Ore Geology Reviews xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Ore Geology Reviews

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

Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier

Andrew S. Merdith [a,b,*], Thomas C.W. Landgrebe [a], R. Dietmar Müller [a]

[a] EarthByte Group, School of Geosciences, The University of Sydney, Madsen Building F09, Australia
[b] National ICT Australia (NICTA), Australian Technology Park, Australia

ARTICLE INFO

Article history:
Received 19 November 2014
Received in revised form 17 March 2015
Accepted 18 March 2015
Available online xxxx

Keywords:
Reject option classifier
Iron ore
Mineral exploration
Predictive targeting
Mixture of Gaussians
Geophysical exploration

ABSTRACT

There has recently been a rapid growth in the amount and quality of digital geological and geophysical data for the majority of the Australian continent. Coupled with an increase in computational power and the rising importance of computational methods, there are new possibilities for a large scale, low expenditure digital exploration of mineral deposits. Here we use a multivariate analysis of geophysical datasets to develop a methodology that utilises machine learning algorithms to build and train two-class classifiers for provincial-scale, greenfield mineral exploration. We use iron ore in Western Australia as a case study, and our selected classifier, a mixture of a Gaussian classifier with reject option, successfully identifies 88% of iron ore locations, and 92% of non-iron ore locations. Parameter optimisation allows the user to choose the suite of variables or parameters, such as classifier and degree of dimensionality reduction, that provide the best classification result. We use randomised hold-out to ensure the generalisation of our classifier, and test it against known ground-truth information to demonstrate its ability to detect iron ore and non-iron ore locations. Our classification strategy is based on the heterogeneous nature of the data, where a well-defined target "iron-ore" class is to be separated from a poorly defined non-target class. We apply a classifier with reject option to known data to create a discriminant function that best separates sampled data, while simultaneously "protecting" against new unseen data by "closing" the domain in feature space occupied by the target class. This shows a substantial 4% improvement in classification performance. Our predictive confidence maps successfully identify known areas of iron ore deposits through the Yilgarn Craton, an area that is not heavily sampled in training, as well as suggesting areas for further exploration throughout the Yilgarn Craton. These areas tend to be more concentrated in the north and west of the Yilgarn Craton, such as around the Twin Peaks mine (~27°S, 116°E) and a series of lineaments running east–west at ~25°S. Within the Pilbara Craton, potential areas for further expansion occur throughout the Marble Bar vicinity between the existing Spinifex Ridge and Abydos mines (21°S, 119–121°E), as well as small, isolated areas north of the Hamersley Group at ~21.5°S, ~118°E. We also test the usefulness of radiometric data for province-scale iron ore exploration, while our selected classifier makes no use of the radiometric data, we demonstrate that there is no performance penalty from including redundant data and features, suggesting that where possible all potentially pertinent data should be included within a data-driven analysis. This methodology lends itself to large scale, reconnaissance mineral explorations, and, through varying the datasets used and the commodity being targeted, predictive confidence maps for a wide range of minerals can be produced.

1. Introduction

The viability of an ore deposit is governed not only by its geological features, but also by economic factors such as global demand and market value. Consequently, economic geology links together frontier geological science with an economic framework; what defines an economically viable ore deposit can alter over time as a consequence of changes in the cost of exploration, extraction and production as well as demand in the global market (Pohl, 2011). As ore deposits, which are both easy to find and extract, are being depleted, there is a rising cost associated with finding new ore deposits using existing exploration techniques. The increase in computational power and in the availability of high-resolution data allows for new methodologies to be developed for the purposes of data-driven mineral exploration, in an effort to reduce exploration costs in finding large ore bodies.

The challenge in developing effective targeting aids that generalise to new geographic locations is in developing methods and models that exploit available data without over-fitting, and lend themselves to continuous improvement as more/higher-quality data become

* Corresponding author at: EarthByte, School of Geosciences, The University of Sydney, Madsen Building F09, Australia.
E-mail address: Andrew.merdith@sydney.edu.au (A.S. Merdith).

http://dx.doi.org/10.1016/j.oregeorev.2015.03.014
0169-1368/

available. Of particular importance is understanding the impact that factors such as data redundancy, correlation between datasets, sample-sizes and data-dimensionality have on the effectiveness of the models and outputs generated by their amalgamation. In this paper we have used a multivariate analysis of geophysical databases to develop a methodology that utilises machine learning algorithms (MLAs) to build and train a classifier to predict the presence and absence of iron ore deposits throughout Western Australia. Importantly, our classifier is designed to be applied over large areas of land (on the order of 1000 km), as such, while generalising well for the exploration of iron ore over multiple cratons and orogenies, it is not specifically adapted for local geology and regional–local scale exploration. As such, we utilise a training and evaluation methodology, which takes into account the various aforementioned factors and attempts to make use of available data in the most effective way.

As conceptual targeting of potential sites at a province/district to regional scale is one of the largest challenges facing geoscience (Hronsky and Groves, 2008) this methodology will assist with first order, large-scale exploration. Mineral exploration consists of a number of successive but interlinked stages, starting with planning and large-scale reconnaissance exploration, before moving towards smaller scale appraisals and explorative drilling, and then finally assessment drilling and mine development (Moon and Whateley, 2006; Pohl, 2011). Generally, as the stages progress the associated economic risk decreases, but the expenditure required increases (i.e. more money is spent on stages that have a higher confidence of success) (Moon and Whateley, 2006). The methodology outlined in this paper is designed to fit in the early reconnaissance stage of mineral exploration, assisting and facilitating in the identification of potential target locations for a commodity. The nature of this methodology fits well in the exploration method, as it has a low expenditure. However, unlike existing reconnaissance exploration, we believe that there is less associated risk with this methodology. While formal mineral exploration already consists of the analysis of geophysical data, our approach differs in that we minimise human bias and use computational methods that allow for the combination and analysis of large amounts of high-dimensional data to create a prospectivity map of a target commodity. We argue that studies in the past such as Groves et al. (2000) and Nykänen et al. (2008) in which targeting "layers" are independently combined together are suboptimal. They do not take into account that many geological/geophysical statistics are not independent, resulting in under-exploited data separability (Brown et al., 2000; Porwal et al., 2003; Singer and Kouda, 1999). In this paper a multivariate approach is taken where we explicitly attempt to deal with these issues, thus combining various data sources into a single model that involves feature extraction and classification to both cope with the dependence between variables, and make trade-offs between the number of dimensions, available data and classifier complexity.

2. Background geology

Though iron ore is one of the most economically important natural commodities, there is still some uncertainty about its genesis. This is due, in part, to an absence of modern analogues with respect to both the process of formation (Bekker et al., 2010) and also the scale at which the deposits form (Morris, 1985). Additionally, ambiguities about Archean and Proterozoic geology, climatic conditions and seawater chemistry (e.g. Canfield, 2005; Lyons et al., 2009; Planavsky et al., 2011) have also caused numerous mechanisms and ideas being proposed over time for the source and transport paths of iron, the timing of deposition of iron formations and the subsequent enrichment of iron formation to iron ore. For instance, while it is generally accepted that all major iron ore deposits occurred in an oceanic setting, the source of the iron within the oceans was thought to originate from continental erosion until, an alternative hydrothermal source was proposed (Isley, 1995).

Banded iron formations (BIF) are defined after James and Trendall (1982) as a rock with thin laminations of chert alternating with iron minerals, and can be broadly classified into two categories based on their depositional environment, Algoma-type and Superior-type (Gross, 1980). Algoma-type deposits are found through Archean and Proterozoic formations, and are associated with volcanic centres and exhalative submarine processes, and typically contain some greywacke or volcanic units (Gross, 1980). They are typically found within Archean greenstone belts (Bekker et al., 2010; Goodwin, 1973; Isley and Abbott, 1999) and are usually smaller, both in terms of tonnage of ore (largest deposits around 10^7 Mt) and in spatial extent than Superior-type deposits. Comparably, Superior-type deposits are more common in Proterozoic aged formations and are associated with a near-shore continental-shelf depositional setting, usually found with carbonates, quartzite, black shales, and small amounts of volcanogenic rocks (Gross, 1980). The Superior-type tend to be larger, up to 10^{14} Mt, and also occupy a larger spatial extent (Bekker et al., 2010; Huston and Logan, 2004; Isley, 1995). Both types of iron formation are associated with oxide, silicate and carbonate facies (Gross, 1980; James, 1954), while Algoma-types may be associated with polymetallic sulphide facies if they occurred in close proximity to a volcanic centre.

2.1. Geological setting

Australia is host to both Algoma- and Superior-type deposits, though it is predominantly known for its massive Superior-type deposits occurring throughout the Hamersley Basin in the Pilbara Craton (Fig. 1). The Pilbara Craton consists of a Paleo-Neoarchean core overlain with a strong angular unconformity by Neoarchean–Paleoproterozoic volcano-sedimentary sequences (Blake and Barley, 1992; van Kranendonk et al., 2002). The core consists of a granite–greenstone terrane that outcrops towards the north, collectively called the North Pilbara terrain (Fig. 2). The North Pilbara terrain has been subdivided into three distinct granite–greenstone terranes, the East Pilbara granite–greenstone terrane (3.72–2.85 Ga), the West Pilbara granite–greenstone terrane (3.27–2.92 Ga) and the Kurunna terrane (3.3–3.2 Ga) towards the southeast of the craton, and two intracratonic sedimentary basins, the Mallina Basin (3.01–2.94 Ga) and the Mosquito Creek basin (~3.3–2.9 Ga) (van Kranendonk et al., 2002). Smaller Algoma-type deposits occur in the Eastern Pilbara granite–greenstone terrane amongst the Gorge Creek Group and Cleaverville Formation (Huston and Logan, 2004). The volcano-sedimentary sequences, collectively referred to as the Hamersley province, overlay the southern part of the craton and are of principal interest to this study as they contain some of the largest and richest iron ore deposits in the world. The stratigraphy of the Hamersley province is divided into five key groups. The lower three, the Fortescue Group (2770–2630 Ma), the iron rich Hamersley Group (2630–2470 Ma) and the Turee Creek Group (2470–ca.2350 Ma) all conformably overlay one another and comprise the Mt. Bruce Supergroup (Taylor et al., 2001). The upper two groups, the Lower and Upper Wyloo Groups (2209–2150 Ma and 2000–1800 Ma respectively), are separated from the Mt. Bruce Supergroup by a first order regional unconformity (Taylor et al., 2001). The Fortescue Group is characterised by mafic–clastic sedimentation, while the lower and middle units of the Hamersley Group are indicative of a deep-water environment consisting of volcanoclastic sedimentary and some carbonate sedimentary units and the Turee Creek Group consists of coarser, clastic sedimentary rocks overlaying iron formation, suggesting a transition from a deep to shallow sea environment (Blake and Barley, 1992; Simonson et al., 1993). Iron ore is found extensively throughout the Hamersley Group. Deformation is more pronounced in the south of the Hamersley province where the younger units outcrop, with the older basal units in the north of the province only being gently folded (Taylor et al., 2001).

