Computers & Geosciences 63 (2014) 22–33

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

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

Matthew J. Cracknell *, Anya M. Reading

University of Tasmania, CODES – ARC Centre of Excellence in Ore Deposits and School of Earth Sciences, Private Bag 126, Hobart, Tasmania 7001, Australia

ARTICLE INFO

Article history:
Received 20 December 2012
Received in revised form
15 October 2013
Accepted 17 October 2013
Available online 26 October 2013

Keywords:
Geological mapping
Remote sensing
Machine learning
Supervised classification
Spatial clustering
Spatial information

ABSTRACT

Machine learning algorithms (MLAs) are a powerful group of data-driven inference tools that offer an automated means of recognizing patterns in high-dimensional data. Hence, there is much scope for the application of MLAs to the rapidly increasing volumes of remotely sensed geophysical data for geological mapping problems. We carry out a rigorous comparison of five MLAs: Naive Bayes, *k*-Nearest Neighbors, Random Forests, Support Vector Machines, and Artificial Neural Networks, in the context of a supervised lithology classification task using widely available and spatially constrained remotely sensed geophysical data. We make a further comparison of MLAs based on their sensitivity to variations in the degree of spatial clustering of training data, and their response to the inclusion of explicit spatial information (spatial coordinates). Our work identifies Random Forests as a good first choice algorithm for the supervised classification of lithology using remotely sensed geophysical data. Random Forests is straightforward to train, computationally efficient, highly stable with respect to variations in classification model parameter values, and as accurate as, or substantially more accurate than the other MLAs trialed. The results of our study indicate that as training data becomes increasingly dispersed across the region under investigation, MLA predictive accuracy improves dramatically. The use of explicit spatial information generates accurate lithology predictions but should be used in conjunction with geophysical data in order to generate geologically plausible predictions. MLAs, such as Random Forests, are valuable tools for generating reliable first-pass predictions for practical geological mapping applications that combine widely available geophysical data.

1. Introduction

Machine learning algorithms (MLAs) use an automatic inductive approach to recognize patterns in data. Once learned, pattern relationships are applied to other similar data in order to generate predictions for data-driven classification and regression problems. MLAs have been shown to perform well in situations involving the prediction of categories from spatially dispersed training data and are especially useful where the process under investigation is complex and/or represented by a high-dimensional input space (Kanevski et al., 2009). In this study we compare MLAs, applied to the task of supervised lithology classification, i.e. geological mapping, using airborne geophysics and multispectral satellite data. The algorithms that we evaluate represent the five general learning strategies employed by MLAs: Naive Bayes (NB) – statistical learning algorithms, *k*-Nearest Neighbors (kNN) – instance-based learners, Random Forests (RF) – logic-based learners, Support Vector Machines (SVM), and Artificial Neural Networks (ANN) – Perceptrons (Kotsiantis, 2007).

The basic premise of supervised classification is that it requires training data containing labeled samples representing what is known about the inference target (Kotsiantis, 2007; Ripley, 1996; Witten and Frank, 2005). MLA architecture and the statistical distributions of observed data guides the training of classification models, which is usually carried out by minimizing a loss (error) function (Kuncheva, 2004; Marsland, 2009). Trained classification models are then applied to similar input variables to predict classes present within the training data (Hastie et al., 2009; Witten and Frank, 2005).

The majority of published research focusing on the use of MLAs for the supervised classification of remote sensing data has been for the prediction of land cover or vegetation classes (e.g., Foody and Mathur, 2004; Ham et al., 2005; Huang et al., 2002; Pal, 2005; Song

* Corresponding author. Tel.: +61 3 6226 2472; fax: +61 3 6226 7662.
E-mail addresses: M.J.Cracknell@utas.edu.au (M.J. Cracknell), anya.reading@utas.edu.au (A.M. Reading).

0098-3004
http://dx.doi.org/10.1016/j.cageo.2013.10.008

et al., 2012; Waske and Braun, 2009). These studies use multi or hyperspectral spectral reflectance imagery as inputs and training data is sourced from manually interpreted classes. MLAs such as RF, SVM and ANN are commonly compared in terms of their predictive accuracies to more traditional methods of classifying remote sensing data such as the Maximum Likelihood Classifier (MLC). In general, RF and SVM outperform ANN and MLC, especially when faced with a limited number of training samples and a large number of inputs and/or classes. Previous investigations into the use of MLAs for supervised classification of lithology (e.g., Leverington, 2010; Leverington and Moon, 2012; Oommen et al., 2008; Waske et al., 2009; Yu et al., 2012) focus on comparing MLAs, such as RF and/or SVM, with more traditional classifiers.

Common to all remote sensing image classification studies is the use of geographical referenced input data containing co-located pixels specified by coordinates linked to a spatial reference frame. Despite this, inputs used in the majority of studies cited do not include reference to the spatial domain. This is equivalent to carrying out the classification task in geographic space where samples are only compared numerically (Gahegan, 2000). To date few investigations have evaluated the performance of MLAs in conjunction with the input of spatial coordinates. Kovacevic et al. (2009), for example, investigated the performance of SVM using Landsat ETM+ multispectral bands and spatial coordinates, concluding that, given training data of suitable quality, there was sufficient information in the spatial coordinates alone to make reliable predictions. However, when applying trained classification model to regions outside the spatial domain of the training data the information in Landsat ETM+ bands became increasingly important.

Our supervised lithology classification example evaluates MLA performance in the economically important Broken Hill area of western New South Wales, a region of Paleoproterozoic metasedimentary, metavolcanic and intrusive rocks with a complex deformation history. In this study, we use airborne geophysics and Landsat ETM+ imagery to classify lithology. Airborne geophysics, unlike satellite spectral reflectance imaging, it is not affected by cloud and/or vegetation cover and represents the characteristics of surface and near surface geological materials (Carneiro et al., 2012). Landsat ETM+ data are freely available and have extensive coverage at medium resolutions over large regions of the globe (Leverington and Moon, 2012). Although hyperspectral data has been shown to generate excellent results in sparsely vegetated regions due to high spectral and spatial resolutions (Waske et al., 2009) this data is limited in its coverage and ability to penetrate dense vegetation for the characterization of geological materials (Leverington and Moon, 2012).

We explore and compare the response of MLAs to variations in the spatial distributions and spatial information content of training data, and their ability to predict lithologies in spatially disjoint regions. We facilitate this comparison by conducting three separate experiments: (1) assessing the sensitivity of MLA performance using different training datasets on test samples not located within training regions; (2) random sampling of multiple training datasets with contrasting spatial distributions; and (3) using three different combinations of input variables, *X* and *Y* spatial coordinates (*XY* Only), geophysical data (i.e. airborne geophysics and Landsat ETM+ imagery; no *XY*), and combining geophysical data and spatial coordinates (all data). These experiments are combined to provide a robust understanding of the capabilities of MLAs when faced with training data collected by geologists in challenging field sampling situations using widely available remotely sensed input data.

1.1. Machine learning for supervised classification

Classification can be defined as mapping from one domain (i.e. input data) to another (target classes) via a discrimination function $y=f(\mathbf{x})$. Inputs are represented as d vectors of the form $\langle x_1, x_2, \ldots, x_d\rangle$ and y is a finite set of c class labels $\{y_1, y_2, \ldots, y_c\}$. Given instances of $\mathbf{x}$ and y, supervised machine learning attempts to induce or train a classification model f', which is an approximation of the discrimination function, $\hat{y}=f'(\mathbf{x})$, and maps input data to target classes (Gahegan, 2000; Hastie et al., 2009; Kovacevic et al., 2009). In practice, as we only have class labels for a limited set of data, T, it is necessary to divide available data into separate groups for training and evaluating MLAs. Training data, T_a, are used to optimize and train classification models via methods such as cross-validation. Test data, T_b, contains an independent set of samples not previously seen by the classifier and is used to provide an unbiased estimate of classifier performance (Hastie et al., 2009; Witten and Frank, 2005).

The task of MLA supervised classification can be divided into three general stages, (1) data pre-processing, (2) classification model training, and (3) prediction evaluation. Data pre-processing aims to compile, correct, transform or subset available data into a representative set of inputs. Pre-processing is motivated by the need to prepare data so that it contains information relevant to the intended application (Guyon, 2008; Hastie et al., 2009).

MLAs require the selection of one or more algorithm specific parameters that are adjusted to optimize their performance given the available data and intended application (Guyon, 2009). With only T_a available for training and estimating performance, the use of techniques such as k-fold cross-validation is required. Trained classification model performance is usually estimated by summing or averaging over the results of k folds. Parameters that generate the best performing classifier, given the conditions imposed, are used to train a MLA using all of the samples in T_a (Guyon, 2009; Hastie et al., 2009).

An unbiased evaluation of the ability of MLAs to classify samples not used during training, i.e. to generalize, is achieved using T_b (Hastie et al., 2009; Witten and Frank, 2005). Classifier performance metrics such as overall accuracy and kappa (Lu and Weng, 2007) are easily interpretable and commonly used measures of MLA performance for remote sensing applications (Congalton and Green, 1998).

