

## Case History

# Lithologic mapping using Random Forests applied to geophysical and remote-sensing data: A demonstration study from the Eastern Goldfields of Australia

Stephen Kuhn<sup>1</sup>, Matthew J. Cracknell<sup>1</sup>, and Anya M. Reading<sup>1</sup>

### ABSTRACT

The Eastern Goldfields of Western Australia is one of the world's premier gold-producing regions; however, large areas of prospective bedrock are under cover and lack detailed lithologic mapping. Away from the near-mine environment, exploration for new gold prospects requires mapping geology using the limited data available with robust estimates of uncertainty. We used the machine learning algorithm Random Forests (RF) to classify the lithology of an underexplored area adjacent to the historically significant Junction gold mine, using geophysical and remote-sensing data, with no geochemical sampling available at this reconnaissance stage. Using a sparse training sample, 1.6% of the total ground area, we produce a refined lithologic map. The classification is stable, despite including parts of the study area with later intrusions and variable cover depth, and it preserves the

stratigraphic units defined in the training data. We assess the uncertainty associated with this new RF classification using information entropy, identifying those areas of the refined map that are most likely to be incorrectly classified. We find that information entropy correlates well with inaccuracy, providing a mechanism for explorers to direct future expenditure toward areas most likely to be incorrectly mapped or geologically complex. We conclude that the method can be an effective additional tool available to geoscientists in a greenfield, orogenic gold setting when confronted with limited data. We determine that the method could be used either to substantially improve an existing map, or produce a new map, taking sparse observations as a starting point. It can be implemented in similar situations (with limited outcrop information and no geochemical data) as an objective, data-driven alternative to conventional interpretation with the additional value of quantifying uncertainty.

### INTRODUCTION

With the increasing cost and difficulty of new discovery in areas with substantial amounts of cover, there is a need for improved approaches to mineral exploration. In the Eastern Goldfields of Australia, very detailed geologic, geochemical, and geophysical data sets exist near mines. There is, however, a sharp transition into adjacent greenfield areas where such data are not available and the geology is significantly less well-constrained. Geophysical and

remote-sensing data are widely available at a reasonable resolution either in the form of government or multient data sets or as a first-pass acquisition performed by explorers when new ground is acquired. Machine learning presents an attractive way forward, facilitating the use of these data to improve a preliminary lithology map or to produce a starting map from limited observations: in each case, improving an explorer's ability to identify targets. Previous studies, however (e.g., Waske et al., 2009; Cracknell et al., 2014; Harris and Grunsky, 2015), have primarily used a richer and more diverse set of

Manuscript received by the Editor 4 September 2017; revised manuscript received 13 February 2018; published ahead of production 19 March 2018; published online 31 May 2018.

<sup>1</sup>University of Tasmania, School of Physical Sciences (Earth Sciences) and CODES Centre of Excellence in Ore Deposits, Canberra, Australia and University of Tasmania, ARC Industrial Transformation Research Hub for Transforming the Mining Value Chain, Canberra, Australia. E-mail: stephen.kuhn@utas.edu.au; m.j.cracknell@utas.edu.au; anya.reading@utas.edu.au.

© The Authors. Published by the Society of Exploration Geophysicists. All article content, except where otherwise noted (including republished material), is licensed under a Creative Commons Attribution 4.0 Unported License (CC BY). See <http://creativecommons.org/licenses/by/4.0/>. Distribution or reproduction of this work in whole or in part commercially or noncommercially requires full attribution of the original publication, including its digital object identifier (DOI).

data inputs such as geochemistry or additional spectral information or made use of a different algorithm, such as, for example, support vector machines (SVMs) (Yu et al., 2012) or artificial neural networks (Barnett and Williams, 2009). In this study, we assess the ability of the machine learning algorithm (MLA) Random Forests (RF) to produce a geologic classification using only those geophysical and remote-sensing data that would be available to an explorer in a greenfield, early stage exploration environment.

### Geologic setting

The Heron South project area is located approximately 15 km east of the Junction gold mine, in the St. Ives Goldfield of the Yilgarn Craton, Western Australia (Figures 1 and 2). The St. Ives camp is estimated to contain in excess of 300 t of gold, with orogenic and (to a lesser extent) intrusion related gold deposits hosted throughout the entire local stratigraphy, making it one of Australia's largest gold-producing districts (Crawford, 2011). The Archean (2.7–2.6 Ma)

bedrock stratigraphy comprises a series of mafic-ultramafic volcanic and intrusive units, volcanoclastic sediments, and felsic intrusions crosscut by Proterozoic-age basaltic dikes. The region has undergone pervasive, regional greenschist to lower amphibolite metamorphism. The St. Ives Goldfield is bound to the west and east by the Merougil and Boulder-Lefroy Fault Zones, respectively.

The region was subject to several distinct phases of deformation between 2675 and 2620 Ma. Until recently, the deformational framework for the region had largely focused on compressional events (e.g., Ngyuen, 1997; Swager, 1997; Connors et al., 2002). The more recent study by Blewett et al. (2010) includes events related to extension, important in understanding the formation of the younger volcano-sedimentary units of the region (Squire et al. 2010). The revised framework proposed by Blewett et al. (2010) is as follows: D<sub>1</sub> is characterized by east–northeast/west–southwest extension. The D<sub>2</sub> period represents a phase of east–northeast/west–southwest contraction that caused regional north–northwest-trending folds and reactivation of faults produced during D<sub>1</sub> as thrusts.

This was followed by D<sub>3</sub>, a period of extension on the same orientation. The D<sub>4a</sub> period of contraction tightened existing north–northwest folds and was followed by D<sub>4b</sub>, a period of sinistral transpression. During this deformation event, existing structures, such as the Boulder Lefroy Fault Zone, which passes through Heron South, were reactivated as sinistral strike slip faults. Localized deflections, step-overs, and local higher order structures produced during D<sub>4b</sub> are associated with the main mineralizing event in the region (Cox and Ruming, 2004; Blewett et al., 2010; Miller et al., 2010). The D<sub>5</sub> period of dextral transtension produced north–northeast-trending strike-slip high-angle faults. These structures may also be associated with a gold mineralizing event at various sites in the region (Connors et al., 2002; Ruming, 2006; Blewett et al., 2010; Miller et al., 2010). The D<sub>6</sub> period is not documented in the St. Ives Goldfield. The D<sub>7</sub> period of contraction is associated with the emplacement of dominantly east–northeast-trending Proterozoic dikes, which occur in abundance in the study area. The Heron South project area is proximal to the Boulder-Lefroy Fault Zone, which passes through the southwest of the project in a north–northwest orientation.