The Yilgarn Craton is a large, Archean aged section of crust within Western Australia, to the south of the Pilbara Craton. Similar to the Pilbara Craton, it is comprised predominantly of Mesoarchean low-grade metamorphosed granite–greenstone belts, though it also contains

Please cite this article as: Merdith, A.S., et al., Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier, Ore Geol. Rev. (2015), http://dx.doi.org/10.1016/j.oregeorev.2015.03.014

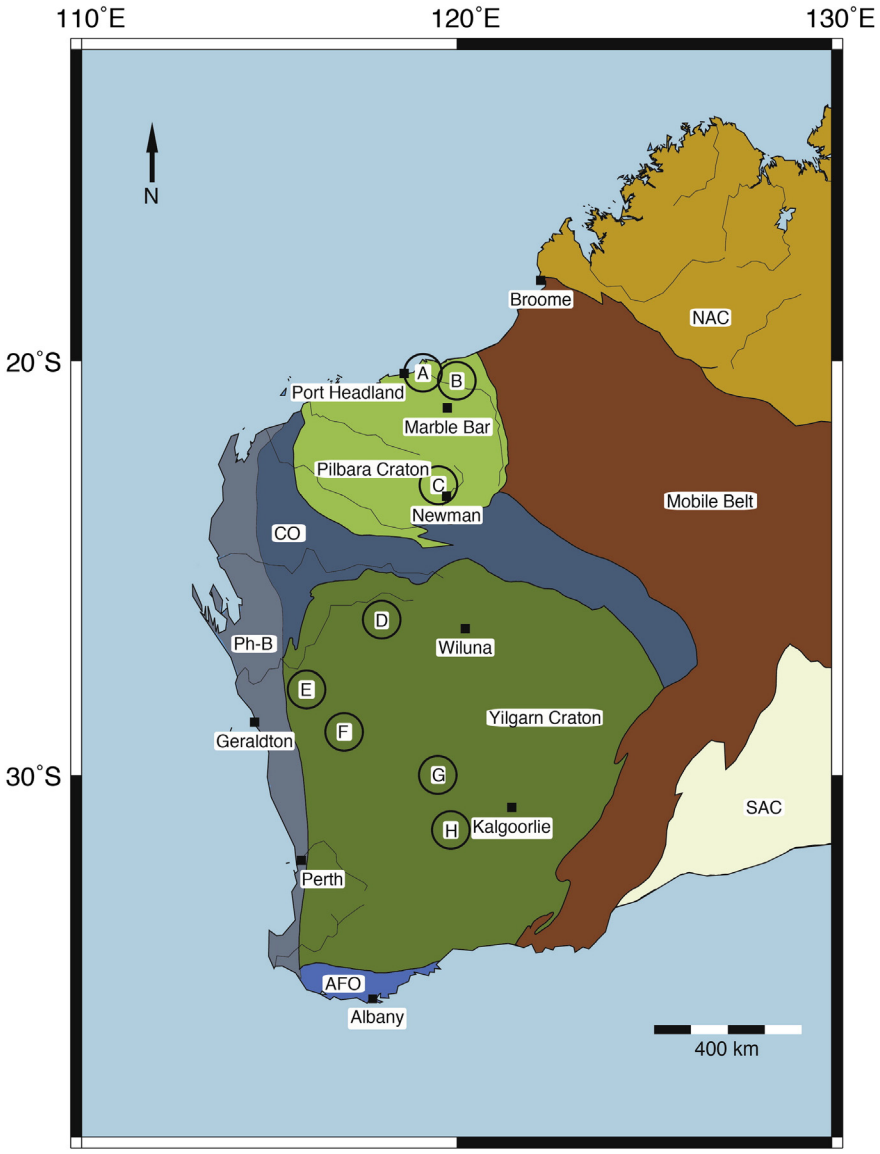

Fig. 1. Map of Western Australia showing geological provinces and major towns and cities. (AFO) Albany-Fraser Orogen; (CO) Capricorn Orogen; (NAC) North Australian Craton; (Ph-B) Phanerozoic Basin; (SAC) South Australian Craton. Black circles represent some iron deposits (A) Pardoo-Mt Goldsworthy; (B) Cattle Gorge; (C), Mt. Whaleback; (D) Madoonga; (E) Tallering Peak; (F) Karara; (G) Windarling; and (H) Koolyanobbing. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

an Eo-Paleoarchean gneissetic component in the north west (Griffin et al., 2004; Myers, 1993). A number of studies aiming to subdivide the craton into discrete geological domains have been attempted using geological, geophysical and geochemical data (e.g. Gee et al., 1981; Griffin et al., 2004; Myers, 1993; Whitaker, 2001), though it is considerably more difficult than in the case of the Pilbara Craton due to the thick regolith that covers large areas of the craton (Anand and Paine, 2002; Dauth, 1997). Key terranes outlined by Myers (1993) include the West Yilgarn Superterrane (3.8–3.7 Ga core in the northwest, intruded by 3.0 Ga granitoids, and 3.0 Ga core in the west and southwest, intruded by 2.9 Ga granitoids), the West Central Yilgarn Superterrane (2.8–2.7 Ga greenstones), the East Central Yilgarn Superterrane (2.7 Ga greenstones) and the East Yilgarn Superterrane (2.8 Ga greenstones). In the Yilgarn Craton, iron ore is restricted to the greenstone belts, such as the Woongan Hill Terrane, Noganyer Formation and Koolyanobbing Greenstone Belt (Huston and Logan, 2004; Isley and Abbott, 1999) and also throughout the Yilgarn Craton within the Woongan Hill Terrane and the Noganyer Formation (both greenstone belts) (Huston and Logan, 2004; Isley and Abbott, 1999) (Fig. 3). Recent advances in iron ore exploration have acknowledged the importance of these smaller Algoma-type deposits in the Yilgarn Craton to the Australian iron industry, and a large body of research is being built up about their deposition, geological properties, structural controls on ore genesis and exploration strategies (e.g. Angerer and Hagemann, 2010; Duuring and Hagemann, 2013; Lascelles, 2006).

2.2. Models of formation

While classification of the depositional environment for BIF is widely accepted, the process(-es) by which enrichment occurs (from ~15% to ~55% iron) is more uncertain. Broadly, methods of enrichment are divided into three main models; syngenetic, supergene and hypogene–supergene models. The syngenetic models advocate that the conversion from BIF to ore is completed during diagenesis. This can occur during an extensional event through the creation of a dilational space within the chert–iron bands, allowing for the incremental removal of silica and concentration of iron where extensional disturbances during diagenesis allowed for iron enrichment to occur (Findlay, 1994), or through breaks in the iron laminae during diagenesis allowing for the escape of gelatinous, amorphous

Please cite this article as: Merdith, A.S., et al., Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier, Ore Geol. Rev. (2015), http://dx.doi.org/10.1016/j.oregeorev.2015.03.014

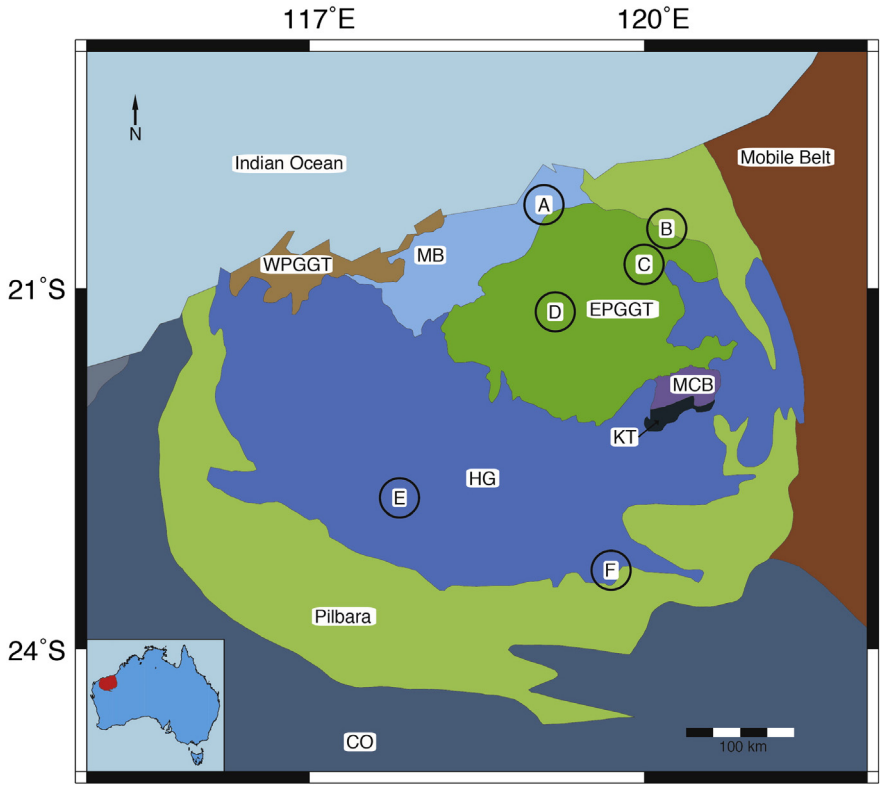

Fig. 2. Overview of key geological terranes in the Pilbara Craton based on Hickman (2004). (CO) Capricorn Orogen; (EPGGT) Eastern Pilbara Granite–Greenstone Terrane; (HG) Hamersley Province; (KT) Kurunna Terrane; (MB) Mallina Basin; (MCB) Mosquito Creek Basin; (WPGGT) Western Pilbara Granite–Greenstone Terrane. Black circles represent some iron deposits (A) Pardoo-Mt Goldsworthy; (B) Cattle Gorge; (C) Spinifex Ridge; (D) Abydos; (E) Mt. Tom Price; and (F) Mt. Whaleback. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

silica leaving chert free BIF (Lascelles, 2007). As silica removal occurs prior to lithification, these models solve the recurring problem of silica removal that both supergene and hypogene–supergene models must account for.

