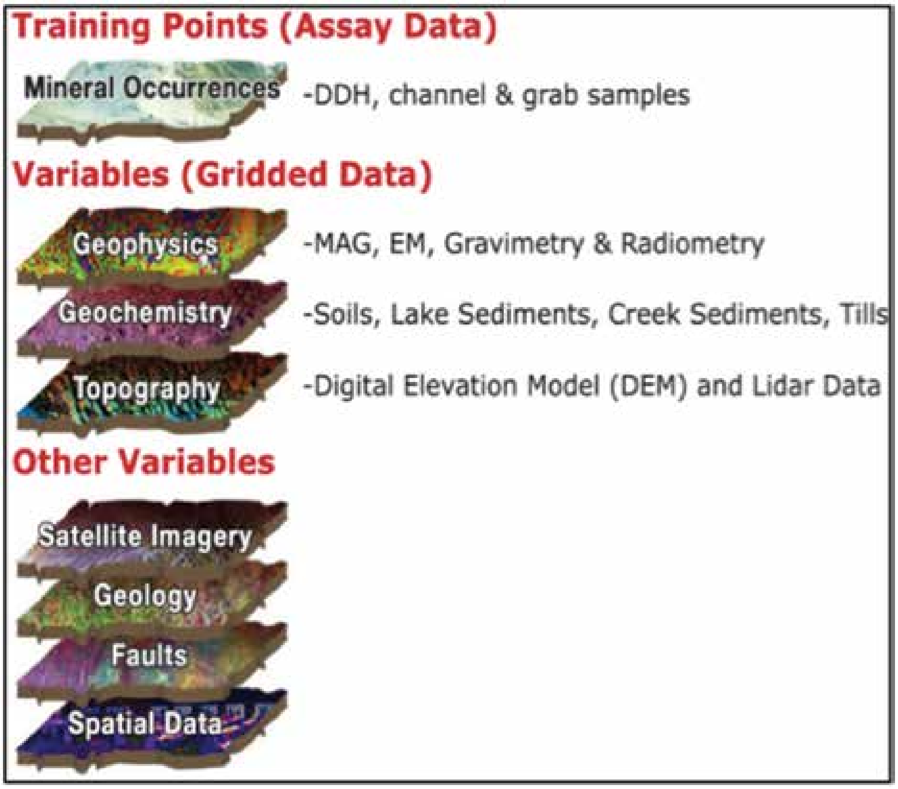
Mining exploration in the era of big data

A recent *Mining Journal* survey of more than 70 mining companies and individuals identified big data analytics, the so-called 'Internet of things,' at the top of 10 technologies that will "rock the mining world" by 2025. The accurate analysis of data will lead to better decisions, resulting in cost reductions and reduced risk. Diagnos has developed its computer-aided resources detection system (CARDS) artificial intelligence (AI) system to improve predictive analytics for exploration and evaluating geology and mineralization.

Since 2004, Diagnos has developed CARDS, a new system that uses artificial intelligence in order to create data-driven models of mineral prospectivity. CARDS uses data mining techniques such as:

- Decision trees.
- Ensemble methods.
- Clustering.
- Random forest, spatial regressions, boosting, etc.



The CARDS computer system uses many layers of data.

These are used on geophysical, geochemical and geological data to identify new targets for further exploration of undiscovered deposit-type locations. Data-driven mineral prospectivity mapping is appropriate in areas representing moderately to well sampled (or so-called brownfields) mineralized landscapes.