The geology of the study area is split by the Boulder Lefroy Fault Zone into a western and an eastern region. The western area forms part of the main St. Ives sequences and contains thick successions of Paringa Basalt and Black Flag Group volcano-sedimentary sequences. The eastern area contains north–south-striking, steeply dipping packages of mafic-ultramafic and sedimentary units impinged between larger granitoid bodies (Figure 2). It is anticipated that these units are correlates of the main stratigraphic sequence mapped at St. Ives, however, this has not yet been confirmed. For the purpose of this study, these units have been defined by the interpreted geologic

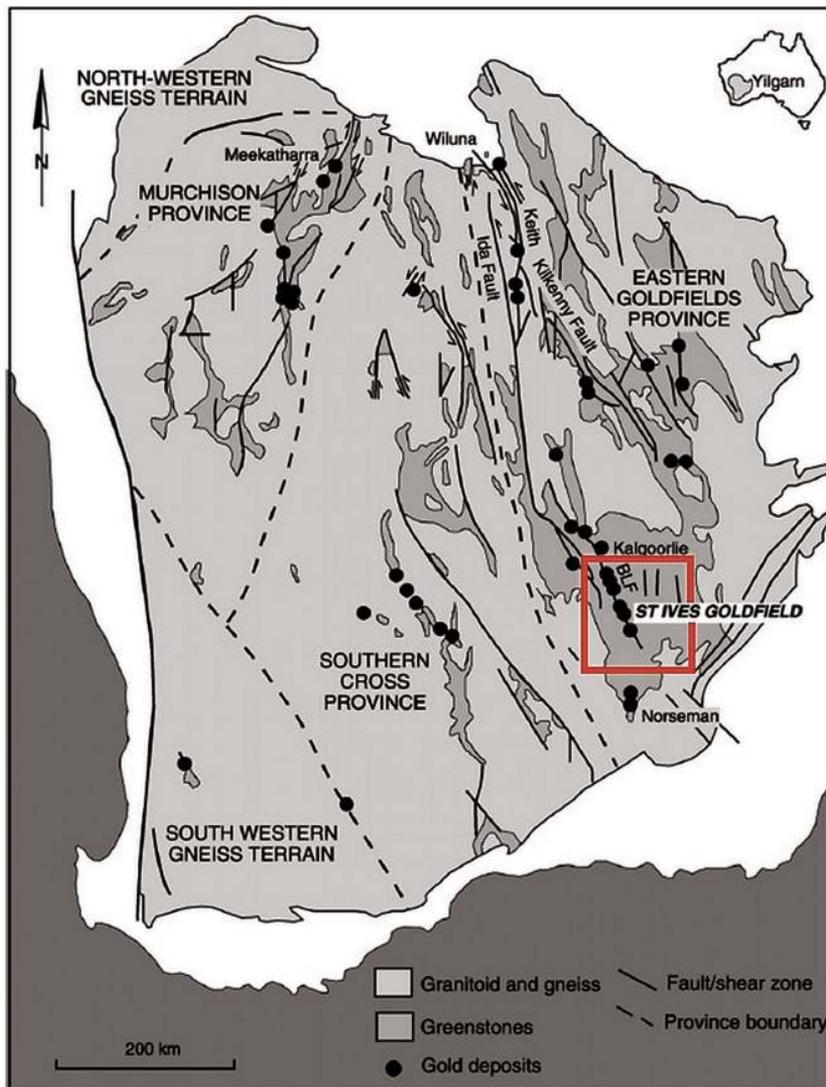


Figure 1. Schematic representation of the Yilgarn Craton including the location of major gold deposits. The approximate location of the Heron South project is shown in red (modified after Cox and Ruming, 2004).

map of the St. Ives Goldfield (Figure 3), as stratigraphically distinct. Stratigraphic labels can be assigned to these units as geochemical and geochronological information becomes available allowing these units to be amalgamated or subdivided as required at a future date.

## RF

RF (Breiman, 2001, pp. 5–32) is a supervised ensemble classification algorithm and an extension of the decision tree method. This classifier constructs a “forest” comprising many decision trees (Figure 4), allowing for superior performance and lower sensitivity to over-fitting compared with single classifiers (Hastie et al., 2009, pp. 587–604). Randomness is introduced at two stages during implementation of the algorithm. First, a process of bootstrap aggregation, known as bagging (Breiman, 1996, pp. 123–124) is used to modulate the training data ( $T_a$ ) available to each decision tree. Bagging obtains for each tree, via random sampling with replacement, a subset of  $T_a$  equal in size to  $T_a$ . This duplicates some samples and will not select others. An average of approximately 63.2% of instances is included in each training subset, whereas the remaining, or “out-of-bag,” samples (approximately 37.8%) are used for validation. The second form of randomization involves the selection of variables available to the classifier to split each node. At each node, a random subset of input variables selected from all available input variables. The number of variables in this subset is predefined and consistent across the forest. At every node, the randomly selected variables are then ranked by their ability to produce a split threshold that maximizes the homogeneity of child nodes (Figure 4) relative to the parent node. The decrease in the Gini index (equation 1), as implemented by Breiman et al. (1984), provides this measure. The Gini index is an expression of information purity given by

$$\text{Gini}(t) = \sum_{c=1}^j \mathbf{g}_c(1 - \mathbf{g}_c), \quad (1)$$

where  $\mathbf{g}_c$  is an expression of the relative frequency of each class  $c$ , of a set comprising  $j$  classes, at a given node  $t$ ;  $\mathbf{g}_c$  is given by

$$\mathbf{g}_c = \frac{n_c}{n}, \quad (2)$$

where  $n_c$  is the number of samples comprising class  $c$  at a given node and  $n$  is the total number of samples comprising that node. Using this measure, the variable that produces the greatest improvement in homogeneity in child nodes relative to the parent node is used to split the node at the threshold that produced the best split. This is repeated at every node until sufficient depth is reached to produce nodes with complete homogeneity (or approached to within a defined toler-

ance). The class assigned by RF to each sample is determined by a majority vote compiled from the output of all classification trees (Breiman, 2001, p. 6).

Many studies have noted a point of diminishing returns, necessitating a forest be grown to a certain extent where a stable error minima is approached, beyond which, additional trees are redundant (e.g., Waske et al., 2009; Cracknell et al., 2014; Rodriguez-Galiano et al., 2014; Harris and Grunsky, 2015). RF has been shown to achieve equal or better accuracy to other classification algorithms with the advantage that parameter selection is relatively straightforward (e.g., Hastie et al., 2009; Cracknell and Reading, 2013). The process of RF training can be performed on any PC with specifi-