Supergene models for iron ore formation have been proposed since the beginning of the twentieth century (e.g. Leith, 1903) and have grown in prominence and support since the midtwentieth century (e.g. Macleod, 1966; Morris, 1980), though the focus of these models

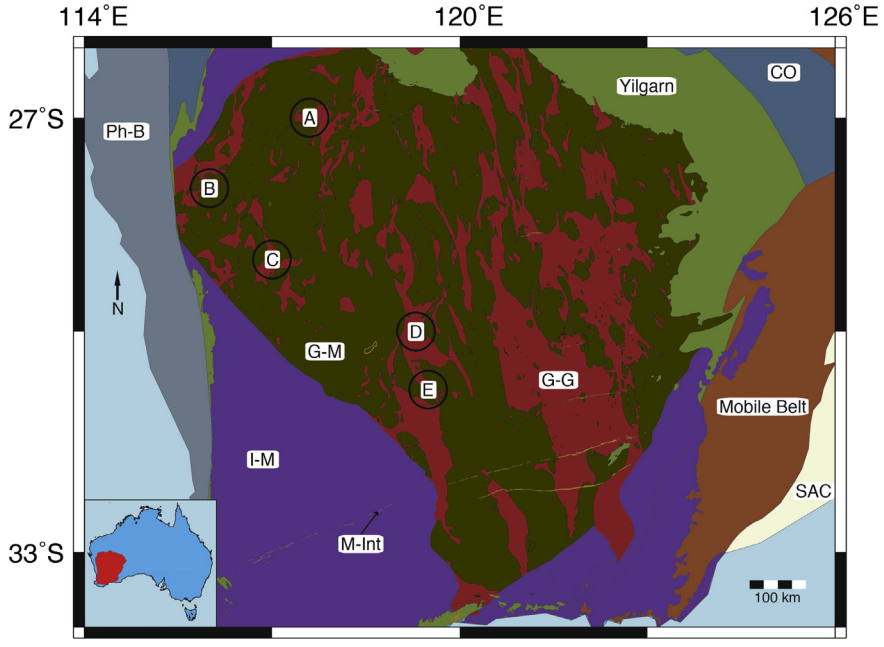

Fig. 3. Overview of key geological features in the Yilgarn Craton (excluding regolith) based on Tyler and Hocking (2007, 2008). (CO) Capricorn Orogen; (G–G) Granite–greenstone belts; (G–M) granite–mafic complexes; (I–M) igneous–metamorphic complexes; (M-Int) mafic intrusions; (Ph-B) Phanerozoic Basins; (SAC) South Australian Craton. Black circles represent some iron deposits (A) Madoonga; (B) Tallering Peak; (C) Karara; (D) Windarling; and (E) Koolyanobbing. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Please cite this article as: Merdith, A.S., et al., Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier, Ore Geol. Rev. (2015), http://dx.doi.org/10.1016/j.oregeorev.2015.03.014

tends to be on Australian and North American deposits. The basis of the supergene models is that surficial processes over long periods of time have caused the enrichment of BIF through preferential leaching of silica and deposition of iron. Advancements in these models have progressed from enrichment occurring continuously over the Proterozoic and Phanerozoic (Macleod, 1966) to it exclusively occurring within shorter time intervals of weathering during the Palaeoproterozoic (Morris, 1985). These weathering events formed martite–goethite (M–G) ores, and, in small number of cases, was followed by a period of low grade metamorphism during the Proterozoic, converting the M–G to martite–microplaty hematite (M–mplH) ores with some residual goethite (Morris and Kneeshaw, 2011). Subsequent erosion during the Mesozoic and Cenozoic exposed this ore to leaching, which removed the residual goethite leaving high quality M–mplH ores (Morris and Kneeshaw, 2011).

A key problem faced by supergene models are mechanisms for removal of large amounts of silica and enrichment of iron at depths. Morris (1985) proposed the development of electrochemical cells allowing for silica at depth to be replaced with iron from the near-surface. In this case the cathode occurs within the vadose zone of the BIF, and the anode at depth, where iron is transported as the soluble Fe^{2+} but oxidised to the insoluble Fe^{3+} at the anode and then precipitated by hydrolysis. This results in an upward growing ore body, and the silica is liberated at depth and then removed via flowing aquifer systems into surface drainage (Morris and Kneeshaw, 2011). Alternatively, infiltration of high-pH, hypersaline brines have also been proposed to assist with the removal of silica, with the concentration of iron occurring through weathering and leaching of near-surface BIF and precipitation at depth (e.g. Evans et al., 2013).

Hypogene–supergene models have only recently been proposed, and have been developed based on the observation of features depictive of high temperature crystallisation as well as supergene enrichment over periods of time (e.g. Barley et al., 1999; Taylor et al., 2001). Taylor et al. (2001) outline four phases of events within the Hamersley Province, most importantly of which is an initial hydrothermal stage where large amounts of silica are stripped and removed from the BIF at depth. Later stages involve supergene processes such as oxidation of remaining iron oxides to martite and microplaty hematite, further removal of gangue material such as carbonates and finally another supergene stage concentrated the ore further with the removal of elements such as calcium and phosphorous (Taylor et al., 2001). A key, fundamental difference between these models and the supergene model are that irrespective of the ubiquity of the supergene processes, ores could only form where first they were hydrothermally altered to remove silica. Hypogene–supergene models have also been developed for deposits throughout the Yilgarn Craton, where early hypogene stages facilitate iron deposition and silica removal, and the supergene conditions occur later on and assist in enriching the ore further (e.g.Duuring and Hagemann, 2013).

The ambiguity of these models is related to a number of key points, and exacerbated by the expanse of time between the key events of formation, enrichment and present day. Primarily, a mechanism by which silica is removed from BIF in order for enrichment to occur is critical in order to explain the ore formation (Barley et al., 1999; Findlay, 1994). Secondly, the extent of deformation, weathering and metamorphism within BIF masks the primary features making it difficult to concretely determine original facies and the number of deformation or weathering events that the BIF has undergone (Klein, 2005; Lascelles, 2007). Finally, the geological controls related to iron ore genesis are important for models to encompass as they link together spatially expansive geological events with broader geological mechanisms, such as weathering, iron sources, fluid transport mechanisms and spatial and temporal constraints on iron formation (Bekker et al., 2010; Gole and Klein, 1981; James and Trendall, 1982; Klein, 2005).

3. Machine learning for mineral exploration

Machine learning is a computational method used to approach automatic classification problems involving multiple input features through learning from representative examples. Essentially, machine learning is characterised by improvement in some criteria with respect to a task over experience or exposure to examples. Its premise is that given a suitable size of data, computers should be able to determine associations and relationships, and learn from examples such that the programmer does not need to specify solutions or know the intended outcome (Alpaydin, 2010; Duda et al., 2000). This is especially important in catering for inter- and intra-class statistical variabilities that become increasingly difficult to deal with as the number of features increase. It has been used extensively in targeted marketing and the retail industry to determine buyer interests (Alpaydin, 2010), credit scoring in the bank industry to assist in risk prediction (Hand, 1998), pattern recognition for features such as facial and character (handwriting) recognition, and in the medical and biological sciences for tumour classification and biometrics.

Broadly, MLAs can be divided into two approaches, namely supervised and unsupervised learning, with both working towards a similar objective i.e. attempting to determine or exploit relationships and associations within data. Supervised learning involves data that have labels associated with it, so that the computer has a set or series of known positive and negative examples from which to learn (in the 2-class case). Unsupervised learning uses data *without* labels, and can be used to find patterns and relationships between examples within large amounts of data via methods such as clustering, density estimation (Bishop, 2006), and multivariate data projection/mapping.

Bonham-Carter (1994) has outlined two conceptual approaches to large-scale data analysis for geosciences, either empirical, data-driven approaches, or conceptual, knowledge-driven approaches. Briefly, knowledge-driven approaches are built based on human expertise, and tend to reflect established relationships between the seed points and data. Comparably, data-driven approaches tend to focus on impartially establishing objective criteria to assist in finding pertinent relationships and associations between seed points (e.g. mineral locations) and datasets. Data-driven approaches typically use MLAs such as neural networks (e.g. Brown et al., 2000; Porwal et al., 2010) or logistic regression (e.g. Chung and Agterberg, 1980) to produce target maps. The two primary concerns with data-driven methods with respect to mineral prospectivity are firstly, that the determined associations are artificial and do not actually reflect the geological relationship between the target points and the data (Carranza et al., 2008). Secondly, as mineral exploration is inherently a function of economics as well as geology, it is critical that the final result provides an acceptable trade-off between both positive and negative examples. That is, a machine learning method must achieve a high separability by reliably identifying positive targets without incurring too many false detections (i.e. areas of low prospectivity). The first concern relates to the concept of representivity, where reliance on known target examples in data is not necessarily accurate due to factors such as preservation bias. We nevertheless argue that the growth in high-resolution digital datasets and known locations is increasing the effectiveness of this approach substantially. We also argue that traditional conceptual models themselves are based on observations on known mineral formations, so the same reasoning holds for these. Manual and heuristic-based approaches are limited in their ability to cope with statistical variability, compounding further as more features/datasets are combined, and may thus be unable to discover co-associations explained therein.

We propose that the most important aspect is to utilise flexible methodologies that can be improved when new knowledge/data becomes available, and we utilise methods such as the classifier with reject option (Landgrebe et al., 2006) to cope better with problems such as poorly represented classes while at the same time capitalising on known information. The second concern is addressed via the methodology used in this

Please cite this article as: Merdith, A.S., et al., Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier, Ore Geol. Rev. (2015), http://dx.doi.org/10.1016/j.oregeorev.2015.03.014

paper, where the separation of positive and negative data is the principal criterion, through training models on both target and outlier examples, and selecting a model that can encompass a complex, heterogeneous outlier class. Both concerns are also addressed, in a geoscientific framework, through ground-truthing, and by testing the classifier in regions (unseen in training) where we expect strong positive and negative responses respectively.

While MLAs have been used extensively in other fields, their application and usefulness to geoscience is lagging, with state of the art methods only recently becoming more commonplace (e.g. Cracknell and Reading, 2014). Previously, both supervised and unsupervised MLAs were used predominantly for land and vegetation classification, and mapping using Landsat and remote imaging (e.g. Huang et al., 2002; Rogan et al., 2008; Wulder et al., 2004). However, increases in available data, computational power increases, and availability of new algorithms are creating opportunities for application to more complex problems.