1.2. Machine learning algorithm theory

1.2.1. Naive Bayes

Naive Bayes (NB) is a well known statistical learning algorithm recommended as a base level classifier for comparison with other algorithms (Guyon, 2009; Henery, 1994). NB estimates class-conditional probabilities by "naively" assuming that for a given class the inputs are independent of each other. This assumption yields a discrimination function indicated by the products of the joint probabilities that the classes are true given the inputs. NB reduces the problem of discriminating classes to finding class conditional marginal densities, which represent the probability that a given sample is one of the possible target classes (Molina et al., 1994). NB performs well against other alternatives unless the data contains correlated inputs (Hastie et al., 2009; Witten and Frank, 2005).

1.2.2. k-Nearest Neighbors

The k-Nearest Neighbors (kNN) algorithm (Cover and Hart, 1967; Fix and Hodges, 1951) is an instance-based learner that does not train a classification model until provided with samples to classify (Kotsiantis, 2007). During classification, individual test samples are compared locally to k neighboring training samples in variable space. Neighbors are commonly identified using a Euclidian distance metric. Predictions are based on a majority vote cast by neighboring samples (Henery, 1994; Kotsiantis, 2007;

Witten and Frank, 2005). As high k can lead to over fitting and model instability, appropriate values must be selected for a given application (Hastie et al., 2009).

1.2.3. Random Forests

Random Forests (RF), developed by Breiman (2001), is an ensemble classification scheme that utilizes a majority vote to predict classes based on the partition of data from multiple decision trees. RF grows multiple trees by randomly subsetting a predefined number of variables to split at each node of the decision trees and by bagging. Bagging generates training data for each tree by sampling with replacement a number of samples equal to the number of samples in the source dataset (Breiman, 1996). RF implements the Gini Index to determine a "best split" threshold of input values for given classes. The Gini Index returns a measure of class heterogeneity within child nodes as compared to the parent node (Breiman et al., 1984; Waske et al., 2009). RF requires the selection of *mtry* which sets the number of possible variables that can be randomly selected for splitting at each node of the trees in the forest.

1.2.4. Support Vector Machines

Support Vector Machines (SVM), formally described by Vapnik (1998), has the ability to define non-linear decision boundaries in high-dimensional variable space by solving a quadratic optimization problem (Hsu et al., 2010; Karatzoglou et al., 2006). Basic SVM theory states that for a non-linearly separable dataset containing points from two classes there are an infinite number of hyperplanes that divide classes. The selection of a hyperplane that optimally separates two classes (i.e. the decision boundary) is carried out using only a subset of training samples known as support vectors. The maximal margin M (distance) between the support vectors is taken to represent the optimal decision boundary. In non-separable linear cases, SVM finds M while incorporating a cost parameter C, which defines a penalty for misclassifying support vectors. High values of C generate complex decision boundaries in order to misclassify as few support vectors as possible (Karatzoglou et al., 2006). For problems where classes are not linearly separable, SVM uses an implicit transformation of input variables using a kernel function. Kernel functions allow SVM to separate non-linearly separable support vectors using a linear hyperplane (Yu et al., 2012). Selection of an appropriate kernel function and kernel width, σ, are required to optimize performance for most applications (Hsu et al., 2010). SVM can be extended to multi-class problems by constructing $c(c\text{-}1)/2$ binary classification models, the so called one-against-one method, that generate predictions based on a majority vote (Hsu and Lin, 2002; Melgani and Bruzzone, 2004).

1.2.5. Artificial Neural Networks

Artificial Neural Networks (ANN) have been widely used in science and engineering problems. They attempt to model the ability of biological nervous systems to recognize patterns and objects. ANN basic architecture consists of networks of primitive functions capable of receiving multiple weighted inputs that are evaluated in terms of their success at discriminating the classes in T_a. Different types of primitive functions and network configurations result in varying models (Hastie et al., 2009; Rojas, 1996). During training network connection weights are adjusted if the separation of inputs and predefined classes incurs an error. Convergence proceeds until the reduction in error between iterations reaches a decay threshold (Kotsiantis, 2007; Rojas, 1996). We use feed-forward networks with a single hidden layer of nodes, a so called Multi-Layer Perceptron (MLP) (Venables and Ripley, 2002), and select one of two possible parameters: *size*, the number nodes in the hidden layer.

1.3. Geology and tectonic setting

This study covers an area of ~160 km^2 located near Broken Hill, far western New South Wales, Australia (Fig. 1). The geology of the Broken Hill Domain (BHD) (Webster, 2004) features an inlier of the Paleoproterozoic Willyama Supergroup (WSG) (Willis et al., 1983). WSG contains a suite of metamorphosed sedimentary,

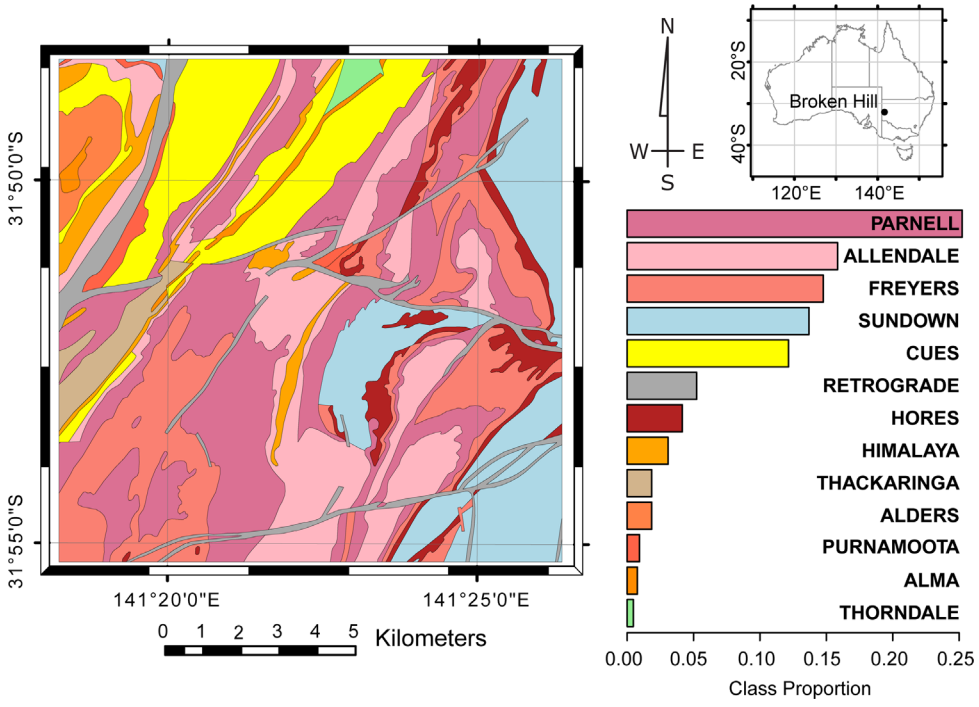

Fig. 1. Reference geological map (after Buckley et al., 2002) and associated class proportions for the 13 lithological classes present within the Broken Hill study area, modified from Cracknell and Reading (2013). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Summary of 13 lithological classes within the Broken Hill study region, complied using information from Willis et al. (1983) and Buckley et al. (2002).

Class Label	Class (undiff.)	Class	Description
Retrograde		Retrograde Schist	Retrograde micaceous schist, original stratigraphic position unknown.
Sundown		Sundown Group	Predominantly non-graphitic meta-sediment. Pelite–psammopelite units most abundant and more common in the lower half. Psammite–psammopelite units more common in the upper half.
Hores	Purnamoota	Hores Gneiss	Mainly garnet-bearing quartzo-feldspathic gneiss. Medium to fine-grained quartz-plagioclase-K-feldspar-biotite garnet gneiss.
Freyers		Freyers Metasediments	Mainly metasediments ranging from well bedded pelitic/psammopelitic schists, with some psammitic intervals, to psammopelitic or psammitic/pelitic metasediments.
Parnell		Parnell Formation	Extensive bodies of basic gneiss, lenticular masses of garnet-bearing quartzo–feldspathic gneiss, and "lode horizon" rocks, intercalated with pelitic to psammopelitic and psammitic metasediments.
Allendale		Allendale Metasediments	Mainly metasediment and metasedimentary composite gneiss. Variable ratio of pelite, psammopelite, psammite. Commonly garnet rich. Sporadic bodies of basic gneiss and quartz-gahnite.
Himalaya	Thackaringa	Himalaya Formation	Extensive bodies of medium-grained saccharoidal leucocratic sodic plagioclase quartz+K-feldspar biotite rocks, with variably developed interbedded metasedimentary composite gneiss, and basic gneiss.
Cues		Cues Formation	Mainly psammopelitic to psammitic composite gneisses or metasediments, with intercalated bodies of basic gneiss.
Alders		Alders Tank Formation	Consists largely of composite gneisses, with little or no basic gneiss, local minor plagioclase-quartz rocks and minor granular quartz-iron oxide/iron sulfide "lode" rocks.
Alma		Alma Gneiss	Mainly medium to coarse-grained quartz–feldspar–biotite+garnet gneiss with abundant to sporadic megacrysts of K-feldspar and plagioclase, or of aggregates of quartz+feldspar.
Thorndale		Thorndale Composite Gneiss	Mainly sedimentary quartz–feldspar–biotite–sillimanite ± garnet ± cordierite composite gneiss, consisting of interlayer psammite and psammopelite, generally with minor pelite and abundant pegmatitic segregations commonly disrupt bedding.