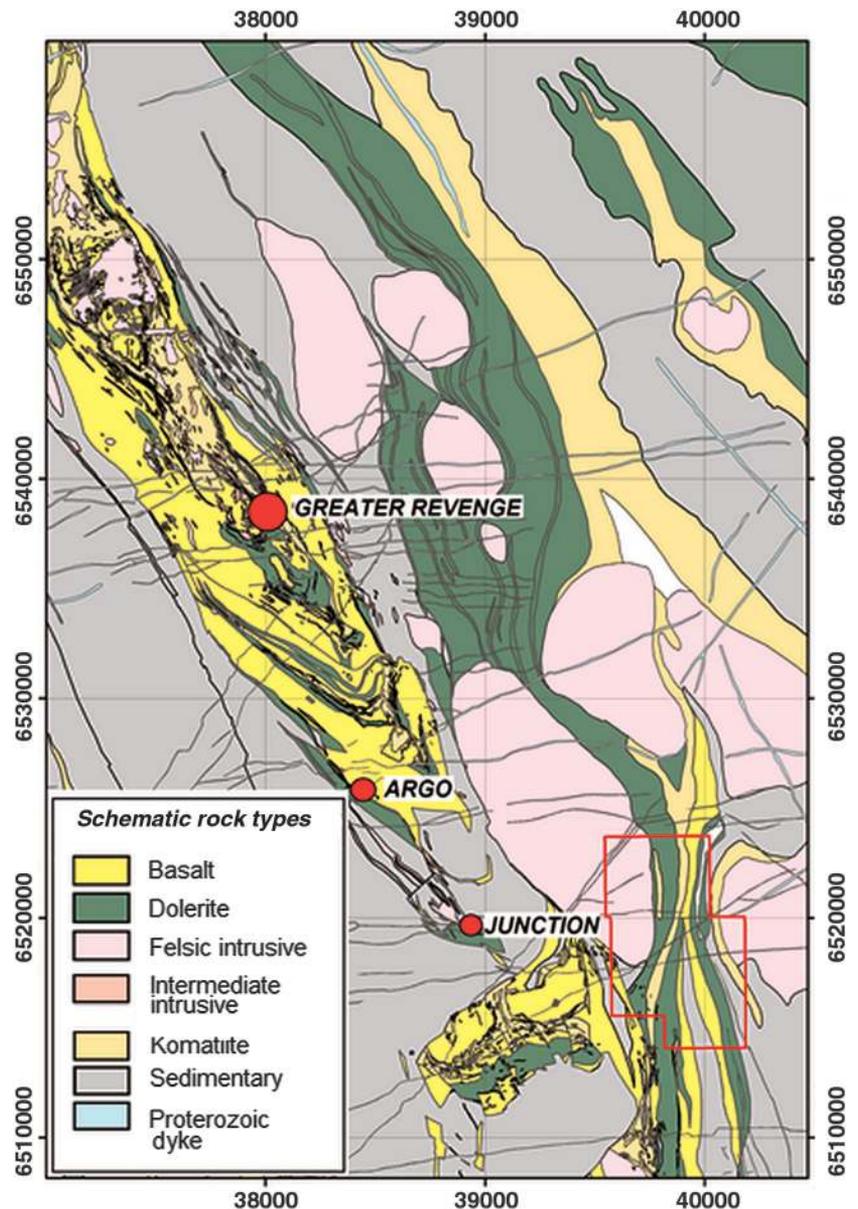


Figure 2. Schematic geology of the St. Ives gold camp including the location and extent of the project (red outline box) relative to several major existing and historical gold mines (indicated by red circles with the mine name adjacent). The project outline (red box) defines the extent of the project in all subsequent figures. Map coordinates are projected using WGS84, UTM grid 51S (m).

cations readily commercially available at the time of this study and does not require specialized equipment. In this study, combined training and cross validation of an RF for any given set of parameters required between 15 and 40 s on a Dell Precision T7610 with an Intel Xeon e2630 processor and 32Gb RAM. This is ideal for uptake by geoscientists because requirements for specialized computing skills and equipment are minimal.

RF has been increasingly applied to the problem of lithologic classification. Waske et al. (2009) compare RF and another popular MLA, SVM (Vapnik, 1995, 1998), in the context of mapping lithology using hyperspectral imagery. They conclude that RF and SVM achieve significantly more accurate results than standard classifiers. Although in that instance, SVM marginally outperformed RF, it is noted by the authors that RF remains an attractive option due to its high accuracy and ease of use. Cracknell and Reading (2014) compare RF with four other MLAs: SVM, Naïve Bayes,  $k$ -nearest neighbors, and artificial neural networks; as applied to lithologic mapping. In their study, RF marginally outperformed

other MLAs. Although there were only small differences in accuracy, Cracknell and Reading (2014) demonstrate that RF was able to produce accurate results with simpler input parameters and at less computational cost than other algorithms evaluated. Another study by Cracknell and Reading (2013) assesses RF and SVM for lithology mapping and identification of lithologic contacts and zones of structural complexity. They discover that RF, in addition to an excellent overall performance, produced more usable outputs. Unlike for SVMs, high uncertainty was spatially associated with incorrect classification and proximal to geologic boundaries and zones of high structural complexity. Cracknell and Reading (2014) note that with increasingly spatially dispersed training data, the comparative performance of RF improved further, widening the gap over other MLAs.

Cracknell and Reading (2014) demonstrate that RF was able to identify and redefine incorrectly mapped features in western Tasmania using 2% of the surface area as training samples. Harris and Grunsky (2015) use a similar approach, applying RF to geologic mapping in northern Canada. They test two  $T_a$  selection scenarios: one based on lake sediment geochemical sample locations and another based on field-mapping observations. Both approaches produced meaningful results with the authors concluding that RF is of value as a first-pass mapping tool or as a means of focusing effort into areas where there is a mismatch between predicted geology and legacy maps.

### Information entropy

There has been increasing effort in the field of mineral exploration to quantify the uncertainty associated with mapping and prediction. One such method, information entropy ( $H$ ) (Shannon, 1948), is defined as

$$H = -k \sum_{i=1}^n p_i \log p_i, \quad (3)$$

where  $p_i$  is the class membership probability at location  $i$ ,  $n$  is the number of candidate classes, and  $k$  is an arbitrary positive constant. The  $k$  and the base of the logarithm can be selected by the user to define the scale. Information entropy has been used to great effect in a “per-voxel” setting to demonstrate how uncertainty is distributed spatially (Wellmann and Regenauer-Lieb, 2012).

In the process of producing a final classification, RF calculates the class-membership probabilities. These are defined as the proportion of trees in a RF that voted for a given candidate class (Hastie et al., 2009). RF class-membership probabilities can be used in equation 3 to calculate  $H$  for each classified instance. The properties of  $H$  for a two-class, binary system are such that a value of 0 corresponds to a 100% probability of one class occurring and a value of 1 corresponds to an equal probability of both represented classes being present. Information entropy in its general form preserves monotonicity such that an increase in the number of candidate classes results in a higher  $H$ . For the purpose of this study, a normalized version of  $H$  has also been used to account for the number of candidate classes by dividing  $H$  by the logarithm of the number of classes present, such that  $H$  assigned to each pixel represents, on a scale of 0–1, the range of minimum to maximum possible  $H$  for that pixel. As such, all pixels are comparable with regard to how close they each internally approach their minimum or maximum

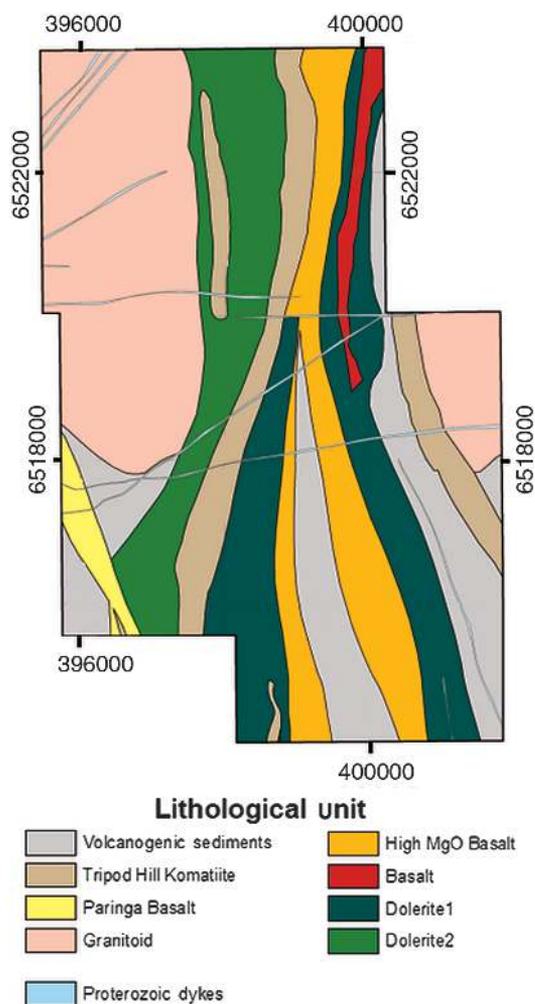


Figure 3. Heron South geology map. In subsequent figures, the lithologic units will be abbreviated as follows: Volcanogenic Sediments (VS), Tripod Hill Komatiite (THK), Paringa Basalt (PB), Granitoid (G), High MgO Basalt (HMgOB), Basalt (B), Dolerite 1 (D1), and Dolerite 2 (D2). The map extent in this figure defines the extent of all subsequent map figures in this paper.

possible  $H$ . For example, for a pixel with two possible and equally probable classes and a pixel with four possible and equally probable classes, they shall both be described as  $H$  being equal to 1.