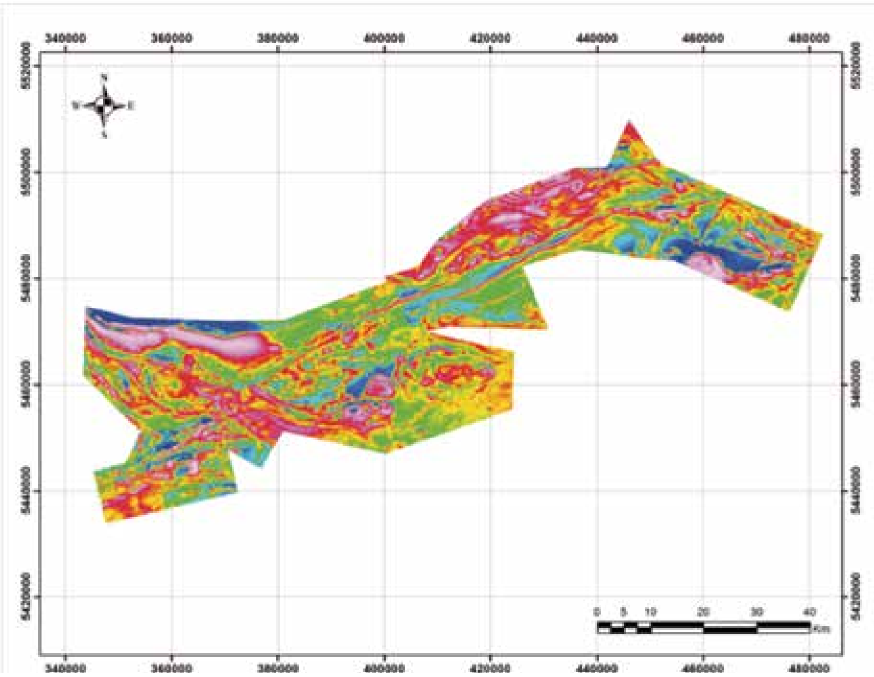
CARDS modeling and prediction system

CARDS is a state-of-the-art computer system that uses the latest artificial intelligence and pattern recognition algorithms to analyze large digital exploration data sets and produce exploration targets. CARDS uses many layers of gridded data (variables) to learn the "signature" of known mineralized sites (positive cells) in a given area. The area is then scored and cells with a high similarity to the sought signature are identified.

The variables used in CARDS prediction models are divided into three categories: 1) the original data layers (primary variables), 2) the calculated data layers (derivative variables) and 3) the neighborhood data layers (neighborhood variables).

When taken alone, the primary data layers may contain only part of the information. In a gridded data layer, single point readings have little meaning, as they do not take spatial patterns into account. The neighborhood around each individual cell also contains important information and patterns. For example, there is no good reason for mineralization to occur at a single elevation. But when all of the cells of the topography grid are combined, patterns such as linear features, drainage patterns, circular patterns, etc. can appear and, in some cases, be an indicator of structure or lithology. The same logic applies to geophysical grids; it might be that certain slopes near high values have statistical significance. Such patterns can be represented by 1) calculating the derivatives of the primary grids and 2) calculating "neighborhood" variables, that allow the characteristics of all cells within a specified distance (neighborhood) to be weighed into the evaluation of each individual cell.

These many extra calculated layers are imputed in CARDS along with the primary



DIAGNOS used the 140 x 80 km MEGATEM survey (50 m resolution) available through the Ministry of Natural Resources of Quebec. The survey, which included magnetic and apparent conductivity grids, was also accompanied by a digital elevation model (SRTM) of the region.

layers creating an important training database. Each cell in this database is identified as positive or unknown, based on drill hole and rock sample assays, and linked to its own set of characteristics (primary, derivative and neighbouring variables). Several algorithms are then used to identify the unknown cells that have a set of characteristics most similar to the signature of the positive cells.

The quality and usefulness of results derived from CARDS modeling is dependent on a variety of factors including the coverage, quantity, variety and quality of geoscientific and historical exploration data processed. In addition, where interpreted data is used, it is also dependent on the adequacy of the interpretation.

Targets generated by CARDS should be evaluated in conjunction with all readily available geological data in the evaluation of the economic potential of a property as well as in the outlining of exploration targets.

Modeling

Aggregation of GEO-referenced models (AGEO). The AGEO algorithm, developed at Diagnos, is the main prediction algorithm used during the modeling phase. Based on ensemble learning methods[1] and semi-supervised learning methods[2], AGEO uses multiple classifiers, called decision trees [3], to discriminate between labeled (positive) and unlabeled (unknown) cells. The results of each classifier are then aggregated to produce the final model results. The advantage of using a decision tree based algorithm is that this type of prediction model permits the identification of the most important or discriminant variables. The importance of a variable may be due to its (possibly complex) interaction with other variables, but in the main, variables that appear frequently and in the top levels of AGEO's decision trees are more important.

As the modeling progresses, data mining experts of the CARDS team constantly evaluate the performance of the AGEO models in collaboration with the geoscientific team. This evaluation is based on the importance of variables in the decision trees and on the comparison with other statistic models. By coupling the modeling and model evaluation phases, certain aspects of the model can be controlled. For example, if a data layer considered weak by the geoscientific team appears to be too discriminant, it can be removed from the final model.

Class clustering (C-Cluster). C-Cluster algorithm was developed at Diagnos, and is based on resampling methods and clustering algorithm.