Groves et al. (2000) explored numerical modelling of stress mapping to predict locations of gold throughout the Yilgarn. They used a knowledge driven approach that does not require a training set of known gold locations, rather using expert knowledge on the timing of gold formation and its relationship to strain and stress regimes to determine areas of higher prospectivity. Similarly, shape analyses of geological data were used by Gardoll et al. (2000) for prediction of gold deposits in the Kalgoorlie Terrane. Their approach followed the assumption that gold deposits typically exhibit specific and identifiable geological structures, allowing for reliable detection through pattern recognition. A classification stage was also incorporated into their work by comparing geometries to a known gold deposit in order to assist with accurate prediction. Comparably, Holden et al. (2008, 2012) used a data-driven approach by developing an automated image processing technique using pattern recognition for predicting the location of gold deposits throughout the Yilgarn Craton in Western Australia and the Superior Craton in Canada. By training an algorithm on linear features present within an aeromagnetic dataset, they removed the need for knowing both the detailed geology of a target area, and a detailed model of formation, rather automatically exploiting unlabelled geophysical data. A data driven approach using support vector machines was undertaken by Abedi et al. (2012) for multiclass prediction of boreholes drilled into porphyry copper deposits on a local scale in the Kerman Province in Iran. Their approach involved a synthesis of geological, geophysical and geochemical data and was used to help identify areas for further drilling.

More recently, a number of studies have begun exploring a range of methods by which machine learning techniques can be applied to geological problems. Carranza et al. (2014) used evidential belief functions for predictive mapping of orogenic gold in the Giyani greenstone belt (GGB). Their approach determined relationships between local geological data and spatially coherent deposit occurrences to determine the percent of the GGB that is prospective for orogenic gold. Comparably, Porwal et al. (2014) used a knowledge driven approach employing fuzzy inference systems for uranium exploration in the Yeelirrie area (Western Australia). Their methodology involved establishing a series of expert fuzzy-"if" rules governing the relationship between uranium occurrence and geological and geophysical data for determining uranium prospectivity. Additionally, Andrada de Palomera et al. (2014) used a weights-of-evidence approach for gold–silver deposits in Argentina at both regional and district scales. Their approach encompassed data such as clay alteration, ASTER and lineament densities, and indicated the importance of high resolution data at smaller scales for generating more accurate prospectivity maps.

A data-mining approach to mineral exploration has become more widely used over the past few years. This approach is more useful for resources or areas of land for which little is known, as this makes it implausible to use a knowledge-driven model. For either case, large datasets are used in association with known deposits to find commonality (Tan et al., 2006) either spatially (Merdith et al., 2013) or temporally (Landgrebe et al., 2013). The commonality is then used to define a model of exploration and typically produce a prospectivity map of a target area (e.g. Carranza, 2011; Cracknell et al., 2013; Merdith et al., 2013).

4. Data and methodology

Our method takes a holistic approach to mineral exploration and adopts a multivariate, data-driven methodology using supervised MLAs. Our analysis uses twelve features calculated from six different geophysical datasets, and is flexible such that a user can choose to add or remove features that they deem important (e.g. distance to fault, degree of weathering, filtered magnetic data, shape analysis, shear and stress mapping etc.). This is to allow for the appropriation of the methodology to a wide array of minerals, such as by using shape analysis for gold exploration, and also to other geographical locations where access to some data may not be possible. The basis of the approach is to achieve good generalisation i.e. high accuracy not only on data used in training, but also on new unseen data. As such, the approach lends itself to large-scale analysis (on the order of 1000 km) for potential mineral deposits.

An important consideration is the need to protect against unseen or unsampled (non-target) classes, as even though the classifier may be trained on data that represents a target class (i.e. mineral deposits) well, it may be difficult to sample the non-targets in a representative fashion since it is highly heterogeneous (Landgrebe et al., 2006). We argue that posing the classification problem in this fashion suits the imprecise nature of geological data, where the notion of representivity has practical limitations. Consequently, our selected MLA for this study (a mixture of Gaussians with distance-based reject option classifier preceded by principal component analysis) seeks to minimise the impact that an unseen class may have during testing and classification by utilising a dual-thresholding approach for the purposes of discrimination (against known data) and rejection (against unseen data) respectively. This strategy of using dual thresholding of both the classifier outputs, and the domain occupied by each class results in a unique feature in which examples either far from the classifier decision boundary, or outside the region enclosed by the rejection threshold are rejected (Landgrebe et al., 2006). Such a strategy was first proposed by Dubuisson and Masson (1993) by introducing a distance-based reject option, which allows for explicit classification rejection for problems wherein some data falls outside the domain in feature space occupied by training samples. This is built upon the reject option first introduced by Chow (1970), but in that case, classified samples with insufficient confidence regarding class membership were rejected. Thus a safeguard was introduced to reduce misclassifications by allowing for the rejection of ambiguous data, rather than forcing a classification decision.

McLachlan and Basford (1987) outline the usefulness of mixture models for classifying or clustering multivariate data, with particular relevance to the geophysical datasets used in this study, exhibiting substantial variability. Typically a single Gaussian is unable to fully capture and describe the distribution of 'real' datasets (Bishop, 2006), but this can be overcome by incorporating a superposition of two (or more) Gaussians, allowing more complex data distributions to be approximated. As this is a less common technique for geosciences, we compare our result with a baseline classifier, a support vector machine (Vapnik, 1995). Subsequent to establishing the base classifier, we implement the distance-based reject option, thereby highlighting the importance of closing the feature space.

Recently Cracknell and Reading (2014) applied an ensemble Random Forest classifier to tackle a high dimensional, multi-class (eight classes) scenario over a small geographical area, well suited for an ensemble classifier such as Random Forests. Conversely, our scenario is only a two-class problem and has much lower dimensionality (12 dimensions) with emphasis on recognising unseen outlier classes over a large geographical area. Our experimentation involving several

Please cite this article as: Merdith, A.S., et al., Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier, Ore Geol. Rev. (2015), http://dx.doi.org/10.1016/j.oregeorev.2015.03.014

classification models were returning high results (80–90% classification accuracies) with little differences between competing classifiers, indicating that we were successfully exploiting all available discriminatory information. If we were to include other features or classes (e.g. discriminate between Algoma and Superior type deposits) that increase dimensionality, then an ensemble classifier (e.g. Random Forests) could become a more effective choice. We note that we have preferred the use of density-based classifiers rather than a discriminant classifier so that the known target data domain is preserved and rejection of unknown outlier data can be performed in a more optimal fashion.

Here we use iron ore as a case study to highlight the ease and usefulness of our chosen methodology for mineral exploration. Two key underlying assumptions of this approach are that firstly each dataset being used is independent and represents a discrete class of information about iron ore. Secondly, we assume that a geophysical characterisation of the location of iron ore deposits can be determined by a unique combination of geophysical features. Due to the predominance of iron ore being found in Western Australia, as well as computational demands, we restricted the area of study to Western Australia, though it can be easily adapted to analyse other areas. Finally, Australia presents itself as a unique and useful case study for any analysis on mineral location and prospectivity, especially considering that high quality data is made publicly available. Australia is well endowed with natural deposits, and is the only example of a country occupying the entire extent of a tectonic plate, allowing for continuous datasets that characterise the entire continental crust of a tectonic plate in high resolution.

4.1. Dataset and feature selection

As the purpose of this investigation is province-scale, greenfield reconnaissance exploration, we suggest that using as many (potentially) pertinent datasets as possible is advantageous in order to maximise discriminatory information, with the notion that each dataset may contain independent information. As such we focus on developing an algorithm that generalises well for use in large target areas, while being mindful of computational limitations. We utilise six geophysical datasets that are available with relatively high coverage and resolution, namely gravity, magnetics, topography and radiometric signal (3-band), to create an algorithm to predict iron ore occurrence that generalises well across large spatial distances and across a variety of different types of iron deposits (Table 1).

Traditionally geophysics has not been used extensively in iron ore exploration within Australia due to the ease of finding deposits, as iron ore (especially in the Pilbara) tends to be found in continuous to semi-continuous formations that can be followed from outcrops and correlated over large distances (Dentith et al., 1994). Specifically, Dentith et al. (1994) outlines that the two primary uses of geophysics in iron ore exploration are aeromagnetic surveys, which are used to confirm stratigraphic and structural controls on mineralisation, and gamma ray logging of boreholes for correlation of shale layers within the BIF. Gravity surveys have started to be used relatively recently to assist with identifying different ore types on a local-scale based on density contrasts between the ores and surrounding rock, with BHP Billiton using self organising maps to help distinguish ores in the Pilbara using an amalgamation of magnetic and gravity data (Mahanta, 2015). There is a correlation between ridges and outcropping ore bodies (e.g. Morris, 1985) on the basis that iron ore is more weathering-resistant relative to the surrounding rock, suggesting that present day topography could be a proxy for iron ore formation. Dauth (1997) suggests that the regolith regimes of the Yilgarn Craton have typical and identifiable radiometric signatures that can be determined. Additionally, considering the well-established relationship of Yilgarn and non-Hamersley Pilbara iron ore deposits with mafic greenstone belts (Angerer and Hagemann, 2010; Huston and Logan, 2004) we suggest that there may be some relevant discriminatory information contained within the radiometric data, through the determination of a signal for non-iron ore deposits (i.e. assisting in determining where regolith is and is not). We test this by omitting radiometric data from the classification chains and comparing the results to the chains where it is included.

The features that we have extracted from the data for use in the classification chain are mean and contrast-ratio (i.e. local deviation of the data) from a predefined area (1 km and 4 km respectively) around each sample point. Both mean and contrast-ratio were chosen due to the spatial size of iron ore deposits in Western Australia, with sizes ranging between 100s and 1000s of metres in length, width and depth (Angerer and Hagemann, 2010). We expect simple features to result in the most robust representation of these deposits at a coarse resolution on a broad, provincial scale, thus avoiding over-sensitivity brought about by more complex features adapting to local geographic patterns. Importantly, both features are indicative of the present day relationship between iron ore and the geophysical datasets, rather than a reflection or recording of their depositional setting, original mineralogy or enrichment over time. This is either through enrichment (e.g. mineralogical end members reflected in magnetics and gravity) or products of recent geological history and events (e.g. surficial weathering and erosion of regolith and iron formations reflected in topography and radiometrics).