## METHODS

### Data

In this study, 16 geophysical and remote-sensing data sets were used (Figure 5), and interpolated at a grid cell size appropriate (20%–25%) to their respective acquisition line spacing (Table 1). Landsat thematic mapper Landsat Program (2003) and Shuttle Radar Topography Mission (SRTM) products (United States Geological Survey, 2003) were procured in raster format, and their original point separation specifications were preserved (United States Geological Survey [2003] and United States Geological Survey [2006], respectively). Each data set was resampled to a 30 m grid to populate a matrix where each line takes the form:  $x, y, p_1, p_2, \dots, p_n$ , where  $x$  and  $y$  are the spatial coordinates and  $p$  are the various measured properties at each pixel. At the extent of the study area, this comprised approximately 56,000 samples. The compiled data were split into subsets comprising training ( $T_a$ ) and test ( $T_b$ ) data through a process of stratified spatially random sampling. One hundred samples were taken from each of the eight lithologic classes comprising the study area. These 800 samples comprising  $T_a$  represent approximately 1.6% of the total data set (Figure 6). The remaining 98.4% of data  $T_b$  were not shown to the classifier during the training process.

### Variable ranking and selection

RF facilitates several means of ranking the importance of input variables. In this instance, each variable was permuted and the effect on the out-of-bag classification accuracy was measured. Those variables which, when permuted, produced the greatest change to classification accuracy were ranked highest (Table 2). Due to the relatively small number of data sets used in this study, none of the starting input variables were sufficiently well-correlated (as defined by a threshold at a Pearson's correlation coefficient = 0.85) with one another to warrant removal, due to the duplication of information prior to ranking. To optimize the speed and interpretability of results, redundant variables were screened at this stage. Using  $T_a$ , variables were successively added to the classification according to their ranked importance established in the prior step. Accuracy was assessed using a forest comprising 500 classification trees, via 10-fold cross validation (Table 2). The cross-validation accuracy improved with the input of additional variables, albeit at a diminishing rate, until a peak cross-validation accuracy of 79% was achieved via the inclusion of variables ranked one to eight (Table 2). Beyond this point, no increase in cross-validation

accuracy was observed through the inclusion of additional variables; as such, the Landsat data, ranked 9th to 15th, were omitted. This is logical given the sensitivity of reflectance methods to the immediate surface in an area heavily influenced by transported cover. Easting and northing were omitted at this stage to avoid over fitting to the classification based on position.

### Classification and uncertainty

Eight hundred samples comprising 100 from each of the lithologic units defined above (Figure 6) were used to train an RF classifier. Each sample was attributed with the eight nonredundant variables identified during variable ranking. We used an RF comprising 500 trees with no limits on individual tree depth or subsequent pruning. The RF produced under these parameters required 12 s to train. Subsequently, the remaining data comprising  $T_b$ , which do not have an associated class, were shown to the trained classifier and a class prediction for each was made. Class-membership probabilities,

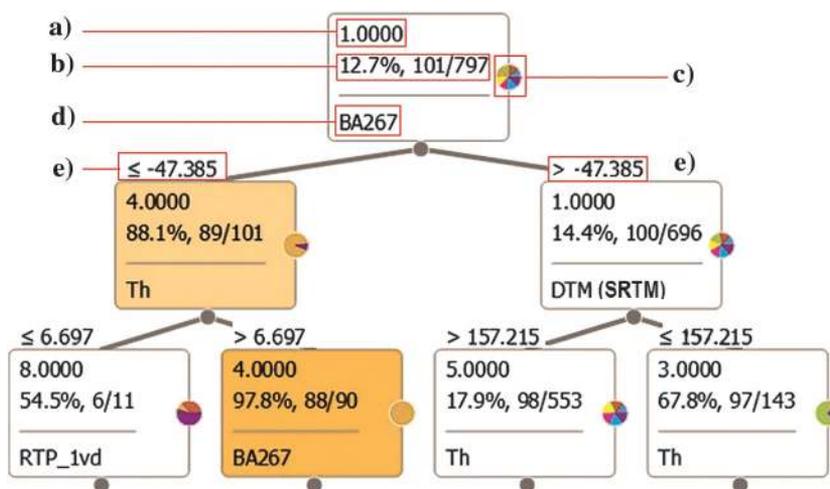


Figure 4. An example showing three levels of a classification tree, showing at each node: (a) the most numerous class, (b) the proportion of samples of the most numerous class, relative to all samples in the node (shown as percentage and count of total), (c) the pie-graph distribution of all classes present, (d) the variable used to split the parent node into child nodes, and (e) the threshold at which that split was executed.

**Table 1. Geophysical and remote sensing data sets used in study, including abbreviations and spatial resolution.**

Data set	Abbreviation	Spacing
Gravity (Bouguer anomaly)	BA267	200 × 200 m
First vertical derivative of gravity	BA267_1vd	200 × 200 m
Airborne magnetics — RTP	RTP	50 m EW flight lines
First vertical derivative of airborne magnetics	RTP_1vd	50 m EW flight lines
Airborne mass spectrometry: Potassium	K	50 m EW flight lines
Airborne mass spectrometry: Thorium	Th	50 m EW flight lines
Airborne mass spectrometry: Uranium	U	50 m EW flight lines
Elevation (digital terrain model)	SRTM	90 m pixel
Landsat thematic mapper: Channels 1–8	LSb1-8	30 m pixel

describing the proportion of trees voting for each class, were retained for the calculation and assessment of  $H$ .

### RESULTS

RFs produced a new version of the geologic map (Figure 7a), correctly predicting mapped geology in 76.8% of  $T_b$  instances. The remainder of the samples can be categorized either as incorrect predictions or as showing new information not previously mapped, or incorrectly mapped in the starting product. When plotted, class probabilities produced by RF (e.g., Figure 7b–7d) show the spatial distribution of lithology-dependent class-membership probabilities. Areas where a class has a very high probability of occupying an area with little likelihood of another class being present such as, for example, the central zone of D2 (Figure 7c) are apparent. There are, however, regions where multiple classes compete such that the class that ultimately is predicted displays a marginally higher probability than its competition (e.g., Figure 7b and 7d).

The confusion matrix in Table 3 indicates, on a per-class basis, the distribution of correct and incorrect classification percentages

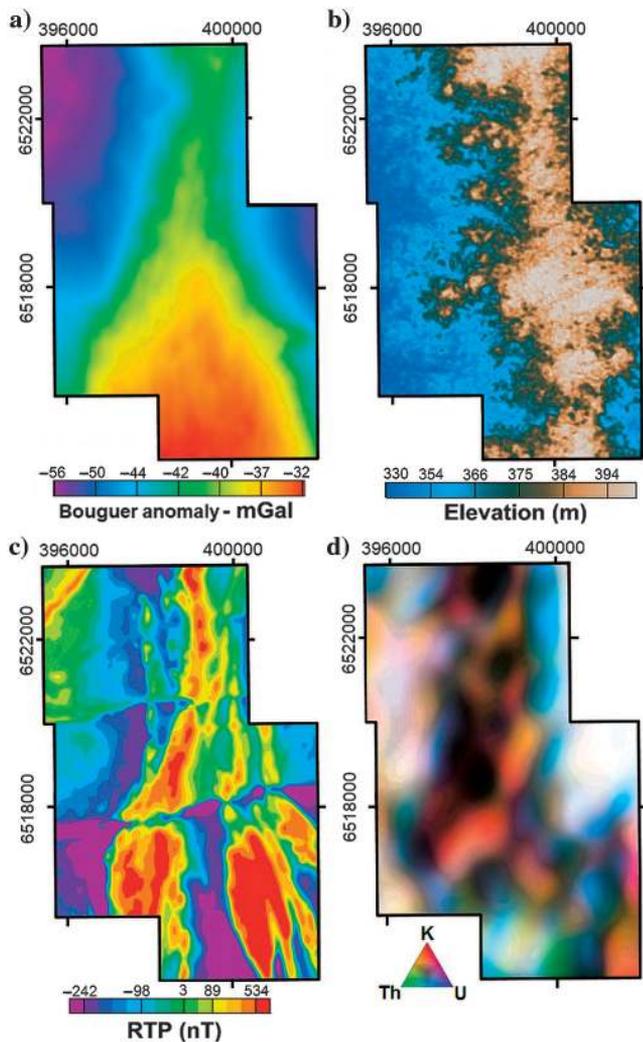


Figure 5. Examples of input data: (a) Bouguer anomaly, (b) elevation, (c) RTP total magnetic intensity, and (d) ternary radiometric image.