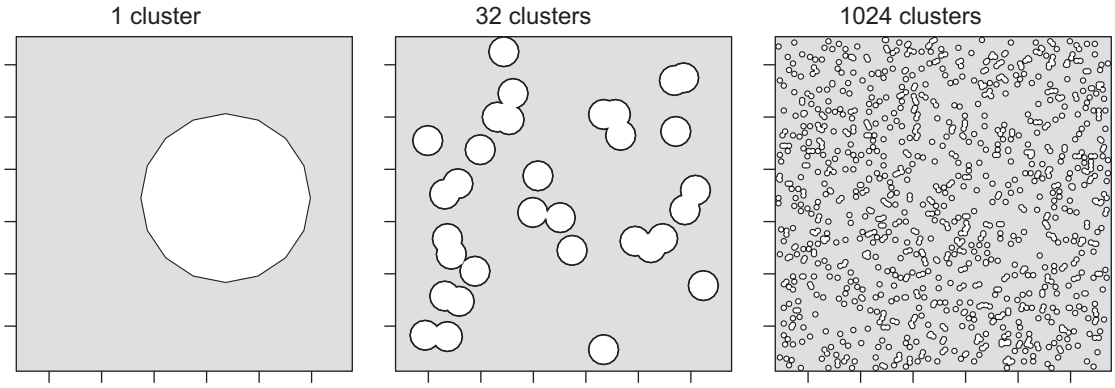

Fig. 2. Example of T_a spatial distributions for 1, 32 and 1024 clusters. Gray areas outside T_a regions were used to obtain T_b for MLA accuracy and kappa comparisons.

volcanic and intrusive rocks, deposited between 1710 and 1704 ± 3 Ma (Page et al., 2005a; Page et al., 2005b). WSG features complex lithology distributions resulting from a long history of folding, shearing, faulting and metamorphism (Stevens, 1986; Webster, 2004). BHD is of significant economic importance as it is the location of the largest and richest known Broken Hill Type stratiform and stratabound Pb–Zn–Ag deposit in the world (Webster, 2004; Willis et al., 1983).

Within the study area defined for this experiment are 13 lithology classes that form a chronological sequence younging from west to east. In general, WSG basal units are dominated by quartzo-feldspathic composite gneisses (i.e., Thorndale Composite Gneiss and Thackaringa Group), and are overlain by dominantly psammitic and pelitic metasedimentary rocks (i.e., Allendale Metasediments, Purnamoota Subgroup and Sundown Group). Table 1 provides a detailed summary of lithology classes present within the study area.

BHD deformation history can be summarized into four events (Stevens, 1986; Webster, 2004). The first two events of the Olarian Orogeny (1600–1590 Ma; Page et al., 2005a; Page et al., 2005b) are associated with amphibolite–granulite facies metamorphism and north-northeast to south–southwest trending regional fabric. A third event is characterized by localized planar or curviplanar zones of Retrograde Schist. These zones fulfill the role of faults in the BHD and display well developed and intense schistosity, strongly deformed metasediment bedding, and generally displace the units they intersect (Stevens, 1986). The fourth deformation event, associated with the Delamerian Orogeny (458–520 Ma) (Webster, 2004), is interpreted from gentle dome and basin structures.

2. Data

The published 1:250,000 Broken Hill digital geological map, compiled by Buckley et al. (2002), was used to obtain labeled samples representing lithology classes for training and to evaluate MLA predictions. We maintained the number of training samples at 10% (~6500) of the total number of samples in the selected scene. Multiple sets of T_a were randomly sampled from approximately circular regions, randomly seeded across the study area. The number of training regions were varied from 1 to 1024, such that the number of regions was equal to 2^a, where a represents sequential integers from 0 to 10. In all cases, the total coverage of training regions equates to $>10\%$ and $<20\%$ of the study area (Fig. 2).

Spatially disjoint T_b samples were sampled from all other pixels in the study area not contained within T_a, equating to $>80\%$ of the total number of samples. We randomly sample 10 sets of T_a and T_b for each combination of training clusters in order to avoid

any bias associated with individual groups of T_a or T_b. Due to the natural variability of lithological class distributions, T_a sampled from low numbers of training regions often did not contain representatives for all 13 lithological classes. This leads to biased accuracy assessments due to the incorporation of samples that cannot be accurately classified. Therefore, those classes not represented in T_a were eliminated from their corresponding T_b when evaluating MLA predictions.

Airborne geophysical data used in this study contained a Digital Elevation Model (DEM ASL), Total Magnetic Intensity TMI (nT), and four Gamma-Ray Spectrometry (GRS) channels comprising Potassium (K%), Thorium (Th ppm), Uranium (U ppm), and Total Count channels. The Landsat 7 ETM+ data contained 8 bands supplied with Level 1 processing applied. Landsat 7 ETM+ band 8, which covers the spectral bandwidths of bands 2, 3, and 4 (Williams, 2009) was not included.

3. Methods

3.1. Pre-processing

Geophysical data were transformed to a common projected coordinate system WGS84 UTM zone 54S, using bilinear interpolation. All inputs were resampled to a common extent (12.8 × 12.8 km) and cell resolution (50 m × 50 m), resulting in image dimensions of 256 × 256 pixels (65,536 samples). To enhance their relevance to the task of lithology discrimination, projected data were processed in a variety of ways specific to the geophysical property they represent. An account of pre-processing steps implemented to generate input data is provided in Supplementary Information Section 1 (*S1 – Data*). Spatial coordinates (Easting (m) and Northing (m)), obtained from the location of grid cell centers were included resulting in a total of 27 variables available for input. Processed input data were standardized to zero mean and unit variance. Highly correlated data, with mean Pearson's correlation coefficients >0.8 associated with a large proportion of other data, were eliminated resulting in a total of 17 inputs available for MLA training and prediction. Relative normalized variable importance was calculated by generating Receiver Operating Curves (ROC) (Provost and Fawcett, 1997) for all pair-wise class combinations and obtaining the area under the curve for each pair. The maximum area under the curve across all pair-wise comparisons was used to defined the importance of variables with respect to individual classes (Kuhn et al., 2012).

Table 2
MLA specific parameters evaluated during classifier training. Note RF parameters presented indicate those used for all input variables (All Data), when inputting only spatial coordinates (XY Only) there is only one possible *mtry* value (2).

MLA	Parameter	1	2	3	4	5	6	7	8	9	10
NB	*usekernel*	FALSE	TRUE	–	–	–	–	–	–	–	–
kNN	*k*	1	3	5	7	9	11	13	15	17	19
RF	*mtry*	2	3	5	6	8	9	11	12	14	16
SVM	*C*	0.25	0.5	1	2	4	8	16	32	64	128
ANN	*size*	5	8	11	13	16	19	22	24	27	30

3.2. Classification model training

Table 2 indicates the MLA parameter values assessed in this study. Optimal parameters were selected based on maximum mean accuracies resulting from 10-fold cross-validation. MLA classification models were trained using selected parameters on the entire set of samples in T_a prior to prediction evaluation. Information on the R packages and functions used to train MLA classification models and details regarding associated parameters are provided in Supplementary information Section (S2 – *MLA Software and Parameters*).

3.3. Prediction evaluation

Overall accuracy and Cohen's kappa statistic (Cohen, 1960) are commonly used to evaluate classifier performance (Lu and Weng, 2007). Overall accuracy treats predictions as either correct or incorrect and is defined as the number of correctly classified test samples divided by the total number of test samples. The kappa statistic is a measure of the similarity between predictions and observations while correcting for agreement that occurs by chance (Congalton and Green, 1998). We do not use the area under ROC to evaluate MLA predictions because multiclass ROC becomes increasing intractable with a large number of classes (i.e. >8) (Landgrebe and Paclik, 2010). We visualize the spatial distribution of prediction error and assess their geological validity by plotting MLA predictions in the spatial domain and by comparing the locations of misclassified samples.

4. Results

In this section we present the results of our comparison of the five MLAs trialed in this study. Initially, we assess the effect of changes in the spatial distribution of training data on the relative importance of input variables and MLA parameter selection. This is followed by a comparison of MLA test statistics in light of variations in the spatial distribution of training data and the inclusion of explicit spatial information.