The advantage of our approach is that by using large datasets across spatially large areas we can draw more robust statistical conclusions about the commodity being investigated. A drawback of our approach is that while we generalise across large areas well (i.e. Western Australia), our algorithm is not adapted for local geology. Therefore, it is important to stress that we do not necessarily propose using the same datasets and features for finer detail exploration on regional to local-scales. We would however advocate using a similar methodology, but suggest greater differentiation in sample selection coupled with a more precise selection of data and features that are pertinent to the specific deposit types in question and the surrounding geology that is being analysed. For example, a user investigating regional-scale exploration within a greenstone belt would benefit from a sample selection of target/outlier points only occurring within greenstone belts (i.e. only Algoma-type deposits), and the prerogative to remove radiometric and topography data but include ASTER and geochemical data instead.

4.2. Model building

To ensure the robustness of our model we make use of a robust statistical methodology for training and validation (Fig. 4). Firstly, our classification chain incorporates randomised hold-out to ensure that testing data are independent from training data, while simultaneously revealing the stability of the classifiers (i.e. variance or stability). Secondly, known ground-truth data are used to ensure that our model can correctly identify cases of iron ore and also identify cases of non-iron ore. The validation was carried out by selecting two areas of Western Australia with which the algorithm could be applied, i.e. the Pilbara in the northwest of the state, known for its richness of iron ore deposits, and the Yeelirrie, an iron-poor area in central Western Australia (Fig. 5). The success of the classification model is based on its ability to satisfy these ground-truth predictions, such that the Pilbara would satisfy the model's ability to predict iron ore, and the Yeelirrie to predict the absence of iron ore. Finally, we compare the results by overlaying

Table 1
List of geophysical datasets used with their resolution and reference. All datasets were acquired freely from Geoscience Australia.

Dataset	Resolution	Reference
Bouguer gravity	800 m	Bacchin et al. (2008)
Magnetic anomaly	80 m	Milligan and Franklin (2010)
Topography	250 m	Hutchinson et al. (2006)
Radiometric — potassium	100 m	Minty et al. (2010)
Radiometric — thorium	100 m	Minty et al. (2010)
Radiometric — uranium	100 m	Minty et al. (2010)

Please cite this article as: Merdith, A.S., et al., Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier, Ore Geol. Rev. (2015), http://dx.doi.org/10.1016/j.oregeorev.2015.03.014

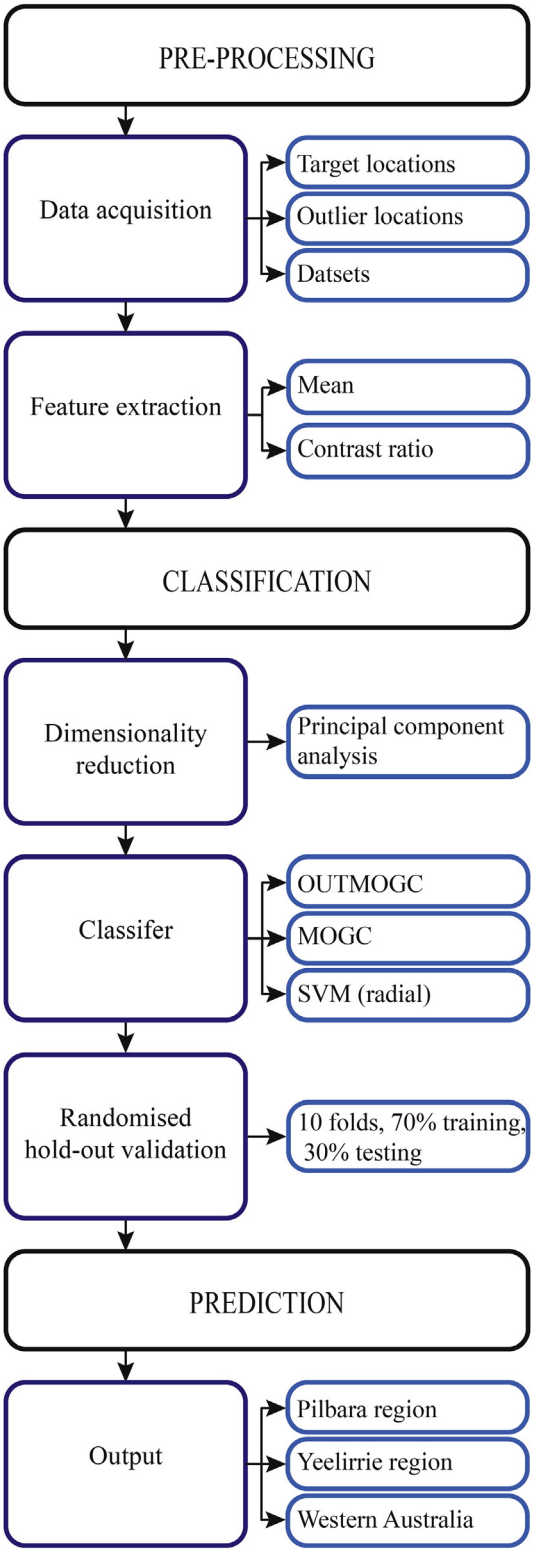

Fig. 4. Scheme of methodology using iron ore and geophysical data in Australia. The method could be applied to any mineral deposit in association with any geological data. Additionally, different features can be extracted and other classification tools can be used depending on what commodity is being targeted. Classifiers tested include a mixture of Gaussians with reject option (OUTMOGC), a mixture of Gaussians (MOGC) and support vector machine (SVM).

an independent dataset of 42 known iron ore camps (five sites occur out of range of data) that were not used in any stage of the classification chain (Ewers et al., 2001).

4.3. Pre-processing

For this study, the OZMIN Mineral Deposits Database (Ewers et al., 2002) was used to find locations of commodities within Australia. Selecting iron ore as our case study, a training set comprising 37 iron ore locations throughout Western Australia was used for the analysis (these points are referred to as the 'target' for the classification model). The OZMIN dataset does not differentiate the type or mineralogy of the deposits; therefore we used all 37 available locations, of which 36 are found within the Pilbara Craton, and one being located in the Yilgarn Craton (Fig. 5).

The rationale for this decision and not separating them based on deposit type is firstly the scope of this experiment, which is to design a classification chain that generalises well for iron ore, rather than trying to identify unique deposits (which would be more appropriate for regional–local scale exploration). Secondly, as discussed above, the features being extracted are not necessarily characteristic of either the deposition of iron formation or early stages of its enrichment, but rather they characterise present day signatures. This is also evident by broad similarities in ore mineralogy for deposits throughout both cratons (e.g. dominance of hematite/martite as the end member of iron ore enrichment (Cooper, 2013a,b; Greentree and Lord, 2007). The classification reject threshold was set such that the most outer 5% of target is excluded, which results in a good compromise between classification and rejection performance.

The non-target data used for classifier model training and validation involved the random selection of 1000 locations throughout Western Australia (though these were filtered down to 590 random locations due to incomplete geophysical data in some locations). These locations are referred to as 'outliers' for the classification-chain, that is they provide an indication of non-iron ore locations that are used to train the classifier to distinguish iron ore from non-iron ore. We follow two criteria outlined in Carranza et al. (2008), i.e. that outlier points are randomly generated, and they must have values for all data being used in the feature vector. A third criteria, i.e. that sampled data points must not lie in the vicinity of the target locations is not necessary to enforce where the number of outliers far exceeds targets, since the statistical classifiers aim to develop models that represent the entire population of data. Note that our approach results in a significant imbalance between the abundance of outlier and target points. Addressing the effects due to this class skew can be achieved using ROC analysis (Metz, 1978), in which the classification thresholds are shifted in accordance with desired error trade-offs (i.e. false positives and false negatives). In this way we can make use of all the target data available and ensure that non-targets are sampled very broadly without introducing artificial techniques such as super sampling or noise generation, with the emphasis of making use of available data.

Two features were extracted from the geophysical data at each sample location (both iron ore and random location); mean and contrast ratio (80th percentile pixel divided by 20th percentile pixel to give an indication of gradient in the data). To determine the mean value of a geophysical dataset at single point locations, a circle of radius of 1 km was constructed around the point and the mean of all the values within this circle was taken. This was to ensure that the extracted values were not over-sensitive to individual data spikes or noise, but rather were more representative of the area from which they were taken. For the contrast ratio, a larger circle, with a radius of 4 km, was used to ensure that any proximity to a geophysical boundary that was present would be included. Furthermore, as the locations are simple 2D points while mineral deposits occur within a spatial 3D framework, by using the radius around each point we are able to extract a more representative characterisation of an actual ore deposit. The features that were extracted from each sample location (both target and outlier) were stored in a 12-dimensional feature vector for classification.

Please cite this article as: Merdith, A.S., et al., Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier, Ore Geol. Rev. (2015), http://dx.doi.org/10.1016/j.oregeorev.2015.03.014