with respect to all other classes. Several classes, namely, the basaltic and granitic units, have been predicted with a high degree of accuracy. One of the doleritic units (D2) is commonly classified by RF as basalt or high-MgO basalt. This suggests that either the classification was incorrect in this instance or, alternatively, areas mapped as dolerite are in fact basalt. There is spatial control on classification accuracy with misclassification more likely when units with similar petrophysical properties occur adjacent to one another. The overlapping petrophysical signals of these classes, particularly in the case of potential field data due to smooth transitions as opposed to sharp boundaries, may be contributing to a reduced ability to make accurate predictions. This is particularly notable where these classes occupy the same areas of the map suggesting that the similarity of properties and spatial proximity are factors.

The spatial distribution of  $H$  (calculated using equation 3) shows very few examples in which a candidate class has a 0 probability of occurrence in a given pixel. By definition, this means that it must be included as a term in the calculation of  $H$ , mitigating the ability to display the monotonic increase in  $H$  that additional possible classes impose. As such, a threshold probability of 2% was selected, below which a class can be considered, for this purpose, to be not present in that pixel. The calculation of  $H$  with this parameter imposed was used to produce a map of the spatial distribution of  $H$  (Figure 8a). Areas in the central north and southwest of the project display the highest  $H$ , indicating that these areas are characterized by a high level of uncertainty across multiple classes that display a relatively

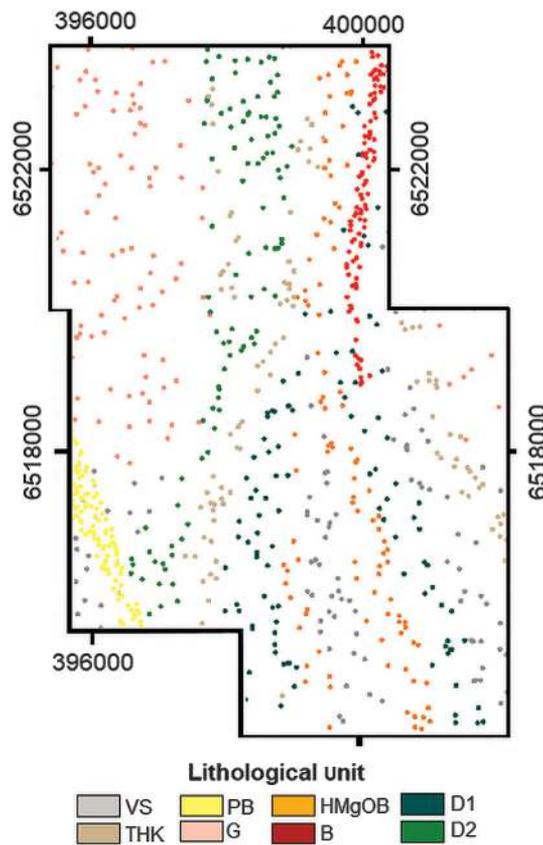


Figure 6. The  $T_a$  location coded by lithology. Note that the sample point diameter has been enlarged by a factor of five for legibility. Legend abbreviations are as described in Figure 3.

high probability of being predicted. Conversely, areas in the east and west of the project extent that are classified as granite coincide with low  $H$ , indicating that, RF classifications can be treated with a high degree of confidence such that no other classes have a high probability of being present. When normalized for the number of possible classes,  $H$  represents the relative minimum to maximum possible  $H$  on a per-pixel basis (Figure 8b). There is a direct relationship between normalized  $H$  and the observed discrepancies between the interpretation map and that produced by RF. This correlation can be qualitatively observed in a visual comparison of Figure 8b and 8d, and it was confirmed quantitatively by Kuhn et al. (2016), who demonstrate statistically distinct populations of  $H$  corresponding to correctly and incorrectly classified sample groups. Both  $H$  and normalized  $H$  can potentially form the basis of the assessment of the quality of RF predictions in the absence of a starting map with which to compare.

## DISCUSSION

In the absence of the information that indicates orogenic gold mineralization directly, the ability to map and interpret geology accurately is a key feature in target identification and the establishment of priority areas for exploration. We have demonstrated in this study that RF was able to classify lithology with an accuracy of approximately 76% relative to an existing interpreted geologic map using only 2% of the available data as training samples. These results are comparable with those achieved by Cracknell et al. (2014), who use a similar approach, achieving 78% accuracy, and

**Table 2. Variable importance rankings as determined by RF and cross-validation accuracy (CV acc). Cross-validation accuracy indicates the accuracy achieved when the corresponding variable is added in addition to higher ranked variables. Abbreviations are as per Table 1. The bold text indicates the first occurrence of peak cross-validation accuracy corresponding to variables selected for classification.**

Variable	Score RF	Rank	CV acc (%)
BA267	20.61	1	35.1
Th	16.54	2	62.1
DTM	14.37	3	69
RTP	12.14	4	74.5
U	7.83	5	76.6
Ba267_1VD	7.18	6	78.4
K	2.69	7	79.5
RTP_1VD	1.52	<b>8</b>	<b>79.8</b>
LSb3	0.97	9	79.7
LSb7	0.69	10	79.1
LSb4	0.61	11	79
LSb5	0.58	12	79.8
LSb2	0.37	13	79.6
LSb8	0.26	14	79.8
LSb1	0.24	15	79.5
LSb6	0.19	16	79.4

they compare favorably with similar implementations using SVM, such as by Yu et al. (2012), who achieve a consistency with the geology map of between 62.2% with a modal convolution filter applied. It is important, however, to note that different data and geologic conditions were encountered in each case. Nevertheless, the results of this study compare positively with similar applications in different settings.

Looking beyond bulk similarities, there is a wide range in performance with regard to predictive power of the RF as applied to individual classes. As shown in Table 3, the VS and D1 classes produced accuracies with respect to the starting geologic map, in the order of 59%, whereas the PB class exceeded 98%. It is likely that this excellent result is due to the spatially discrete and small area defined by the PB class, resulting in a very well-constrained class signature. The poor performance of the VS class is likely due to a highly variable class signature, the result of a wide range of sample locations and, potentially, misidentification in the original map. The D1 class was commonly confused with B (16.5% of instances)