Fig. 3 indicates variations in the importance of input variables with respect to changes in the spatial distribution of training

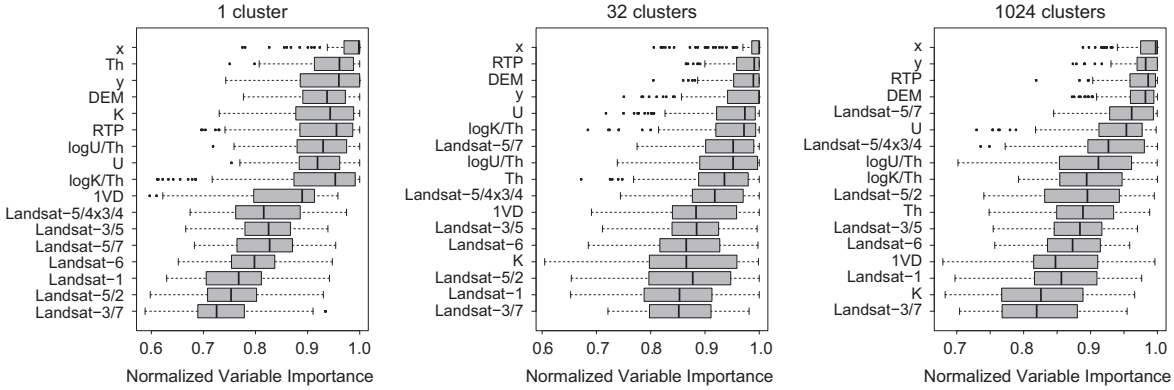

Fig. 3. Mean ranked normalized variable importance for 1, 32, and 1024 T_a clusters using all data after the removal of highly correlated variables. Boxplots indicate the distribution of combined class univariate importance across all 10 sets of T_a.

samples. A large range of importance suggests that the usefulness of a particular input at discriminating individual classes is highly variable. Top ranked variables with relatively narrow distributions include X and Y spatial coordinates, DEM and Reduced to Pole (RTP) magnetics. Less important inputs represent GRS channels and ratios with Landsat ETM+ band 5 variables. Lowest ranked variables are 1st Vertical Derivative (1VD) of RTP TMI, and the remaining Landsat ETM+ bands. As the number of training regions increases, GRS variables decrease in importance while ratios with Landsat ETM+ 5 increase in importance. The X coordinate is consistently the most important variable, reflecting the presence of approximately north–south trending geological structures.

Fig. 4 compares MLA cross-validation accuracy with respect to classification model parameter and number of T_a clusters. NB, SVM and ANN cross-validation accuracies are more sensitive to model parameters than to the number of T_a clusters. In contrast, RF and kNN cross-validation accuracies are relatively stable. Fluctuations in cross-validation accuracy are observed for all MLAs with respect to variations in T_a clusters, indicating the sensitivity of MLA classifiers to individual sets of T_a.

Fig. 5 provides MLA T_b accuracy comparisons with respect to numbers of T_a clusters across the three input variable combinations. All MLAs generate increasingly accurate predictions and exhibit less variation between different T_a for larger numbers of (smaller) T_a clusters. For ≤ 32 T_a clusters and when incorporating spatial coordinates, either exclusively or inclusively, substantial differences between MLA accuracies are not indicated. A <0.10 increase in T_b accuracy is observed when including spatial coordinates and training MLAs using >32 T_a clusters, compared to using only geophysical data as input. NB consistently generates the lowest mean T_b accuracies of all MLAs. Where the number of T_a clusters is >32 and when using only spatial coordinates, kNN, RF and SVM obtain T_b accuracies within 0.05 of each other up to a maximum accuracy of $\sim$0.70–0.75 when using 1024 clusters. When utilizing only geophysical data, and all available data (including spatial coordinates) RF generates substantially higher T_b accuracies than all other MLAs for T_a clusters >32.

Table 3 compares the mean and standard deviation of cross-validation and T_b accuracies and kappa statistics when using all input variables. MLAs exhibit a slight decrease in mean cross-validation accuracy (and kappa) and substantial increase in mean T_b accuracy (and kappa) with increasing numbers of T_a clusters. As the number of T_a clusters increases the difference between mean cross-validation accuracy (and kappa) and mean T_b accuracy (and kappa) decreases, indicating that trained classification models are increasingly over-fitted when presented with smaller numbers of T_a clusters. MLAs were unable to train classification models using the samples provided within T_a or due to cross-validation partitions not containing at least one sample from each class. SVM was more likely to encounter errors during training than other MLAs.

Figs. 6–8 compare the spatial distribution of MLA lithology predictions and location of misclassifications for selected numbers of T_a clusters. Where inputs consist only of spatial coordinates (Fig. 6), T_a must be distributed across the entire study area in order to train MLAs that generate predictions closely mimicking the spatial distribution of lithologies. Despite lower overall accuracies when provided with only geophysical data (Fig. 7), all MLAs generate lithological predictions that indicate major geological structures and trends (i.e. approximate locations of contacts between lithologies); however, the correct lithologies are not always identified. Nevertheless, a large amount of high frequency noise is present in all MLA predictions when spatial coordinates are excluded. Using training data dispersed across the study area and all available data (Fig. 8) not only generates geologically plausible predictions but the degree to which predictions are affected by high frequency noise is greatly reduced.

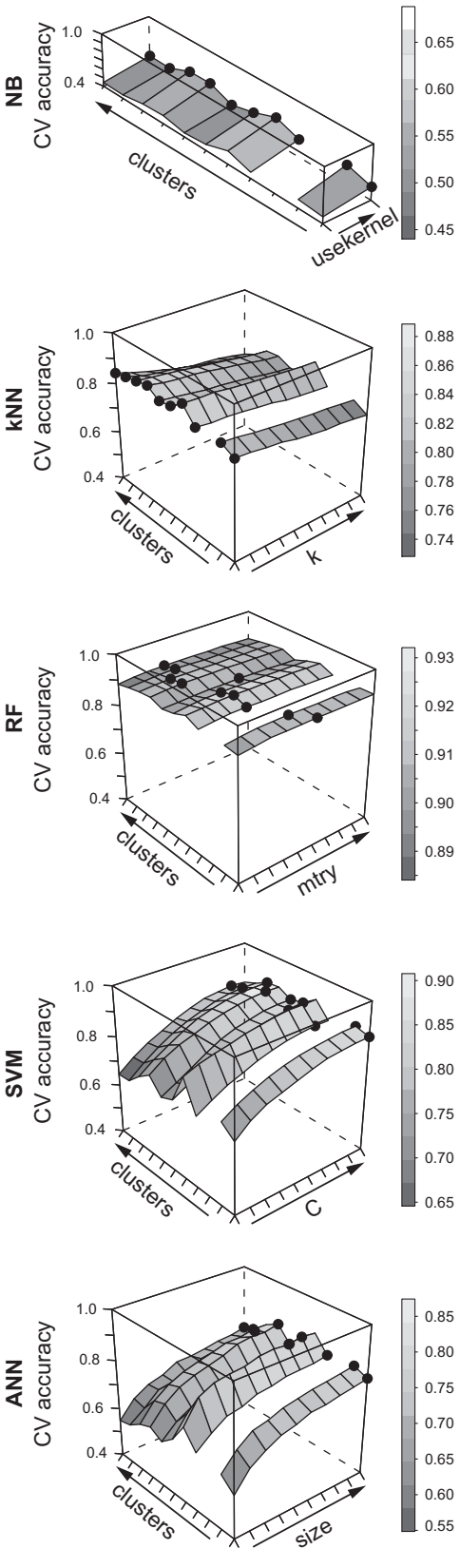

Fig. 4. Comparison of MLA cross-validation accuracies as a function of classification model parameter and number of T_a clusters. Selected parameter value = black circle. Refer to Table 2 for corresponding MLA parameter values.

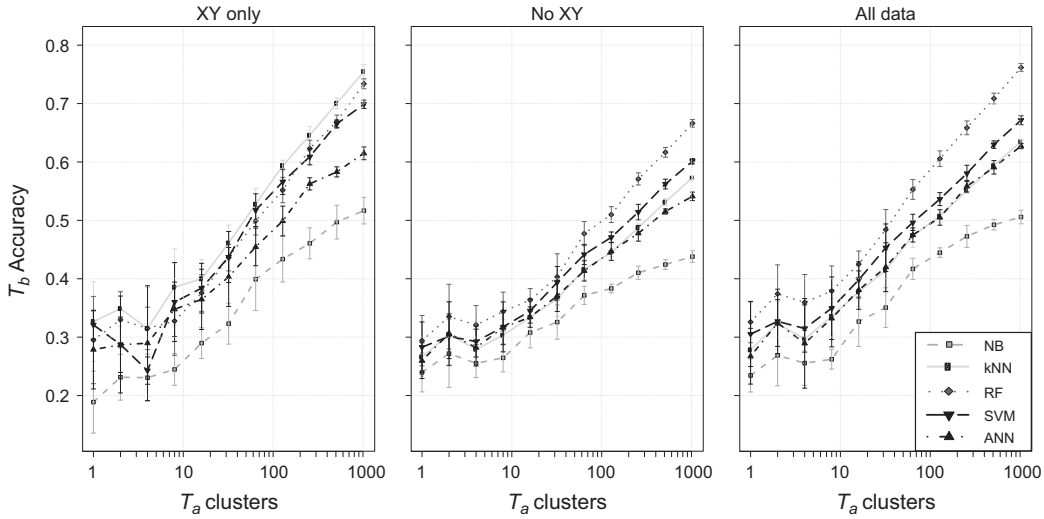

Fig. 5. Comparison of MLA mean T_b accuracy with respect to variations in the number of T_a clusters. Error bars represent one standard deviation from mean accuracy calculated from a maximum of 10 sets of T_b.