This new methodology of class prediction,

---

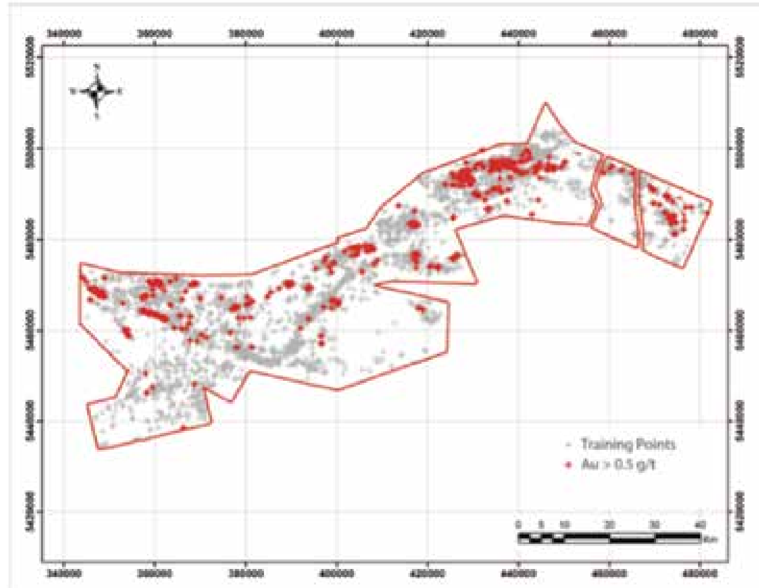
[1] Ensemble learning methods generate many classifiers and aggregate their results. In fact, ensemble methods use multiple models to obtain a better predictive performance than could be obtained from any of the constituent models.

[2] Semi-supervised learning is a class of machine learning techniques that makes use of both labeled and unlabeled data for training, typically a small amount of labeled data with a large amount of unlabeled data. Semi-supervised learning falls between unsupervised learning (without any labeled training data) and supervised learning (with completely labeled training data).

[3] The decision tree represents the classification process as a series of nested choices or questions which enable the identification of the predictable attributes. At each step (node) in the process, a single binary or multinomial question is posed, and the answer determines the next set of choices to be made. The path between the root (first node) and the leaf (terminal node) of the decision tree is an assignment rule of the type "if condition, then conclusion", and the hierarchical rules of the tree constitute the prediction model.

without using negative occurrences in the training process, is tailored to the task of modeling mineral exploration data. The method can best be thought of as a predictive approach, to model and score a wide range of geo-referenced data. The methodology is based on clustering, and in conjunction with resampling techniques, it provides for a method to classify an unknown data points across multiple runs of a clustering algorithm and to assess their similarity with known mineralized data points.

The main task is to complete a cluster analysis for all data points (unknown and known positive occurrences combined) and to calculate the proportion of known positive data points in each cluster. The higher this proportion in a particular cluster is, the higher scored his unknown data points are.



1533 Positive Training Points (Au > 0.5 g/t) were used to generate the Au model (left). Diagnos also compiled 228 positive training points for copper (Cu > 0.1 percent) and 583 positive training points for VMS deposits (copper and zinc > 0.1 percent)

Modeling results and deliverables

Results are presented in the form of mineral prospectivity maps: a raster grid with each point representing a score value. The score represents the probability a point will resemble another group of points where specific mineral occurrences are present. Review of prospectivity maps should be carried out in cooperation with BC Geosciences personnel prior to public presentation.

Mineral prospectivity maps will be submitted in paper format and digital format: Geosoft, ArcGIS or MapInfo. Maps will be accompanied by a report detailing the techniques and methodology used.

Diagnos: Project experience and historical success

Diagnos has completed hundreds of data mining and target generation projects for a variety of mineral exploration companies in its 10 years of experience. Projects were located in Quebec, Ontario, New Brunswick, Newfoundland, Chile, Nevada, Dominican Republic, Mexico, Burkina Faso and Tanzania.

Diagnos has already completed many projects of similar endeavour using private and public data to generate exploration targets in several areas worldwide. The Quest geoscience data are perfect data sets for target generation data mining projects. Diagnos is pleased to present with the consent of its client Metanor a summary of a successful project completed with the use of public data in the following section.

Lebel project history

In 2009, Diagnos modeled the Lebel-Sur-Quévillon region of the Abitibi Greenstone belt, Quebec. This area is reputed for Cu-Zn massive sulfides, Cu-Zn vein deposits and lode gold deposits.

Diagnos compiled 11,470 training points for the Lebel project. These points originated from 4950 diamond drill holes and 6438 grab samples. The assays were extracted from the assessment reports available through the MRN website.