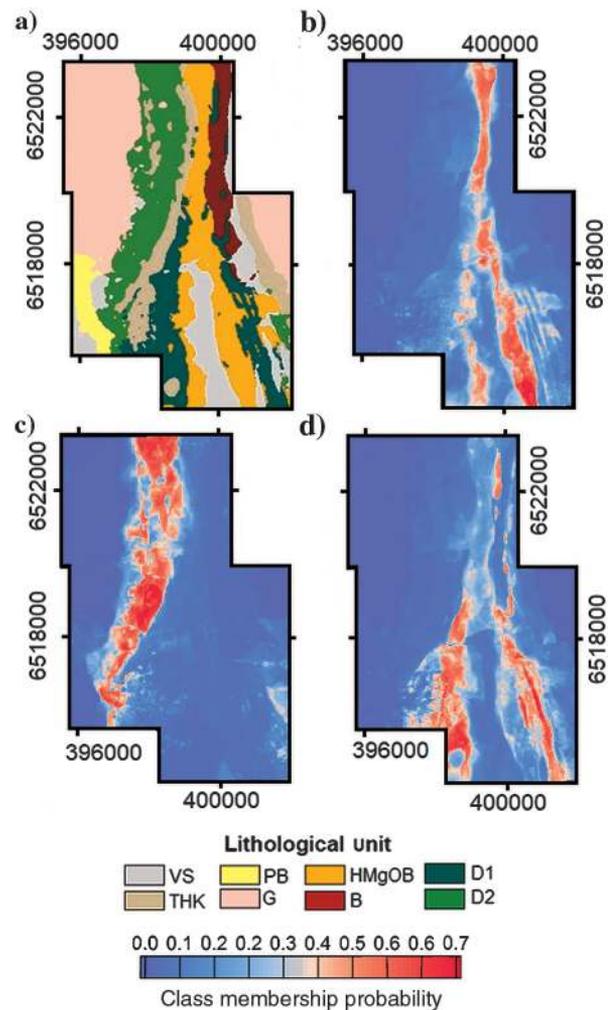


Figure 7. (a) The RF predicted Heron South geology, (b) probability of the THK class, (c) probability of the D2 class, and (d) probability of the D1 class. All class membership probabilities are presented on the same linear scale, shown at the bottom of the image. Lithology abbreviations are described in Figure 3.

and HMgOB (14.9%), which is logical, given the compositional similarity of these mafic units. The D2 class, however, while quite accurately captured at 81.8% was confused most commonly with the THK class at a rate of 10.5% indicating the possibility of unmapped ultramafic material interspersed in the region mapped as D1, or conversely, doleritic intrusions in the THK. Alternatively, this could indicate erroneous mapping of these units in the original geologic interpretation map. The G class was most often confused with the D2 class. This is explained by erroneous mapping in the starting map being repartitioned into the D2 class, which RF extends further to the west, supported by the expression of  $H$  in that region (Figure 8c).

An important component of these results was the observation that RF was able to preserve class labels defined from stratigraphic relationships and distinguishes between equivalent lithologies. In this case, the stratigraphic sequence is not well-characterized and geochemical data were not available to resolve this distinction. Geologic interpretations indicate that multiple dolerites and basalts are present in this region. The contrast between greenstone, felsic to intermediate intrusive bodies, and sedimentary packages is well-expressed in the gravity and magnetic data sets facilitating mapping using these variables via machine learning. It is, however, difficult to distinguish between units of similar composition using these data sets alone. Nevertheless, RF is able to capture this distinction, to the extent that it was present in the training data, and produce a map retaining stratigraphy and not simply amalgamating by rock type. Results produced by RF do not indicate a large-scale revision to the mapping or understanding of the structure in the area. Updates to lithologic boundaries could form the basis of an adjustment to the position of faults subparallel to stratigraphy and those that offset stratigraphy. Knowledge of the position within the stratigraphic column is important in an exploration context given that several models for the stratigraphic position of favorable host units, relative to the timing of gold deposition, have been identified. Again, this is contingent on the congruency of the sampled region. We suggest that

when using geophysical data, the accuracy of RF lithologic predictions cannot be assumed to apply to adjacent terrains. Potential field data in particular are influenced by effects such as cover depth, or the response of deeper sources can produce a shift in absolute signal amplitude, not related to geology as mapped at the surface. As such, the rules defined by RF are only reliably applicable to the domain and from which they were derived. Radiometric data are indicative of surficial features and may be mirrored in adjacent or distant domains, however, it is also likely that these data may be influenced by weathering and vegetation, which differs from the study area. In any event, it is not anticipated that radiometric data alone would be sufficient to propagate mapping to greater distances beyond the sampled region. Our approach is designed as a pragmatic workflow; however, further insights might be gained by more geostatistical- or computer-science-oriented practitioners (e.g., Grunsky and Kjaarsgaard, 2016).

It is important to note that regardless of the physical response, elevation, depth to source, or height of the sensor of a method, RF will preferentially use whichever variables allow the algorithm to most accurately solve the given problem, in this case, lithology. The data sets which are ranked highest and the associated frequency response are entirely determined by those that allow RF to discriminate between the lithologies.

The topographic (SRTM) data set ranked highly among the available input data. Given the contiguity and dominant strike of the geology relative to topography, it is possible that topography is, in fact, serving as a proxy for lithologic position in the landscape. It is also probable that rock composition is one of the controlling factors in preferential weathering and hence topography, although this relationship is not always obvious in the region.

The Bouguer anomaly and reduced-to-pole (RTP) total magnetic intensity data sets were both ranked as more important to the classification than their first vertical derivatives. The most plausible interpretation of this result being that the potential field data are more closely related to rock composition at the scale of this study. The

respective derivatives may define detailed features of the units that could reflect structural or compositional variability. This information is of immense value in accurately mapping and interpreting the regional and within-unit structural complexity of the area, but it does not necessitate a change to the lithologic class at any given location. Should the mapping area be expanded, the effects of regional trends would become more significant with derivatives, as a form of high-pass filter, being required to mitigate the influence of these trends and thus would likely be ranked of higher importance. It is possible that the introduction of additional, textural data, derived from those data sets could have improved results. It is worth noting, however, that of key importance is the ease of use of the method by geoscientists and as such we consider this a good demonstration of the method using readily available data sets, accessible to most projects without additional prerequisite knowledge of geographic information systems (GIS) operations.

The value  $H$  provided an indication of those areas where an operator can be confident of

**Table 3. Confusion matrix comparing mapped class with RF predictions. Values are shown as a percentage of the number of samples of a class present in the interpretation map. The red, yellow, and blue text indicates a recall greater than 50%, 70%, and 80%, respectively.**

		Predicted							
		VS	THK	PB	G	HMgOB	B	D1	D2
Observed	VS	59	3.9	5	3.6	13	2.6	3.6	9.4
	THK	1.8	71.5	0.3	0.9	7.5	0.2	8	9.8
	PB	0.9	0	98.7	0	0	0	0	0.4
	G	1.2	0.4	0.6	89.8	0	0	0	8.1
	HMgOB	7.3	0	0	0	85.4	2.2	5.1	0
	B	0	0.2	0	3.6	0	95.7	0.5	0
	D1	5.8	3.4	0	0.2	14.9	16.5	59.1	0
	D2	0.8	10.5	2	2.2	0.1	0	2.7	81.8



accurate mapping and those areas where they are more likely to be incorrect. Consistent with prior research (Cracknell and Reading, 2013; Cracknell, 2014), high uncertainty was generally observed in proximity to lithologic boundaries and areas of geologic complexity. Kuhn et al. (2016) demonstrate statistically, through examination of the distribution of normalized  $H$  of correctly and incorrectly classified (Figure 8f) samples that  $H$  provides a good, albeit imperfect, proxy for inaccuracy. As such,  $H$  is a valuable tool when mapping in unknown areas and where validation against a known result is not possible. Performing any exploration activity requiring fiscal expenditure through a decision unknowingly underpinned by a type II statistical error in classification has a greater consequence than performing additional study on an area that in fact was mapped correctly. Displaying  $H$  highlights areas that

require additional data collection, such that geoscientists can further validate these areas to within the scope of reasonable due diligence prior to additional expenditure. Conversely, areas producing low  $H$  do not require the same level of attention, and, as such, effort need not be expended here and can be diverted to those areas of higher uncertainty. We believe that  $H$  is therefore a valuable mechanism for quantifying uncertainty given that in addition to a normalized product, the purest form of  $H$  preserves monotonicity and provides a measure of the absolute uncertainty present throughout the classification.