Fig. 9 provides a comparison of the processing time required to train MLAs and generate predictions given input variables and the spatial distribution of training samples. As the number of T_a clusters increases all MLAs incur additional computation cost. Overall, kNN takes consistently less and SVM consistently more time to run than other MLAs. Increases in processing time are observed when more variables are used as input, with NB and RF taking considerably longer to run. ANN processing times do not substantially increase when including additional data. In contrast to other MLAs, NB requires more time to generate predictions than to train classification models.

5. Discussion

5.1. Machine learning algorithms compared

Of the MLAs trialed in this study, RF performed well across several aspects evaluated, such as stability, ease of use, processing time and prediction accuracy. A strong aspect of RF performance was its relative insensitivity to variations in parameter values, indicating that RF reduces the potential for classifier over fitting (Breiman, 2001). The range of RF cross-validation accuracies were less than $\sim$0.05 with increasing *mtry* values. In contrast, SVM (*C*) and ANN (*size*) cross-validation accuracy ranges were $>$0.35 across all parameter values. The relative instability of SVM and ANN highlights the need to effectively search a large parameter space to optimize their performance. In this study, we have simplified comparisons by deliberately limiting the search to one parameter for SVM (σ estimated), and ANN (*decay* constant). In situations where multiple parameters require optimization grid searches can be employed. However, this can be very computationally expensive with fine search grids or, as in the case of coarse search grids, they may not be effective at identifying optimal parameters for a given dataset due to the effect of discretization (Guyon, 2009). Alternatively, a multi-stage nested search could be employed to reduce computational cost and improve parameter selection outcomes (Oommen et al., 2008). A nested search initially identifies a small region of parameter space using a coarse grid then searches this region using a finer grid of parameter values.

The least computationally costly MLA to implement was kNN. RF, ANN and NB each had similar processing times (when using geophysical variables) and took less time to run than SVM. Rankings of MLAs based on processing time were maintained regardless of an increase in the number of T_a clusters or the number of input variables. The use of spatially dispersed training samples resulted in a large increase in the already lengthy SVM processing times.

T_b accuracy and kappa results indicate that RF generated substantially higher (between $\sim$0.05 and 0.1) T_b mean accuracy and kappa than other MLAs when provided with geophysical data represented by spatially dispersed T_a. SVM, kNN and ANN all obtained T_b mean accuracies and kappa statistics within $\sim$0.05, while NB obtained significantly lower test accuracies compared to other MLAs. SVM was more likely to encounter convergence errors when training classification models as compared to other MLAs. The good performance of RF in these experiments suggests that its approach to learning, which resembles an adaptive kNN strategy (Hastie et al., 2009), provides advantages over other MLAs when applied to spatially constrained supervised classification problems. Furthermore, RF is generally insensitive to noisy input data where the number of relevant variables is comparatively large. In these situations, the effect of variations in the information content of T_a on RF classification models is reduced, improving RF generalization on unseen data (Breiman, 2001; Hastie et al., 2009).

5.2. Influence of training data spatial distribution

Previous studies into the application of MLAs to remote sensing supervised classification examined the influence of the number of samples in T_a on overall predictive accuracy (e.g., Gelfort, 2006; Ham et al., 2005; Oommen et al., 2008; Song et al., 2012). These studies found the minimum number of T_a samples required to induce accurate MLA classifiers needs to be between 10% and 25% of the total number of samples. Furthermore, increasing the number of samples in T_a did not lead to improved classification accuracy despite additional computational cost. The experiments conducted in this study clearly indicate that T_a spatial clustering is a limiting factor in the efficacy of MLA to predict spatially distributed phenomena while having little effect on processing time.

As samples within T_a become increasingly dispersed across the region of interest, classifier T_b accuracy (and kappa) improves dramatically. Moreover, the difference between cross-validation performance estimates and spatially disjoint T_b accuracy decreases, indicating better classifier generalization capabilities. Improvements in classifier generalization with highly dispersed T_a are coupled with reduced variability of MLA performance on individual sets of T_a. These findings

Table 3
Comparison of MLA cross-validation and T_b accuracy and kappa using all input variables with respect to different numbers of T_a clusters. T_a n indicates the number successfully trained classification models for a given run used to calculate mean and standard deviation. Selected parameters correspond to the MLA parameter (refer to Table 2) that obtained the maximum mean cross-validation accuracy and *count* indicates the number of times this parameter occurred.

T_a clusters	T_a n	Cross-Validation				Selected parameter	Count	T_b			
		Accuracy		Kappa				Accuracy		Kappa	
		Mean	Standard Deviation	Mean	Standard Deviation			Mean	Standard Deviation	Mean	Standard Deviation
NB											
1	9	0.618	0.069	0.527	0.073	TRUE	9	0.234	0.028	0.082	0.021
2	10	0.644	0.030	0.553	0.040	TRUE	10	0.269	0.052	0.122	0.038
4	9	0.620	0.070	0.534	0.086	TRUE	9	0.255	0.038	0.104	0.032
8	8	0.615	0.017	0.536	0.024	TRUE	8	0.262	0.017	0.132	0.026
16	10	0.638	0.059	0.569	0.071	TRUE	10	0.327	0.043	0.196	0.053
32	10	0.632	0.029	0.568	0.037	TRUE	10	0.351	0.034	0.234	0.043
64	8	0.630	0.052	0.568	0.056	TRUE	8	0.417	0.018	0.319	0.020
128	9	0.609	0.021	0.548	0.025	TRUE	9	0.445	0.008	0.355	0.010
256	10	0.582	0.029	0.516	0.033	TRUE	10	0.473	0.019	0.389	0.020
512	10	0.573	0.018	0.506	0.021	TRUE	10	0.492	0.009	0.413	0.010
1024	10	0.543	0.009	0.472	0.010	TRUE	10	0.506	0.012	0.430	0.013
kNN											
1	9	0.832	0.024	0.787	0.023	1	9	0.278	0.047	0.120	0.053
2	10	0.852	0.019	0.806	0.020	1	9	0.323	0.038	0.166	0.031
4	9	0.850	0.007	0.808	0.014	1	8	0.297	0.019	0.153	0.032
8	9	0.848	0.014	0.808	0.014	1	8	0.336	0.030	0.203	0.033
16	10	0.861	0.016	0.828	0.022	1	8	0.377	0.024	0.256	0.029
32	10	0.865	0.009	0.837	0.011	1	9	0.415	0.039	0.306	0.047
64	8	0.866	0.013	0.840	0.014	1	7	0.474	0.020	0.378	0.022
128	9	0.867	0.005	0.843	0.005	1	9	0.507	0.006	0.420	0.007
256	10	0.860	0.007	0.834	0.009	1	10	0.551	0.005	0.472	0.005
512	10	0.860	0.006	0.835	0.007	1	10	0.594	0.006	0.522	0.007
1024	10	0.835	0.008	0.806	0.010	1	10	0.635	0.004	0.571	0.004
RF											
1	9	0.914	0.010	0.890	0.009	9	4	0.326	0.035	0.148	0.050
2	10	0.921	0.009	0.896	0.010	9	3	0.374	0.050	0.220	0.049
4	9	0.924	0.004	0.902	0.006	6	4	0.359	0.048	0.216	0.054
8	9	0.915	0.007	0.893	0.008	6	3	0.379	0.043	0.242	0.048
16	10	0.918	0.011	0.899	0.013	5	4	0.425	0.022	0.300	0.028
32	10	0.918	0.002	0.901	0.004	6	4	0.484	0.034	0.381	0.041
64	8	0.917	0.007	0.901	0.009	5	2	0.553	0.017	0.466	0.020
128	9	0.915	0.006	0.900	0.008	5	3	0.606	0.013	0.531	0.016
256	10	0.910	0.004	0.893	0.005	6	3	0.658	0.012	0.594	0.014
512	10	0.907	0.003	0.891	0.003	8	4	0.709	0.009	0.654	0.011
1024	10	0.892	0.005	0.872	0.006	9	3	0.762	0.007	0.718	0.008
SVM											
1	9	0.877	0.019	0.843	0.020	64	6	0.305	0.055	0.144	0.059
2	6	0.890	0.011	0.857	0.012	128	4	0.327	0.055	0.175	0.063
4	7	0.891	0.006	0.862	0.007	128	5	0.315	0.040	0.184	0.044
8	7	0.876	0.015	0.846	0.017	128	4	0.350	0.054	0.217	0.057
16	9	0.888	0.011	0.862	0.016	64	5	0.398	0.031	0.280	0.037
32	7	0.882	0.008	0.857	0.009	128	5	0.453	0.041	0.351	0.046
64	8	0.884	0.012	0.860	0.014	64	5	0.497	0.014	0.404	0.015
128	7	0.877	0.008	0.855	0.010	64	4	0.536	0.011	0.453	0.014
256	8	0.870	0.006	0.846	0.008	128	6	0.580	0.014	0.505	0.017
512	10	0.861	0.004	0.836	0.004	128	5	0.630	0.006	0.563	0.007
1024	10	0.832	0.007	0.802	0.009	64	7	0.671	0.008	0.613	0.009
ANN											
1	9	0.816	0.036	0.765	0.039	30	8	0.267	0.048	0.121	0.056
2	10	0.838	0.019	0.788	0.022	30	7	0.324	0.040	0.180	0.027
4	9	0.840	0.024	0.795	0.033	30	5	0.289	0.077	0.172	0.064
8	9	0.828	0.015	0.783	0.015	30	8	0.332	0.049	0.208	0.047
16	10	0.837	0.021	0.798	0.028	30	6	0.381	0.033	0.267	0.035
32	10	0.826	0.015	0.790	0.018	30	8	0.420	0.042	0.314	0.044
64	8	0.824	0.026	0.789	0.030	30	8	0.475	0.012	0.379	0.014
128	9	0.808	0.012	0.772	0.015	30	8	0.505	0.013	0.418	0.017
256	10	0.787	0.012	0.747	0.015	30	8	0.558	0.010	0.479	0.011
512	10	0.765	0.012	0.723	0.013	30	8	0.591	0.012	0.517	0.013
1024	10	0.735	0.008	0.687	0.009	30	6	0.627	0.005	0.561	0.005

imply that for real world mapping applications, spatially scattered field observations are most likely to generate stable MLA predictions that accurately reflect the spatial distributions of lithologies. Given appropriately distributed T_a, the proposed methods are applicable to problems such as generating rapid first-pass predictions in complex geological terranes, reducing the subjectivity with which geological interpretations are formulated, and validating pre-existing interpreted geological maps.

