recall rates for both XGBoost and RF were below 50% and they were unable to efficiently identify gold deposits/showings.

The XGBoost model produced the best results on the dataset. Ranking the features in order of importance using Shapley Additive Explanation (SHAP), which calculates the average contribution of each feature through several permutations, showed that the deep North-South trending shear zones are rated higher than the ones trending East-West. Further, there is a positive correlation between North-South structures and gold deposits/showings (positive classification by the model).

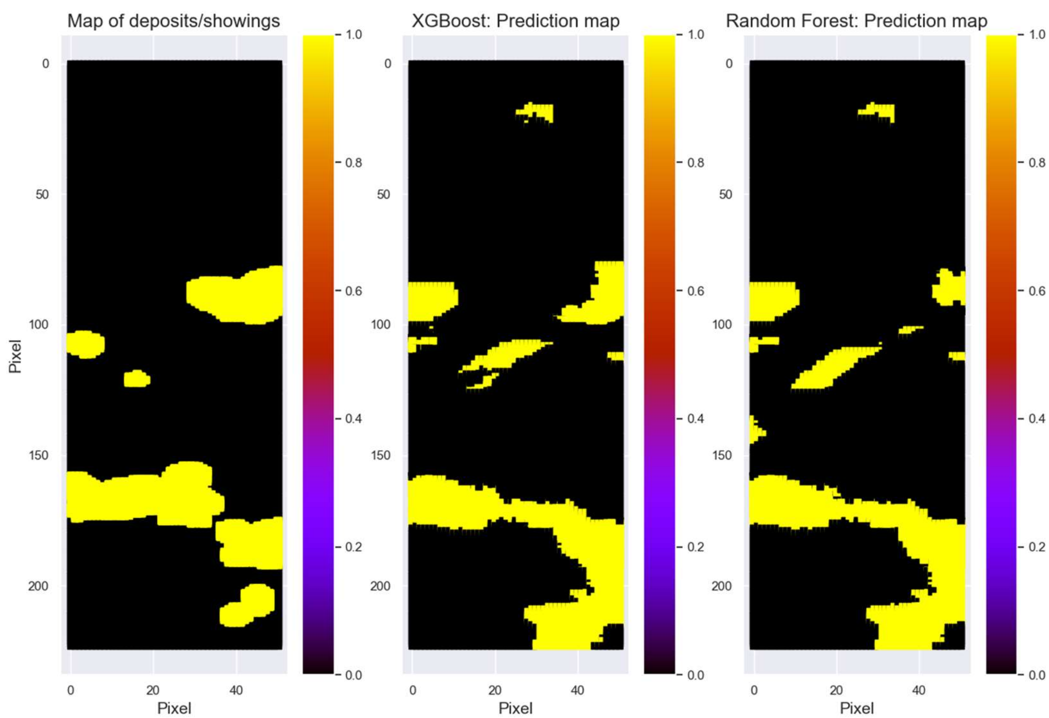

Figure 2 Prediction map of gold deposits and showings of the oversampled dataset. Value of 1 indicates a positive result. Positive results (gold deposit and showings) cover a larger area of the map produced by the RF model compared to the XGBoost model. This is reflected on their respective precision and recall values from Table 1. The RF model has a higher recall rate as the larger coverage includes more true positive pixels but a lower precision as it exaggerates the number of positive results. XGBoost has a higher precision rate but the lower recall rate indicates that it can predict 69.8% of positive results compared to the 88.7% by the RF model.

Table 1 Evaluating Extreme Gradient Boosting and Random Forest classifiers using the training accuracy, testing accuracy, recall, and precision. Training accuracy is the accuracy of the model at predicting the training data, testing accuracy is the accuracy of the model at predicting the test data (data it has not seen before), recall refers to the rate at which the positive outcome is true, and recall, and precision indicates the percentage of results that are relevant.

	XGBoost classifier				*Random Forest classifier*			
	Training Accuracy	Testing Accuracy	Recall	Precision	Training accuracy	Testing Accuracy	Recall	Precision
No sampling	93%	88%	42%	45%	91%	91%	23%	75%
Undersampling	83%	78%	60%	85%	79%	77%	56%	88%
Oversampling	83%	80%	67%	85%	80%	77%	60%	84%

References

Blakely RJ, Connard GG & Curto JB (2016) Tilt Derivative Made Easy. Édit Geosoft).

Card K, Poulsen K, Lucas S & St-Onge M (1998) Geology and mineral deposits of the Superior Province of the Canadian Shield. *Geology of the Precambrian Superior and Grenville Provinces and Precambrian Fossils in North America, Geology of North America* 1:13-204.