The presence of highly magnetic Proterozoic dikes often confounds the ability to interpret Archaean stratigraphy. A manual interpreter may opt to attempt to see past these features in a somewhat subjective manner. It does, however, prohibit the use of absolute levels in classification of individual data sets, such as aeromagnetic imagery, when analyzing only that property, dikes are indistinguishable from other mafic units on a pixel-by-pixel basis. In this instance, our randomly selected training data included several samples of various rock units in the locations where they were intruded by Proterozoic dikes. Because this interaction was represented in the training data, RF was able to consistently map the underlying geologic class and was largely immune to the presence of these features. Looking at  $H$ , we can see that uncertainty does consistently increase by up to approximately 20% (Figure 8b) in areas where dikes intrude other lithologies; however, the correct decision has still been obtained.

It is assumed that classifications produced by RF are deemed incorrect in the event that they do not conform to the geologic map. An interpreted geologic map, however, is a constantly evolving product. The accuracy and level of detail of an interpreted geologic map improves as data of higher resolution and accuracy and better interpretation techniques become available. In a greenfield setting, where a geologic map is based on limited outcrop and interpretation of potential field data sets, it is entirely plausible that it contains errors and/or oversimplifications.

When RF produces a result that differs from the geologic map that the training data are sourced from,  $H$  provides a means to assert whether the RF output or the reference information are likely to be incorrect. In this instance (Figure 8e), we can see that the western boundary of the greenstone package is moved to the west relative to its position in the interpreted map. Low  $H$  at the original boundary suggests that RF predicted with high certainty that this was in fact an area of greenstone. In addition,  $H$  increases toward the predicted contact suggesting greater uncertainty as the transition between rock types was approached and the potential field signals “smear” (e.g., gravity decreasing toward the granitoid body). The relationship observed between the RF uncertainty and the distance to geologic boundaries is consistent with prior observations (e.g., Cracknell et al., 2014). A high  $H$  value is also observed in the southeast region of the study area. It is not possible to determine whether the interpretation map is incorrect; however, the RF classifications and high  $H$  suggest that this rock unit is significantly more complex than is shown. This is a clear example of the benefit of analysis of the RF classification in conjunction with uncertainty, and it may serve to optimize ongoing field efforts, either outcrop mapping or drilling as appropriate.

## CONCLUSION

This study demonstrates that RFs may be applied to reconnaissance-type geophysical data, in the absence of geochemistry, and

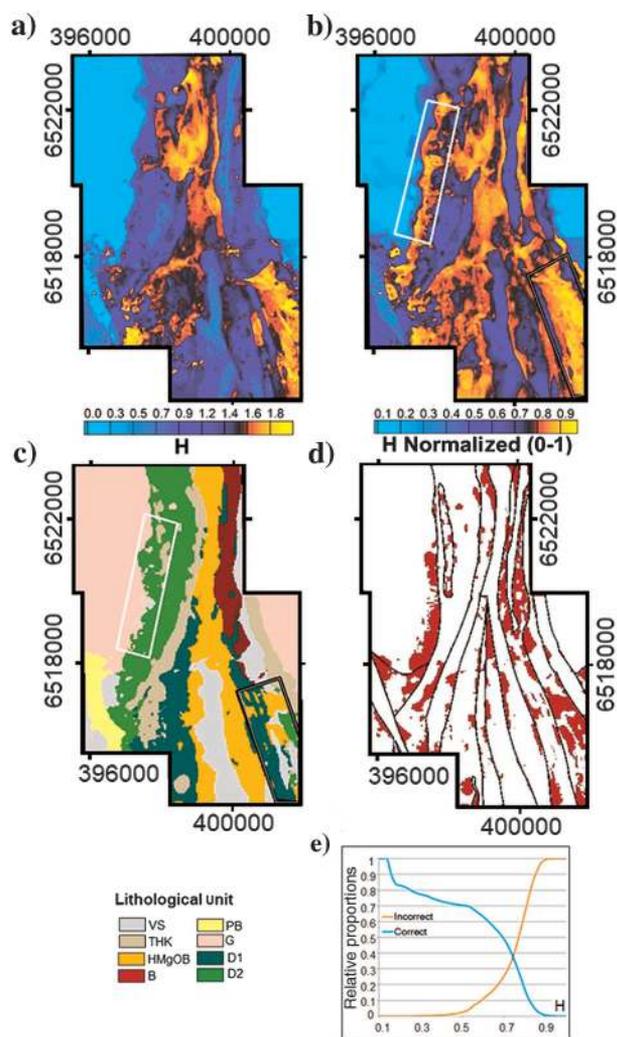


Figure 8. (a) The  $H$  (information entropy). (b)  $H$  normalized per pixel to 0–1. (c) Lithology predictions made by RF. (d) Accuracy relative to starting map (white = correct, red = incorrect). (e) The relative proportion of correctly (blue) and incorrectly (orange) classified samples (blue) at a given threshold of  $H$ . (b and c) The white box indicates a westward extension of D1 predicted by RF and associated high  $H$  increasing toward, and peaking at, the geologic boundary. (b and c) The white-black outlined box indicates a zone of potential geologic complexity associated with high  $H$ . (b-e) Modified from Kuhn et al. (2016).

produce sound lithologic predictions. There are two obvious applications for the use of RF for early-stage geologic mapping. The first is for the refinement of an existing geologic map. The second is for the production of a geologic map from a limited number of observations in the creation of a first-pass map. Sparse outcrop or a broad drilling campaign could provide such starting observations, provided the spatial distribution of the observations adequately samples the project area.

In this demonstration study, RF was able to preserve class labels, i.e., the stratigraphic context, where more than one class comprised the same lithology. This is an important outcome because the timing relationships between mineralization and various stratigraphy are vital information for mineral prospecting. Proterozoic dikes, which are petrophysically indistinguishable from Archean mafic rocks in the study area, confuse aeromagnetic interpretation. RF using a higher dimensional data space can deal with this complication, provided examples of the dikes overprinting the older stratigraphy are sampled in the training data.

Information entropy ( $H$ ) provides a valuable insight into the classification results. The highest  $H$  denotes areas of geologic/geometric complexity and proximity to lithologic boundaries. Where a predicted lithologic boundary significantly differs from the reference map, the behavior of  $H$  proximal to interpreted and predicted boundaries indicates which position is most probable. Statistically distinct populations in  $H$  correlate with correctly and incorrectly classified samples. Through understanding  $H$ , an optimal trade-off, retaining the greatest number of correct samples while discarding incorrect samples can be identified. Understanding the distribution of  $H$  for correct and incorrect sample populations allows a user to define an acceptable trade-off between discarding the maximal number of incorrectly classified samples or retaining a more complete, albeit a potentially less accurate, map. This will reflect the tolerance for risk of each individual explorer/company. The combination of RF classification and uncertainty appraisal allows explorers to critique quantitatively, the validity of map outputs — a quality control measure not available in conventional mapping.

## ACKNOWLEDGMENTS

We would like to thank Gold Fields Ltd. for access to data for the purpose of this study. S. Kuhn is supported by an Australian Postgraduate Award Scholarship from the University of Tasmania. This research was conducted in collaboration with the ARC Industrial Transformation Research Hub for Transforming the Mining Value Chain (project number IH130200004) at the Centre of Excellence in Ore Deposits, University of Tasmania. The views expressed herein are those of the authors and are not necessarily those of the Australian Research Council. We used the Orange software package (Demsar et al., 2013) for RF classification. Preprocessing, interpolation, and plotting were performed using Geosoft Oasis Montaj and ESRI ArcGIS. D. Douth is thanked for his input on the geologic and structural setting of the project. We thank the assistant editor, associate editor, and three reviewers for their suggestions, which have significantly improved the manuscript.