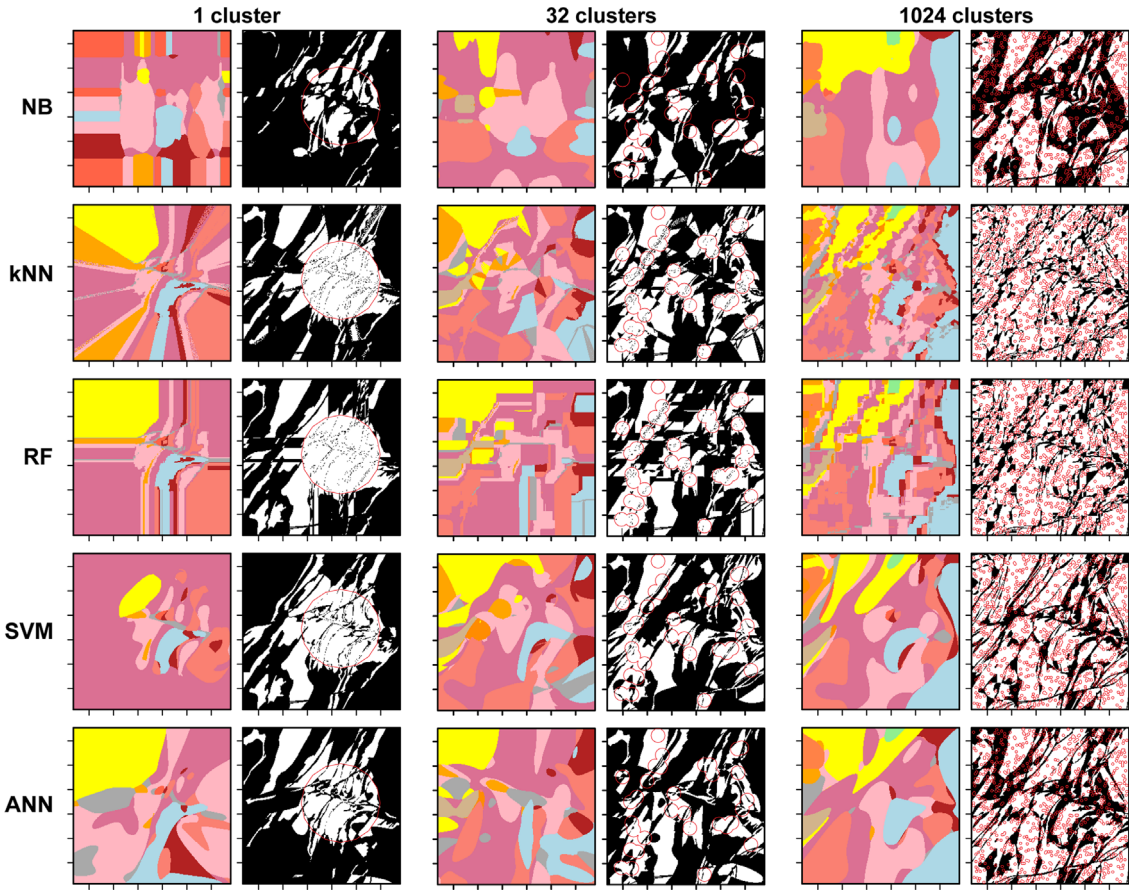

Fig. 6. Visualization of the spatial distribution of MLA lithology class predictions (refer to Fig. 1 for reference map and key to class labels), and misclassified samples (black pixels) for 1, 32 and 1024 T_a clusters (red circles) using X and Y spatial coordinates (XY Only) as inputs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

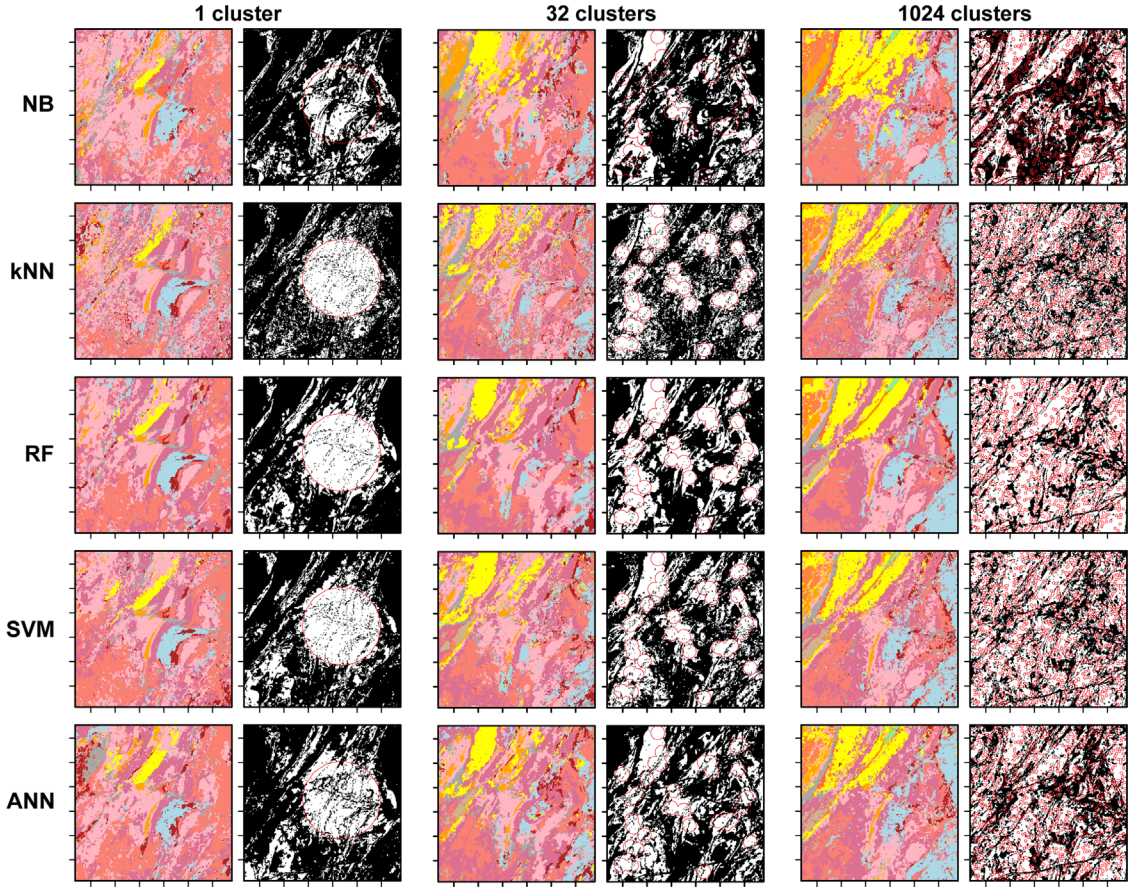

Fig. 7. Visualization of the spatial distribution of MLA lithology class predictions (refer to Fig. 1 for reference map and key to class labels), and misclassified samples (black pixels) for 1, 32 and 1024 T_a clusters (red circles) using geophysical data (No XY) as inputs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