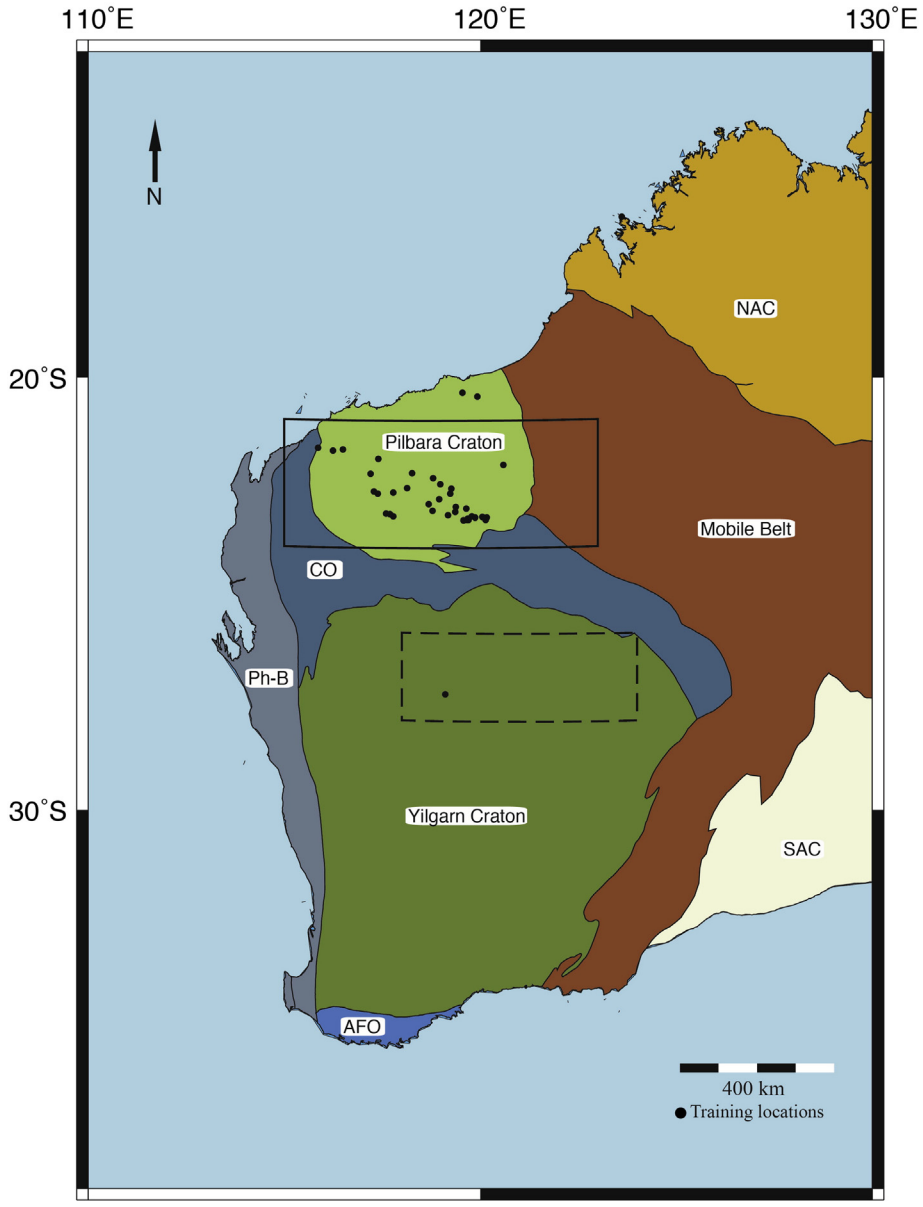

Fig. 5. Map of Western Australia showing geological provinces (see Fig. 1 for key), ground truth areas and the training locations of iron ore used for construction of the classifier. The two boxes indicate the extent of our analyses that were used for ground truth validation. The solid line box represents the Pilbara, known for its richness in iron deposits, while the dashed line box represents the Yeelirrie, an iron-poor region in the Yilgarn Craton.

4.4. Classification

There were three key stages of operation performed in the training and testing of our model. Firstly, dimensionality reduction algorithms, such as principal component analysis (PCA), were used to reduce the dimensionality to better cope with redundant correlated data while maximising variability (Jolliffe, 2002). Secondly, we tested three different classification algorithms to determine which returned the best result. All classification tools that were used are from the freely available PRTools (Duin et al., 2007) and DD_tools (Tax, 2014) libraries for Matlab. We selected a mixture of Gaussian distribution with reject option classifier (OUTMOGC) for this classification, though we compared the results with other classifiers, in particular, support vector machines (SVM) and a mixture of Gaussian classifier (MOGC). Finally, for each classification chain, the data underwent a 10-fold randomised hold-out (70% training, 30% testing) in order to assess accuracy and ensure that our classification scheme and choice of classifier was robust. We optimised each classifier by varying key parameters, such as number of Gaussians per class, and the sigmoid, *'r'*, for the SVM. Additionally, for each of the Gaussian classification chains we varied the degree of dimensionality reduction through PCA to between two and five dimensions (i.e. projecting down from 12D to 2–5D). Note that we also investigated other dimensionality reduction techniques such as the Fisher discrimination, with poor results. Finally, in order to investigate the value of the radiometric data, we also ran the same classification chains but omitting the radiometric features, using instead just the mean and gradient values of gravity, magnetics and topography.

Importantly, the flexibility of this methodology allows a user to optimise the various parameters to compare between the abilities of models to successfully classify a target, and then pick the most appropriate classifier and parameters for the commodity being targeted.

4.5. Prediction

Once the final classification chain has been selected, the classifier is trained using all available training data, and applied to the areas of

Please cite this article as: Merdith, A.S., et al., Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier, Ore Geol. Rev. (2015), http://dx.doi.org/10.1016/j.oregeorev.2015.03.014

interest by computing the probability of iron ore occurring at each pixel. Instead of forcing a classification decision, the "soft" classifier outputs (these are posterior probabilities for density based classifiers) for each location are used. For effective visualisation we converted these classifier outputs into a predictive confidence map by data scaling, thus depicting the fit to the classification model as a proxy to the likelihood of iron ore occurring. For this case study the trained classifier chain was applied to the Pilbara and Yeelirrie regions to test our ground-truth predictions (Figs. 6 and 7) and determine the success of our model. Once satisfied that the model correctly identifies both target and outlier classes, it was applied to a larger area for explorative purposes. In this case, it was applied to the entirety of Western Australia. Computational time using Matlab was 24 h on a personal laptop with a 2.6 GHz Intel Core i7 processor for ~2.2 million computations. Use of computational parallelisation computers will allow for larger, continental scale analysis in shorter times.

5. Results

5.1. Classifier results

A selection of results from the classification chains of the various classifiers is in Table 2 (for a full table of results, including other classifiers, see Supplementary material). The mixture of Gaussians with reject option (two Gaussians permitted for target and two for outlier) coupled with a PCA projection to 4 dimensions (D) was the best performing classifier, successfully identifying 87.6% of target locations and 92.7% of outlier locations, with a standard deviation of 2.6% and 7.2% respectively (9.8% mean class-weighted error rate). This is an improvement of 4% (based on error rate) over the next most effective classifier (MOGC, one Gaussian, PCA to 5D), which achieved 97.8% for target locations and 74.5% for outlier locations (standard deviations of 1.3% and 8.6% respectively) with 13.8% error rate. Comparably, the highest performance of the SVM using a radial-basis kernel ($r = 9.1$) was substantially lower, with 98.6% for target and 68.2% for outlier (standard deviations of 0.8% and 7.7% respectively) with a 16.6% error rate.

5.2. Omitting the six radiometric features

The results of the best performance of each classification chain by omitting the radiometric features are highlighted in Table 2. The mixture of Gaussians with reject option (two Gaussians for target and outlier, PCA to 4D) was the once again the most successful classifier, obtaining the same results when omitting the radiometric features as when they are included (87.6% and 92.7% success for target and outlier respectively). The mixture of Gaussians performed highest when optimised with two Gaussians coupled with PCA to 3D, achieving 97.3% on target and 70.9% on outlier (standard deviation of 1.3% and 12.7% respectively). This resulted in a classification error rate of 15.9% (compared to 9.8% achieved with the radiometric features included). The SVM was optimised with $r = 9.0$ achieving 98.5% on target and 65.5% on outlier (standard deviation of 0.9% and 8.4% respectively) and 18.0% error.

5.3. Predictive confidence maps

Our classification model indicates areas of high and low prospectivity of iron ore throughout Western Australia. Figs. 6 and 7 depict the results in the Pilbara Craton and Yeelirrie region, respectively. The Pilbara Craton (Fig. 6) is dominated by a thick band (~400 km long and ~150 km wide) of elevated prospectivity that trends northwest to southeast, and is easily identifiable as corresponding to the Mount Bruce Supergroup. Within this band there are a number of thin lineaments (~10 km wide and ~40–100 km long) of extremely high prospectivity. Additionally, there are a number of smaller features that depict an elevated and extremely high probability of iron ore. These are found predominantly in the central northern area of the Pilbara region. Further east than ~121°E the area is dominated by low probability of iron ore occurring, though there are a few smaller 'islands' of higher probability (e.g. at 22.75°S, 123°E). Comparably, the Yeelirrie region (Fig. 7) is dominated by low probability for iron ore with small islands (trending north–south) of higher probability occurring throughout the entire region.

The final output of the classification model is an application of the algorithm to the entirety of Western Australia for greenfield and first order, province-scale reconnaissance exploration (Fig. 8). Outside both the Pilbara and Yeelirrie regions, the majority of Western Australia has a low probability of iron ore, though there are notable areas of higher probability. There are a number of areas throughout the Yilgarn Craton (~25–30°S, 115–120°E) that have high probability, as well further south on the southern coast (~32.5–35°S) amongst the Albany-Fraser Orogen.

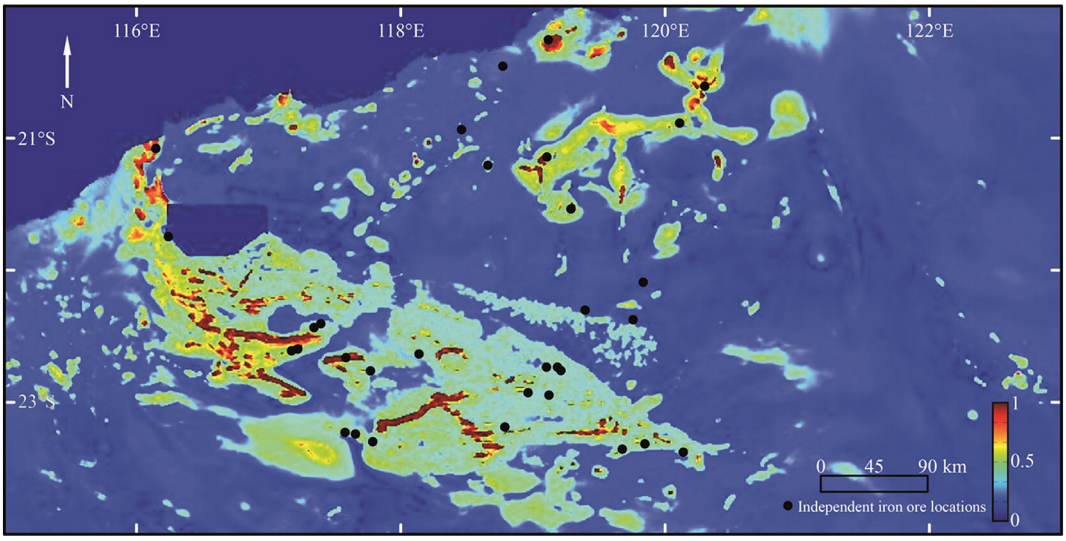

Fig. 6. The output from the trained model for the Pilbara region as a predictive confidence map depicting probability of iron ore occurrence. The independent iron ore locations (Ewers et al., 2001) have not been used in any stage of the model. As predicted by observations, large areas in the Pilbara Craton are highly probable for iron ore, indicating that our classifier can successfully distinguish positive examples.