On behalf of Metanor Resources Inc., Diagnos conducted an exploration campaign to validate several gold targets surrounding its Bachelor Lake Mine. Three grab samples collected 2.5 km west of the mine revealed assay values between 11.03 and 14.8 g/t Au.

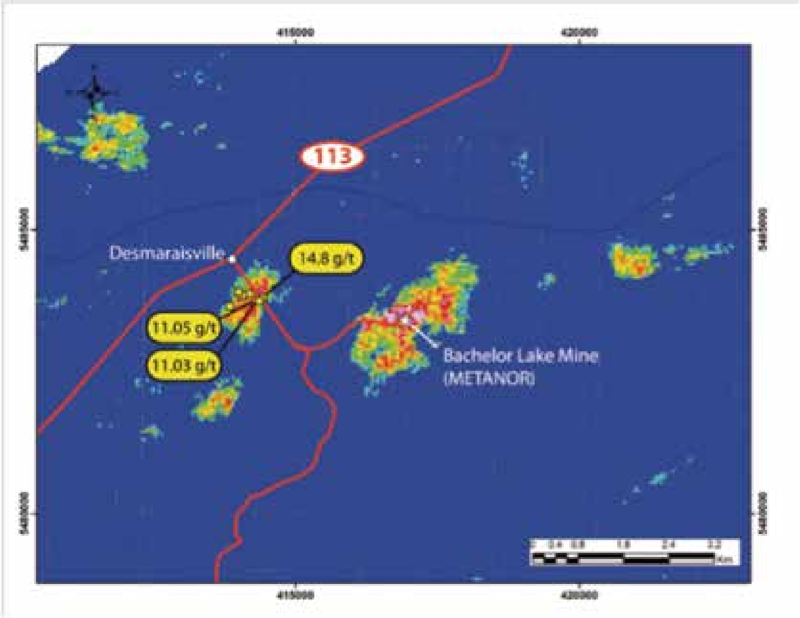
The mineralized quartz vein, which was partly hidden under the access road to the mine, was discovered in six days of work.

A 1527.5-m drilling program was conducted in 2010 to investigate this new zone. The gold -earing structure was intercepted in eight different drill holes and mineralization was followed over a horizontal distance of 250 m to the west.

The best intersection was 8.48 g/t Au over 2.65 m (HW-10-05).

In 2011, Diagnos was mandated by its Ontario partner to generate target maps for gold, copper, silver, zinc and nickel over an area of 330,900 km^2 in the northwestern part of Quebec. The following year, a total of 5,242 claims (2,335 km^2) were acquired to cover the most promising and unexplored targets. Diagnos was responsible for managing the entire land package and finally conducted a \$2.2-million exploration program during the summer of 2013 in order to validate some of these targets.

In 2013, Diagnos also executed a CARDS analysis using the public geological and



Diagnos conducted an exploration campaign to validate several gold targets surrounding Metanor Resources' Bachelor Lake Mine.

geophysical databases available for the Bathurst Mining Camp, New Brunswick. This enormous database was composed of 26,887 training points from drill hole and rock samples as well as 50 m multi-parameter airborne surveys (magnetic, electromagnetic gravimetric and radiometric). The database was combined with privately owned heliborne surveys conducted by the Bathurst option joint venture between Votorantim Metals Canada Inc. Glencore Canada Corp. (formerly Xstrata Canada Corp.) and El Nino Ventures Inc. to generate copper, zinc and silver target maps over the entire camp. The same year, El Nino announced that it had intersected mineralization of interest while drilling a CARDS target located in the south central part of the Bathurst Mining Camp.

CARDS detects patterns in a multidimensional dataset composed of several hundred layers that are too difficult to be perceived by humans. It exceeds all other statistical tools because the grid pixels are not only analysed individually, but the spatial relationship between each pixel and its surroundings (neighborhood) is also taken into account during the modeling process. Furthermore, the system withdraws all negative mineralized occurrences before creating pattern signatures and finally combines several models to avoid over-learning. This new technology is still subject to research and development and is continuously being improved by a multidisciplinary team composed of data scientists, geologists and geophysicists.

The final target maps, which are validated by professional geologist for quality control, are easily importable into most GIS as a product to support geologists and geophysicists in the exploration process. This data not only adds value to different layers of existent geophysical, geochemical and geological data, but can also be used as a marketing tool to promote exploration investments.

Diagnos will be presenting at SME's first Big Data/Smart Data Conference Oct. 14, 2014 at the Princeton Club in New York. ■