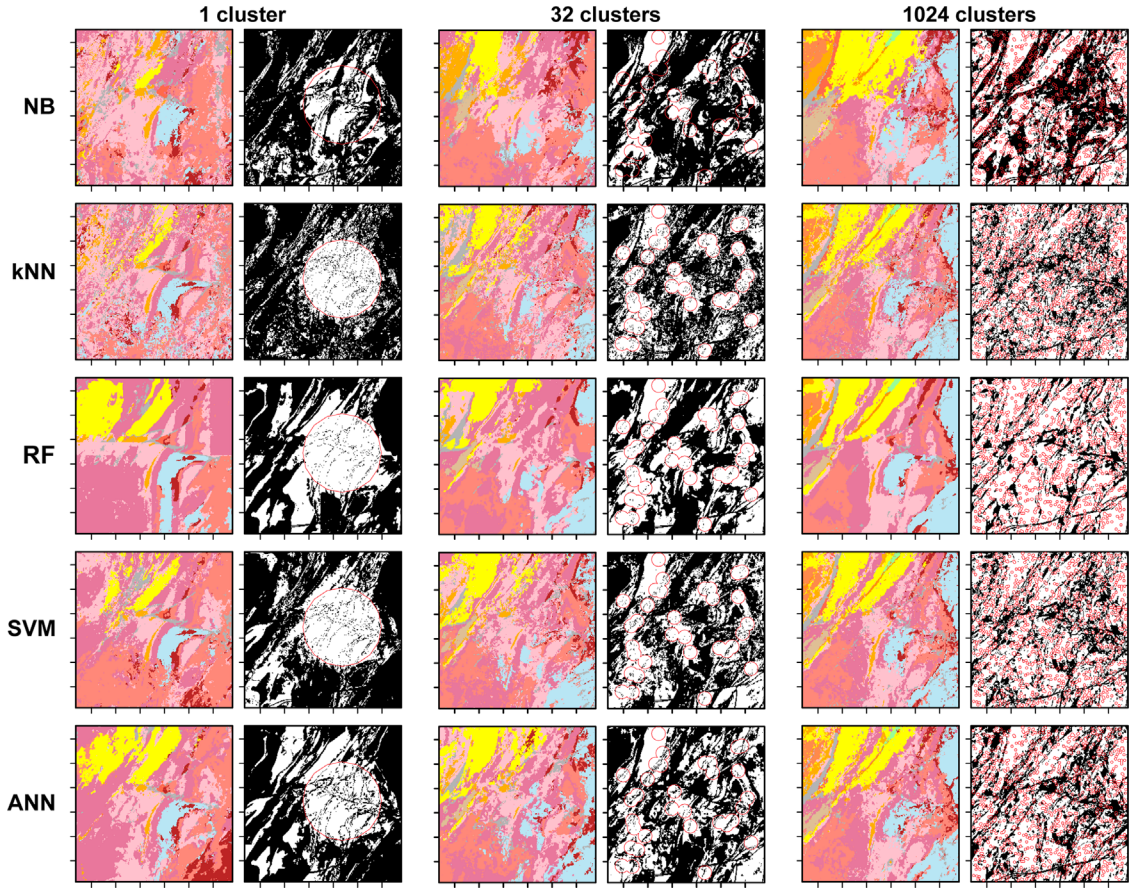

Fig. 8. Visualization of the spatial distribution of MLA lithology class predictions (refer to Fig. 1 for reference map and key to class labels), and misclassified samples (black pixels) for 1, 32 and 1024 T_a clusters (red circles) using spatial coordinates and geophysical data (all data) as inputs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

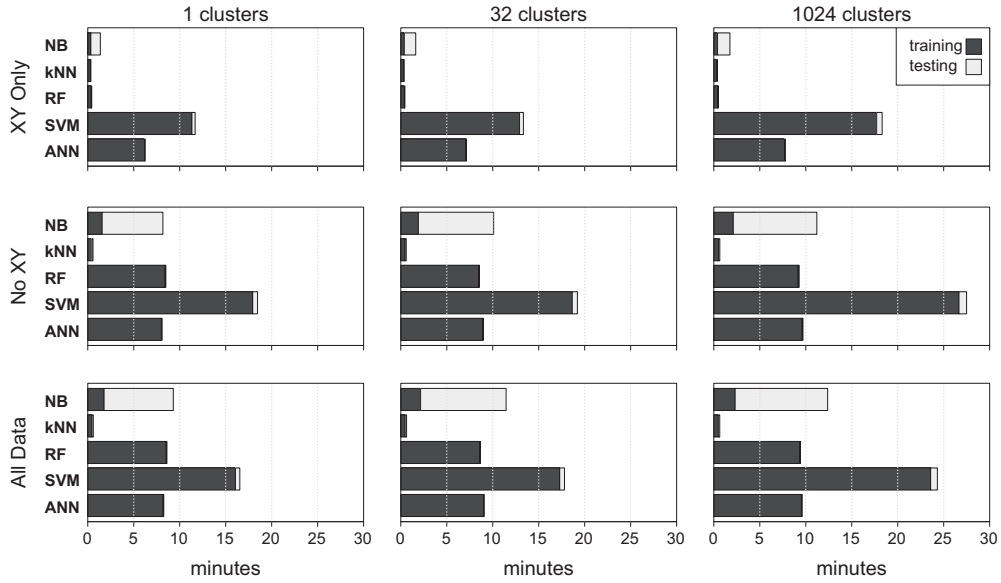

Fig. 9. Comparison of MLA training (using 10-fold cross-validation) and prediction processing times (minutes). Bars represent the mean time taken to train a MLA classification model and generate predictions for a maximum of 10 sets of T_a and T_b. All processes were executed using a DELL Desktop PC 64-bit Intel Dual Core CPU @ 3.00 GHz and 8 GB of RAM. Cross-validation folds and T_b predictions (divided into groups of 10,000) were run independently using Message Passing Interface parallel processing on 2 cores.

5.3. *Using spatially constrained data*

Spatial dependencies (spatial autocorrelation), i.e. closer observations are more related than those farther apart, are commonly encountered within spatial data (Anselin, 1995; Getis, 2010; Lloyd, 2011). When incorporating spatial coordinates as input data for training and prediction, MLAs are provided with explicit information linking location and lithology classes. Spatial values, coupled

with randomly distributed T_a samples over the entire study area, resulted in a best-case scenario in terms of MLA test accuracy. However, geologically plausible predictions were only achieved using input variables containing information that reflected the physical properties of the lithologies under investigation. This observation suggests that standard measures of classifier performance, such as accuracy and kappa, do not necessarily provide an indication of the validity of MLA classifications in the spatial domain. Therefore, the use of spatial coordinates should be approached with caution. If, for instance, training data were collected only from field observations, these data are likely to be spatially clustered due outcrop exposure, accessibility, and time constraints. Using only spatial coordinates with highly spatially clustered T_a samples generates classifiers that incorrectly learn a finite range of coordinate values for a particular class and were unable to generalize well to spatially disjoint regions (Gahegan, 2000).

An alternative to the use of spatial coordinates as a means of providing MLAs with spatial context would be to incorporate some measure of local spatial relationships during pre-processing using neighborhood models (Gahegan, 2000). In its simplest form, neighborhood models generate new variables by deriving statistical measures of the values of proximal observations such as mean or standard deviation. In this way neighborhood models represent local spatial similarity or dissimilarity (Lloyd, 2011). An alternative to pre-processing inputs to include information on the spatial context of observations is to code some notion of spatial dependency into the MLA itself. Li et al. (2012) developed and tested a SVM algorithm that directly utilized spatial information to improve classifier accuracy. Their SVM variant takes analyzed hold-out test sample results obtained in geographical space, and using a local neighborhood identifies the support vectors that are likely to be misclassified. The SVM hyperplane boundary is then adjusted to include these support vectors. This process proceeds iteratively until the change in prediction accuracy reaches some minimum threshold.

6. Conclusions

We compared five machine learning algorithms (MLAs) in terms their performance with respect to a supervised lithology classification problem in a complex metamorphosed geological terrane. These MLAs, Naive Bayes, *k*-Nearest Neighbors, Random Forests, Support Vector Machines and Artificial Neural Networks, represent the five general machine learning strategies for automated data-driven inference. MLA comparisons included their sensitivity to variations in spatial distribution of training data, and response to the inclusion of explicit spatial information.

We conclude that Random Forests is a good first choice MLA for multiclass inference using widely available high-dimensional multisource remotely sensed geophysical variables. Random Forests classification models are, in this case, easy to train, stable across a range of model parameter values, computationally efficient, and when faced with spatially dispersed training data, substantially more accurate than other MLAs. These traits, coupled with the insensitively of Random Forests to noise and over fitting, indicate that it is well suited to remote sensing lithological classification applications.

As the spatial distribution of training data becomes more dispersed across the region under investigation, MLA predictive accuracy (and kappa) increases. In addition, the sensitivity of MLA categorical predictions to different training datasets decreases. The inclusion of explicit spatial information (i.e. spatial coordinates) proved to generate highly accurate MLA predictions when training data was dispersed across the study area. Despite resulting in lower test accuracy (and kappa), the use of geophysical data provided MLAs with information that characterized geological structural trends. Therefore, combining spatial and geophysical data is beneficial, especially in situations where training data is moderately spatially clustered. Our results illustrate the use of machine learning techniques to address data-driven spatial inference problems and will generalize readily to other geoscience applications.

Acknowledgments

We thank Malcolm Sambridge and Ron Berry for their constructive comments and discussion which significantly improved the draft manuscript. Airborne geophysics data were sourced from Geoscience Australia and Landsat ETM+ data from the United States Geological Survey. This research was conducted at the Australian Research Council Centre of Excellence in Ore Deposits (CODES) under Project no. P3A3A. M. Cracknell was supported through a University of Tasmania Elite Research Ph.D. Scholarship. We thank the editor and Milos Kovacevic for their constructive comments that have strengthened the scientific content and clarity of presentation.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.cageo.2013.10.008.