Please cite this article as: Merdith, A.S., et al., Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier, Ore Geol. Rev. (2015), http://dx.doi.org/10.1016/j.oregeorev.2015.03.014

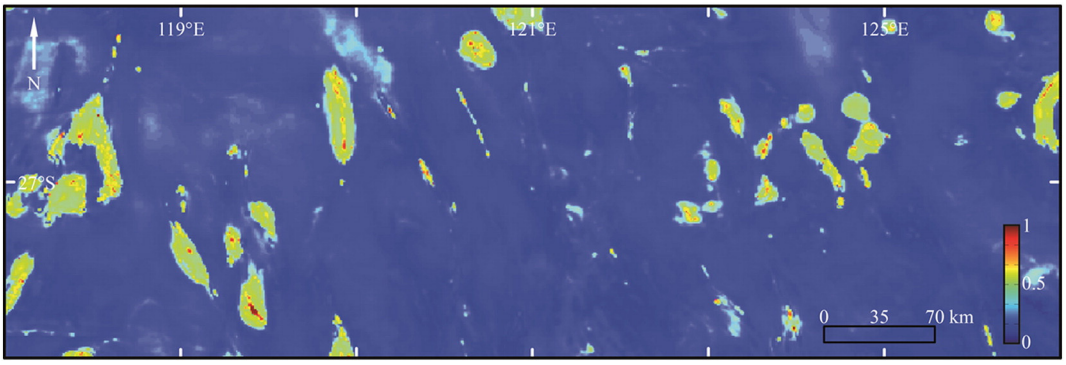

Fig. 7. The output from the trained model for the Yeelirrie region as a predictive confidence map depicting probability of iron ore occurrence. As predicted by observations, the Yeelirrie region is predominantly dull, indicating that our classifier can successfully find negative examples.

5.4. Areas of high prospectivity

In order to highlight potential areas for exploration the confidence map is bracketed into 15 intervals based on the calculated probability of iron ore occurring at each individual pixel, thus allowing for sufficiently fine resolution to identify key areas. This facilitates ranking and isolating areas of Western Australia based on their expected iron ore occurrence by depicting the areas of highest through to lowest prospectivity. Divisions 1–14 contain the same number of points (88,520 points, or ~7.04% of the total area), while the final division, division 15, contains 18,795 points, or ~1.49% of the total area and represents the areas of highest confidence as predicted by the classifier. Fig. 9 depicts the division of Western Australia into the 15 intervals, with Fig. 9a–e showing Western Australia divided into selection of three intervals and Fig. 9f showing all the divisions. The areas of lowest confidence of iron ore (Fig. 9a, intervals 1–3, and 9b, intervals 4–6) are typically concentrated around the Proterozoic mobile belts towards the east of the Pilbara Craton in the north, and large areas of the central Yilgarn Craton in the south. The areas of medium confidence (Fig. 9c, intervals 7–9) occur throughout the Capricorn and Albany-Fraser Orogeny as well towards the periphery of the Pilbara. Finally, the areas of highest confidence (Fig. 9d, intervals 10–12, and 9e, intervals 13–15) are reserved for the bulk of the Hamersley Group in the Pilbara Craton, with a number of areas scattered throughout the central-western area of the Yilgarn Craton.

6. Discussion

6.1. Classification chains

The best performing classifier chain used the mixture of Gaussian classifier with reject option. By comparing the results of the mixture of Gaussian classifiers we can determine the usefulness of the reject option (i.e. dual thresholding and closing the 'feature space'). Overall, there is an 18% improvement between the MOGC and OUTMOGC for predicting outlier locations, indicating that by closing the feature-space around the target locations is a very effective strategy for dealing with the poorly defined non-target class. This is of paramount importance for the mineral exploration industry, helping to establish more selective prospectivity maps, thus lowering the risk associated with the early stages of an exploration project and providing confidence

Table 2
Selection of results for various classification chains using the mixture of Gaussians classifier (MOGC), outlier mixture of Gaussians with reject option classifier (OUTMOGC) and support vector machine classifier (SVM). Principal component analysis (PCA) was varied for both the MOGC and OUTMOGC and not used for the SVM. The brackets after the MOGC and OUTMOGC indicate the number of Gaussians used to describe the data; in the case of OUTMOGC they are read as number of Gaussians to describe [target outlier]. Results for all classifiers are read as [true positive, false positive] in the top row and [false negative, true negative] in the bottom row. The standard deviations (SD) of each result are indicated underneath, respectively. Best performing results are in bold type.

PCA		2				3				4				5			
		12 features		6 features (grav/mag/topo)		12 features		6 features (grav/mag/topo)		12 features		6 features (grav/mag/topo)		12 features		6 features (grav/mag/topo)	
MOGC [1]	Results	0.987	0.013	0.9831	0.0169	0.9836	0.0164	0.9695	0.0305	0.9836	0.0164	0.9672	0.0328	**0.978**	**0.022**	0.9582	0.0418
		0.4091	0.5909	0.3818	0.6182	0.3727	0.6273	0.3455	0.6545	0.3364	0.6636	0.2909	0.7091	**0.2545**	**0.7455**	0.2727	0.7273
	SD	0.0084	0.0084	0.0096	0.0096	0.0094	0.0094	0.0205	0.0205	0.0068	0.0068	0.0213	0.0213	0.0128	0.0128	0.0241	0.0241
		0.0773	0.0773	0.0835	0.0835	0.0671	0.0671	0.0939	0.0939	0.1054	0.1054	0.1118	0.1118	0.0862	0.0862	0.1134	0.1134
MOGC [2]	Results	0.9881	0.0119	0.9763	0.0237	0.987	0.013	**0.9729**	**0.0271**	0.9864	0.0136	0.9582	0.0418	0.9802	0.0198	0.9537	0.0463
		0.4273	0.5727	0.3909	0.6091	0.4364	0.5636	**0.2909**	**0.7091**	0.3455	0.6545	0.2818	0.7182	0.3273	0.6727	0.3	0.7
	SD	0.0073	0.0073	0.0145	0.0145	0.0083	0.0083	0.0133	0.0133	0.0081	0.0081	0.0261	0.0261	0.0115	0.0115	0.0289	0.0289
		0.0748	0.0748	0.0862	0.0862	0.0717	0.0717	0.1271	0.1271	0.1342	0.1342	0.1317	0.1317	0.1926	0.1926	0.0862	0.0862
OUTMOGC [2 2]	Results	0.8503	0.1497	0.8503	0.1497	0.8746	0.1254	0.8746	0.1254	**0.8757**	**0.1243**	**0.8757**	**0.1243**	0.8559	0.1441	0.8554	0.1446
		0.1	0.9	0.1	0.9	0.0909	0.9091	0.0909	0.9091	**0.0727**	**0.9273**	**0.0727**	**0.9273**	0.0909	0.9091	0.0909	0.9091
	SD	0.0259	0.0259	0.0259	0.0259	0.0194	0.0194	0.0194	0.0194	0.0257	0.0257	0.0257	0.0257	0.0304	0.0304	0.0306	0.0306
		0.0671	0.0671	0.0671	0.0671	0.0606	0.0606	0.0606	0.0606	0.0717	0.0717	0.0717	0.0717	0.0606	0.0606	0.0606	0.0606
Radial parameter		r8.9				r9				r9.1				r9.2			
SVM (radial)	Results	0.9853	0.0147	**0.9853**	**0.0147**	0.9836	0.0164	**0.9853**	**0.0147**	**0.9864**	**0.0136**	0.9853	0.0147	0.9853	0.0147	0.9876	0.0124
		0.3273	0.6727	**0.3545**	**0.6455**	0.3182	0.6818	**0.3455**	**0.6545**	**0.3182**	**0.6818**	0.4273	0.5727	0.3273	0.6727	0.4636	0.5364
	SD	0.0076	0.0076	0.0089	0.0089	0.0094	0.0094	0.0089	0.0089	0.0085	0.0085	0.0085	0.0085	0.0089	0.0089	0.0052	0.0052
		0.0977	0.0977	0.1	0.1	0.0982	0.0982	0.0835	0.0835	0.0773	0.0773	0.1821	0.1821	0.0977	0.0977	0.145	0.145

Please cite this article as: Merdith, A.S., et al., Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier, Ore Geol. Rev. (2015), http://dx.doi.org/10.1016/j.oregeorev.2015.03.014