## REFERENCES

- Barnett, C. T., and P. M. Williams, 2009, Using geochemistry and neural networks to map geology under glacial cover: Geoscience BC Report, 2009-03, [http://www.geosciencebc.com/i/project\\_data/QUESTdata/GBCReport\\_2009-3/GBC\\_Report\\_2009-3.pdf](http://www.geosciencebc.com/i/project_data/QUESTdata/GBCReport_2009-3/GBC_Report_2009-3.pdf), accessed 15 November 2017.
- Blewett, R. S., R. Squire, J. M. Miller, P. A. Henson, and D. C. Champion, 2010, Architecture and geodynamic evolution of the St. Ives Goldfield, eastern Yilgarn Craton, Western Australia: *Precambrian Research*, **183**, 275–291, doi: [10.1016/j.precamres.2010.07.017](https://doi.org/10.1016/j.precamres.2010.07.017).
- Breiman, L., 1996, Bagging predictors: *Machine Learning*, **24**, 123–140.
- Breiman, L., 2001, Random forests: *Machine Learning*, **45**, 5–32, doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone, 1984, Classification and regression trees: Wadsworth International Group, Wadsworth & Brooks/Cole Statistics/Probability Series.
- Connors, K., E. Stolz, and J. E. Hanneson, 2002, Early fault architecture at St. Ives: Implications for Au and Ni mineralisation, in S. Vearncombe, ed., Applied structural geology for mineral exploration and mining: Australian Institute of Geoscientists, 29–31.
- Cox, S. F., and K. Ruming, 2004, The St. Ives mesothermal gold system, Western Australia: A case of golden aftershocks?: *Journal of Structural Geology*, **26**, 1109–1125, doi: [10.1016/j.jsg.2003.11.025](https://doi.org/10.1016/j.jsg.2003.11.025).
- Cracknell, M. J., 2014, Machine learning for geological mapping: Algorithms and applications: Ph.D. thesis, University of Tasmania.
- Cracknell, M. J., and A. M. Reading, 2013, The upside of uncertainty: Identification of lithology contact zones from airborne geophysics and satellite data using random forests and support vector machines: *Geophysics*, **78**, no. 3, WB113–WB126, doi: [10.1190/geo2012-0411.1](https://doi.org/10.1190/geo2012-0411.1).
- Cracknell, M. J., and A. M. Reading, 2014, Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information: *Computers and Geosciences*, **63**, 22–33, doi: [10.1016/j.cageo.2013.10.008](https://doi.org/10.1016/j.cageo.2013.10.008).
- Cracknell, M. J., A. M. Reading, and A. W. McNeill, 2014, Mapping geology and volcanic-hosted massive sulfide alteration in the Hellyer-Mt Charter region, Tasmania, using Random Forests™ and self-organizing maps: *Australian Journal of Earth Sciences*, **61**, 287–304, doi: [10.1080/08120099.2014.858081](https://doi.org/10.1080/08120099.2014.858081).
- Crawford, M., 2011, Dynamic coupling between deformation processes, fluid-rock interaction, and gold deposition in the Argo gold deposit, St. Ives, Western Australia: Ph.D. thesis, Australian National University.
- Demsar, J., T. Curk, A. Erjavec, C. Gorup, T. Hocevar, M. Milutinovic, M. Mozina, M. Polajnar, M. Toplak, A. Staric, M. Stajdohar, L. Umek, L. Zagar, J. Zbonar, M. Zitnik, and B. Zupan, 2013, Orange: Data mining toolbox in python: *Journal of Machine Learning Research*, **14**, 2349–2353.
- Grunsky, E. C., and B. A. Kjarsgaard, 2016, Recognizing and validating structural processes in geochemical data, in J. A. Martin-Fernandez and S. Thio-Hermestrosa, eds., *Compositional data analysis: Springer Proceedings in Mathematics and Statistics*, 85–116.
- Harris, J. R., and E. C. Grunsky, 2015, Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data: *Computers and Geosciences*, **80**, 9–25, doi: [10.1016/j.cageo.2015.03.013](https://doi.org/10.1016/j.cageo.2015.03.013).
- Hastie, T., R. Tibshirani, and J. H. Friedman, 2009, *The elements of statistical learning: Data mining, inference and prediction*: Springer, Springer series in statistics.
- Kuhn, S., M. J. Cracknell, and A. M. Reading, 2016, Lithological mapping via random forests: Information entropy as a proxy for inaccuracy: 25th International Geophysical Conference and Exhibition, ASEG, Extended Abstracts, 1–4.
- Miller, J., R. Blewett, J. Tunjic, and K. Connors, 2010, The role of early formed structures on the development of the world class St. Ives Goldfield, Yilgarn, WA: *Precambrian Research*, **183**, 292–315, doi: [10.1016/j.precamres.2010.08.002](https://doi.org/10.1016/j.precamres.2010.08.002).
- Ngyuen, T. P., 1997, Structural controls on gold mineralisation at the revenge mine and its tectonic setting in the Lake Lefroy Area, Kambalda, Western Australia: Ph.D. thesis, University of Western Australia.
- Rodriguez-Galiano, V. F., M. Chica-Olmo, and M. Chica-Rivas, 2014, Predictive modeling of gold potential with the integration of multisource information based on random forest: A case study on the Rodalquilar area, southern Spain: *International Journal of Geographical Information Science*, **28**, 1336–1354, doi: [10.1080/13658816.2014.885527](https://doi.org/10.1080/13658816.2014.885527).
- Ruming, K. J., 2006, Controls on lode gold mineralisation in the victory thrust complex, St. Ives Goldfield, Western Australia: Ph.D. thesis, University of Newcastle.
- Shannon, C. E., 1948, A mathematical theory of communication: *Bell Systems Technical Journal*, **27**, 379–423, doi: [10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x).
- Squire, R. J., C. M. Allen, R. A. Cas, I. H. Campbell, R. S. Blewett, and A. A. Nemchin, 2010, Two cycles of voluminous pyroclastic volcanism and sedimentation related to episodic granite emplacement during the late Archean: Eastern Yilgarn Craton, Western Australia: *Precambrian Research*, **183**, 251–274, doi: [10.1016/j.precamres.2010.08.009](https://doi.org/10.1016/j.precamres.2010.08.009).
- Swager, C. P., 1997, Tectono-stratigraphy of Late Archaean Greenstone Terranes in the Southern Eastern Goldfields, Western Australia: *Precambrian Research*, **83**, 11–42, doi: [10.1016/S0301-9268\(97\)00003-X](https://doi.org/10.1016/S0301-9268(97)00003-X).

- United States Geological Survey, 2003, National Aeronautics and Space Administration Shuttle Radar Topography Mission, 3 Arc Second, Scene S32E121.SRTMGL3; USGS, Sioux Falls.
- United States Geological Survey, 2006, National Aeronautics and Space Administration Landsat Program, Landsat ETM+ scene L71008058\_05820031026, SLC-Off; USGS, Sioux Falls.
- Vapnik, V. N., 1995, *The nature of statistical learning theory*: Springer.
- Vapnik, V. N., 1998, *Statistical learning theory*: John Wiley & Sons Inc.
- Waske, B., J. A. Benediktsson, K. Arnason, and J. R. Sveinsson, 2009, Mapping of hyperspectral AVIRIS data using machine-learning algorithms: *Canadian Journal of Remote Sensing*, **35**, S106–S116, doi: [10.5589/m09-018](https://doi.org/10.5589/m09-018).
- Wellmann, J. F., and K. Regenauer-Lieb, 2012, Uncertainties have a meaning: Information entropy as a quality measure for 3-D geological models: *Tectonophysics*, **526-529**, 207–216, doi: [10.1016/j.tecto.2011.05.001](https://doi.org/10.1016/j.tecto.2011.05.001).
- Yu, L., A. Porwal, E. J. Holden, and M. C. Dentith, 2012, Towards automatic lithological classification from remote sensing data using support vector machines: *Computers & Geosciences*, **45**, 229–239, doi: [10.1016/j.cageo.2011.11.019](https://doi.org/10.1016/j.cageo.2011.11.019).