References

Anselin, L., 1995. Local indicators of spatial association – LISA. Geogr. Anal. 27, 93–115.
Breiman, L., 1996. Bagging predictors. Machine Learn. 24, 123–140.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworths & Brooks/Cole Advanced Books & Software, Pacific Grove, USA p. 358.
Buckley, P.M., Moriarty, T., Needham, J. (compilers), 2002. Broken Hill Geoscience Database, 2nd ed. Geological Survey of New SouthWales , Sydney.
Carneiro, C.C., Fraser, S.J., Croacutesta, A.P., Silva, A.M., Barros, C.E.M., 2012. Semiautomated geologic mapping using self-organizing maps and airborne geophysics in the Brazilian Amazon. Geophysics 77, K17–K24.
Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46.
Congalton, R.G., Green, K., 1998. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, first edn. Lewis Publications, Boca Raton p. 137.
Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE Trans. Inf. Theory 13, 21–27.
Cracknell, M.J., Reading, A.M., 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, WB113–WB126.
Fix, E., Hodges, J.L., 1951. Discriminatory analysis. Nonparametric discrimination; Consistency properties. U.S. Air Force, School of Aviation Medicine, Randolph Field, Texas.
Foody, G.M., Mathur, A., 2004. A relative evaluation of multiclass image classification by support vector machines. IEEE Trans. Geosci. Remote Sens. 42, 1335–1343.
Gahegan, M., 2000. On the application of inductive machine learning tools to geographical analysis. Geogr. Anal. 32, 113–139.
Gelfort, R., 2006. On Classification of Logging Data, Department of Energy and Economics. Clausthal University of Technology, Germany, p. 131.
Getis, A., 2010. Spatial autocorrelation. In: Fisher, M.M., Getis, A. (Eds.), Handbook of Applied Spatial Analysis: Software, Tools, Methods and Applications. Springer-Verlag, Berlin, pp. 255–278.
Guyon, I., 2008. Practical feature selection: from correlation to causality. In: Fogelman-Soulié, F., Perrotta, D., Piskorski, J., Steinberger, R. (Eds.), Mining Massive Data Sets for Security – Advances in Data Mining, Search, Social Networks and Text Mining, and their Applications to Security. IOS Press, Amsterdam, pp. 27–43.
Guyon, I., 2009. A practical guide to model selection. In: Marie, J. (Ed.), Proceedings of the Machine Learning Summer School. Canberra, Australia, January 26 - February 6, Springer Text in Statistics, Springer p.37.

Ham, J., Yangchi, C., Crawford, M.M., Ghosh, J., 2005. Investigation of the random forest framework for classification of hyperspectral data. IEEE Trans. Geosci. Remote Sens. 43, 492–501.
Hastie, T., Tibshirani, R., Friedman, J.H., 2009. The elements of statistical learning: data mining, Inference and Prediction, 2nd edn. Springer, New York, USA p. 533.
Henery, R.J., 1994. Classification. In: Michie, D., Spiegelhalter, D.J., Taylor, C.C. (Eds.), Machine Learning, Neural and Statistical Classification. Ellis Horwood, New York, pp. 6–16.
Hsu, C.-W., Lin, C.-J., 2002. A comparison of methods for multiclass support vector machines. IEEE Trans. Neural Netw. 13, 415–425.
Hsu, C.-W., Chang, C.-C., Lin, C.-J., 2010. A Practical Guide to Support Vector ClassificationDepartment of Computer Science, National Taiwan University, Taipei, Taiwan16.
Huang, C., Davis, L.S., Townshend, J.R.G., 2002. An assessment of support vector machines for land cover classification. Int. J. Remote Sens. 23, 725–749.
Kanevski, M., Pozdnoukhov, A., Timonin, V., 2009. Machine learning for spatial environmental data: theory, Applications and Software. CRC Press, Boca Raton, USA (368 pp.).
Karatzoglou, A., Meyer, D., Hornik, K., 2006. Support vector machines in R. J. Stat. Softw. 15, 28.
Kotsiantis, S.B., 2007. Supervised machine learning: a review of classification techniques. Informatica 31, 249–268.
Kovacevic, M., Bajat, B., Trivic, B., Pavlovic, R., 2009. Geological units classification of multispectral images by using Support Vector Machines. In: Proceedings of the International Conference on Intelligent Networking and Collaborative Systems, IEEE, pp. 267–272.
Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., 2012. caret: Classifcation and Regression Training, R Package Version 5.15-023.
Kuncheva, L., 2004. Combining Pattern Classifiers: Methods and Algorithms. John Wiley & Sons p. 376.
Landgrebe, T.C.W., Paclik, P., 2010. The ROC skeleton for multiclass ROC estimation. Pattern Recogn. Lett. 31, 949–958.
Leverington, D.W., 2010. Discrimination of sedimentary lithologies using Hyperion and Landsat Thematic Mapper data: a case study at Melville Island, Canadian High Arctic. Int. J. Remote Sens. 31, 233–260.
Leverington, D.W., Moon, W.M., 2012. Landsat-TM-based discrimination of lithological units associated with the Purtuniq Ophiolite, Quebec, Canada. Remote Sens. 4, 1208–1231.
Li, C.-H., Kuo, B.-C., Lin, C.-T., Huang, C.-S., 2012. A spatial-contextual Support Vector Machine for remotely sensed image classification. IEEE Trans. Geosci. Remote Sens. 50, 784–799.
Lloyd, C.D., 2011. Local Models for Spatial Analysis, 2nd edn. CRC Press, Taylor & Francis Group, Boco Raton, USA p. 336.
Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. Int. J. Remote Sens. 28, 823–870.
Marsland, S., 2009. Machine Learning: An Algorithmic Perspective. Chapman & Hall/CRC (406 pp.).
Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. IEEE Trans. Geosci. Remote Sens. 42, 1778–1790.
Molina, R., P´erez de la Blanca, N., Taylor, C.C., 1994. Modern statistical techniques. In: Michie, D., Spiegelhalter, D.J., Taylor, C.C. (Eds.), Machine Learning. Neural and Statistical Classification. Ellis Horwood, New York, pp. 29–49.
Oommen, T., Misra, D., Twarakavi, N.K.C., Prakash, A., Sahoo, B., Bandopadhyay, S., 2008. An objective analysis of support vector machine based classification for remote sensing. Math. Geosci. 40, 409–424.
Page, R.W., Conor, C.H.H., Stevens, B.P.J., Gibson, G.M., Preiss, W.V., Southgate, P.N., 2005a. Correlation of Olary and Broken Hill Domains, Curnamona Province: possible relationship to Mount Isa and other North Australian Pb–Zn–Ag-bearing successions. Econ. Geol. 100, 663–676.
Page, R.W., Stevens, B.P.J., Gibson, G.M., 2005b. Geochronology of the sequence hosting the Broken Hill Pb–Zn–Ag orebody, Australia. Econ. Geol. 100, 633–661.
Pal, M., 2005. Random forest classifier for remote sensing classification. Int. J. Remote Sens. 26, 217–222.
Provost, F., Fawcett, T., 1997. Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions, Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD 97). American Association for Artificial Intelligence, Huntington Beach, CA, pp. 43–48.
Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge, UK p. 403.
Rojas, R., 1996. Neural Netwoks: A Systematic Introduction. Springer-Verlag, Berlin p. 502.
Song, X., Duan, Z., Jiang, X., 2012. Comparison of artificial neural networks and support vector machine classifiers for land cover classification in Northern China using a SPOT-5 HRG image. Int. J. Remote Sens. 33, 3301–3320.
Stevens, B.P.J., 1986. Post-depositional history of the Willyama Supergroup in the Broken Hill Block, NSW. Aust. J. Earth Sci. 33, 73–98.
Vapnik, V.N., 1998. Statistical Learning Theory. John Wiley & Sons, Inc., New York, USA p. 736.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, 4th edn. Springer, New York, USA (495 pp.).
Waske, B., Braun, M., 2009. Classifier ensembles for land cover mapping using multitemporal SAR imagery. ISPRS J. Photogramm. Remote Sens. 64, 450–457.
Waske, B., Benediktsson, J.A., Árnason, K., Sveinsson, J.R., 2009. Mapping of hyperspectral AVIRIS data using machine-learning algorithms. Can. J. Remote Sens. 35, 106–116.
Webster, A.E., 2004. The Structural Evolution of the Broken Hill Pb–Zn–Ag deposit, New South Wales, Australia, ARC Centre for Excellence in Ore Deposit ResearchUniversity of Tasmania, Hobart p. 430.
Williams, D., 2009. Landsat 7 Science Data User's Handbook. National Aeronautics and Space Administration, Greenbelt, Maryland p. 186.
Willis, I.L., Brown, R.E., Stroud, W.J., Stevens, B.P.J., 1983. The early proterozoic Willyama supergroup: stratigraphic subdivision and interpretation of high to low-grade metamorphic rocks in the Broken Hill Block, New South Wales. J. Geol. Soc. Aust. 30, 195–224.
Witten, I.H., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques, 2nd edn. Elsevier/Morgan Kaufman, San Fransisco, USA p. 525.
Yu, L., Porwal, A., Holden, E.J., Dentith, M.C., 2012. Towards automatic lithological classification from remote sensing data using support vector machines. Comput. Geosci. 45, 229–239.