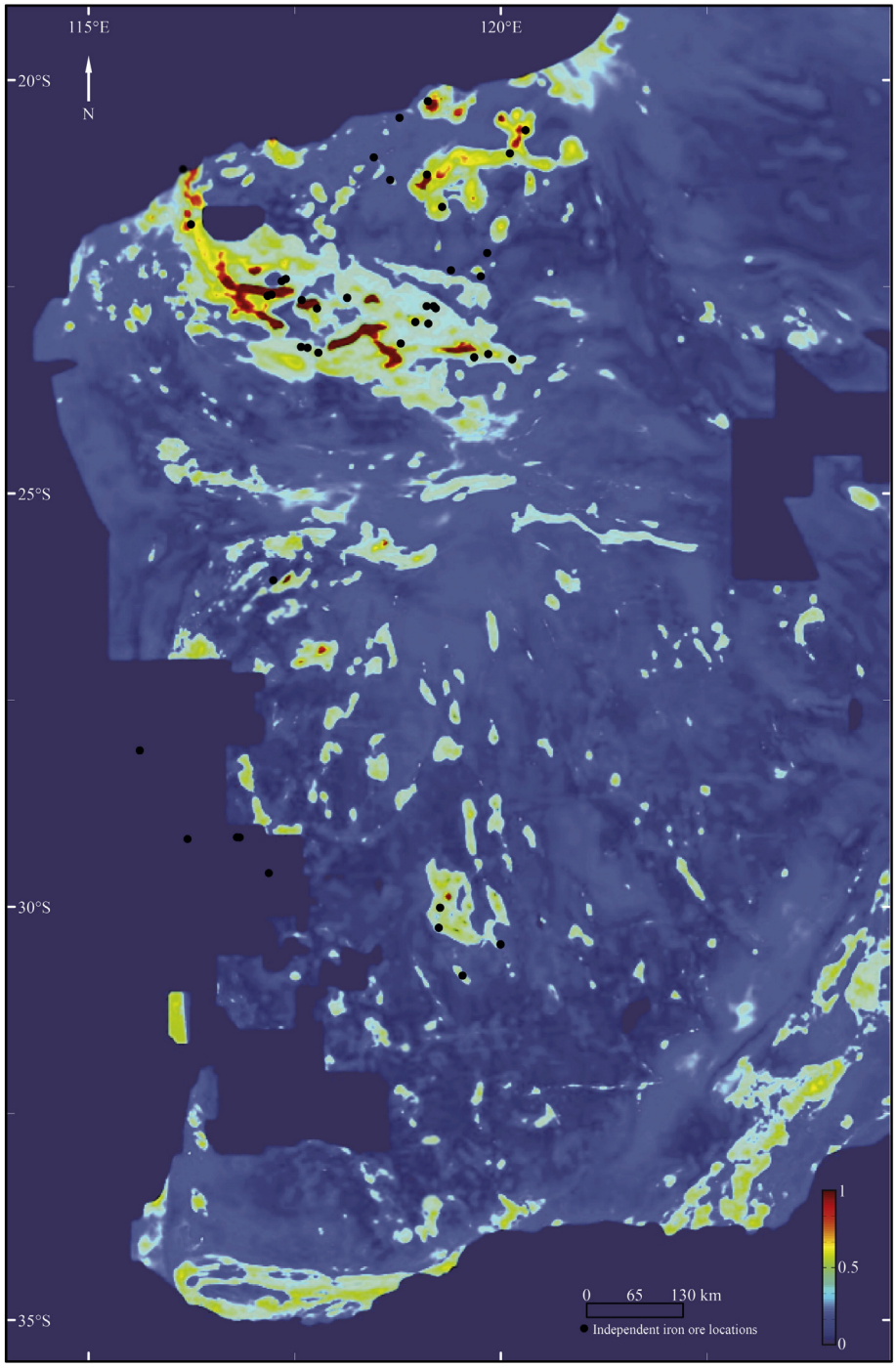

Fig. 8. The output from the trained model for Western Australia. The independent iron ore locations (Ewers et al., 2001) have not been used in any stage of the model. Due to some of the geophysical datasets being incomplete, some areas have not been computed.

that the final map accurately portrays areas of both high and low prospectivity for the targeted mineral. The mixture of Gaussian classifier also substantially outperforms SVM for this application, attributed to the large degree of variability and class overlap exhibited by the chosen features. These are clearly modelled well using the mixture model algorithm, which is a density-based approach (i.e. places more emphasis on regions in feature space with more abundant data), as opposed to SVM which is discriminative (focusing on data on the overlap region only). We note that increasing the dimensionality by selecting more features such as through including geochemical or geological data could improve the performance of the SVM relative to both the MOGC and OUTMOGC.

6.2. Value of radiometric data

The comparison between the 12 feature classification chains (where radiometric data was included) and six feature chains (where it was omitted) shows no improvement in our best performing classifier, OUTMOGC, and a 4–5% improvement in performance for the 12 feature classification chains using both the MOGC and SVM classifier. This suggests that there is no benefit in including radiometric data for provincial-scale iron ore exploration, though some lesser overall performance classifiers may see some improvement and demonstrates that redundant features can be incorporated within the workflow without loss of classification performance. This indicates that, assuming

Please cite this article as: Merdith, A.S., et al., Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier, Ore Geol. Rev. (2015), http://dx.doi.org/10.1016/j.oregeorev.2015.03.014

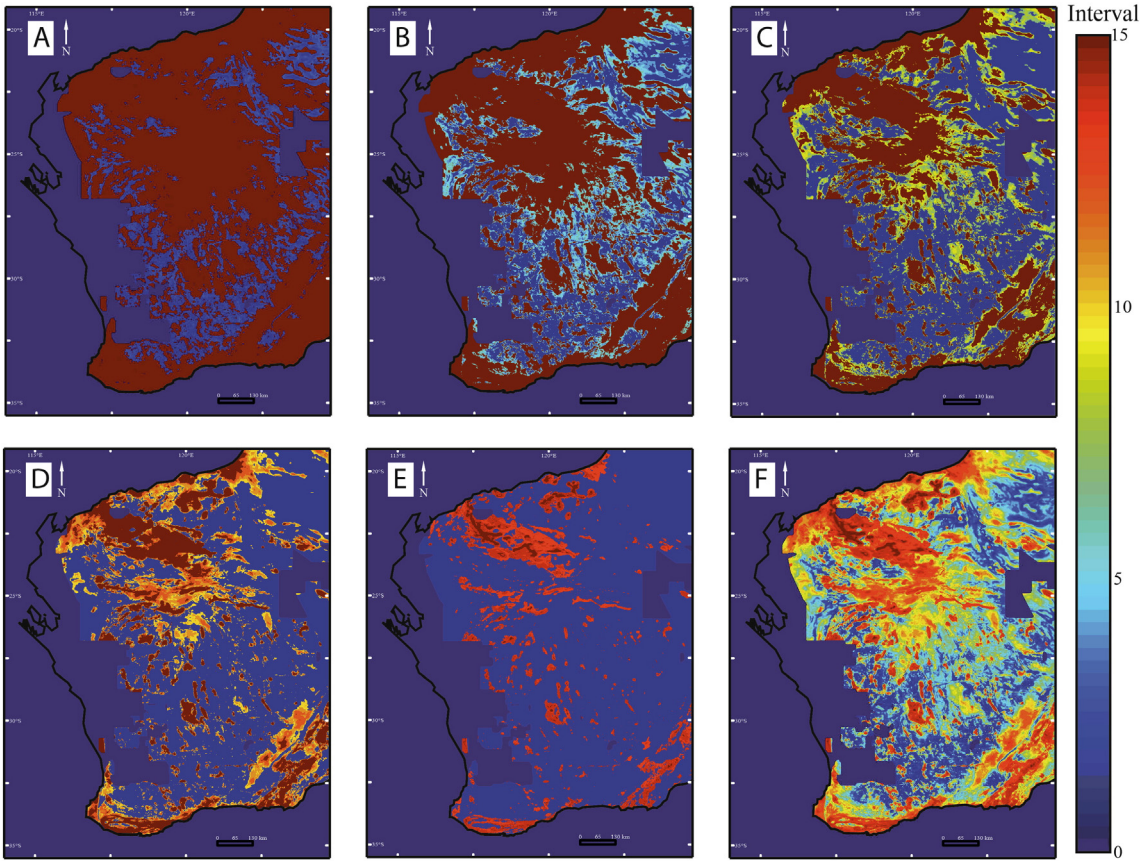

Fig. 9. Predictive confidence map for Western Australia using bracketed intervals based on the probability of a pixel to contain iron ore; (A) intervals 1–3 (lowest probability, 0–21.1%); (B) intervals 4–6 (low probability, 21.1–42.2%); (C) intervals 7–9 (medium probability, 42.2–63.3%); (D) intervals 10–12 (high probability, 63.3–84.4%); (E) intervals 13–15 (highest probability, 84.4–100%), and; (F) all intervals. The dull blue colour (interval 0) represents areas where there is no geophysical data. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

computational cost is minimal, a user faced with a decision of whether or not to include a dataset in an analysis (i.e. unsure on whether a dataset contains useful discriminatory information) can include the data to maximise potentially useful discriminatory information without risking a decrease in the classification ability.

6.3. Exploration for iron ore

It is apparent from our results that our classifier has successfully distinguished a number of known iron ore deposits throughout Western Australia. Within a mineral exploration framework, the focus of our results is on their predictive success i.e. their generalisation performance, how confident we are that what we predict as iron ore is actually iron ore, such that we can trust it to accurately predict potential areas for further exploration. The success of our methodology is based on its ability to satisfy two criteria. Firstly, the satisfaction of our ground-truth hypotheses that the Pilbara is iron rich and Yeelirrie iron poor; and, secondly, that it can accurately predict the locations of existing deposits that it wasn't explicitly trained on. The first criteria establishes the veracity of our classifier's ability to recognise both iron ore and non-iron ore cases and should give an indication as to whether the algorithm is using artefacts or meaningful associations, while the second criteria builds confidence in its ability to predict in unseen areas.

The predictive confidence maps of the Pilbara and Yeelirrie regions (Figs. 6 and 7) demonstrate that our classifier is capable of distinguishing both iron ore and non-iron ore, satisfying the first criteria. The Pilbara Craton (Fig. 6) is relatively bright, showing large areas of elevated probability of iron ore. Comparably, the Yeelirrie region (Fig. 7) is relatively dull, with a few small islands of higher probability.

The results for the entire of Western Australia (Fig. 8) show an abundance of highly prospective areas throughout the Yilgarn Craton. Comparisons of our map to published maps of both areas (Cooper, 2013a, 2013b) indicate a notable similarity between areas of high probability of iron ore and already established iron ore mines, satisfying the second criteria. In the Yilgarn Craton our map correctly identifies most key iron ore localities (Fig. 10), though it misses the Tallering Peak, Karara and Extension Hill Hematite mines due to incomplete geophysical data, causing them to be omitted from the analysis. The map also suggests some new areas for greenfield exploration such as a series of narrow E–W ridges at 25°S, an area near the Twin Peaks mine, around 27°S, 116.4°E, and a few small areas around the Steeple Hill mine in the southeast of the Yilgarn Craton.

Outside the Hamersley Basin in the Pilbara region, the Marble Bar area stands outs as an area of high prospectivity, specifically a thin ridge between the current mines of Abydos and Spinifex Ridge, located between 119 and 120°E and at about 21°S. Additionally, further north along the coastline there is a distinct area of higher probability of iron ore (~20°S, 119–121°E) as well as a number of small, scattered areas just north of the Hamersley Basin (21–21.5°S, 118°E).

Our classifier is trained predominantly on iron ore deposits from the Pilbara region, which are typically volcano-sedimentary rocks, yet it successfully detects iron ore found throughout the Yilgarn Craton, which tends to host iron ore in mafic-greenstone belts. Additionally, our classifier successfully picks up all styles of mineralisation, including: supergene enriched hematite and hematite–goethite; primary banded-iron formation deposits (magnetite); stratabound clastic-hosted deposits (both supergene enriched and magnetite); and, orthomagnetic layered mafic intrusives (Fe and V–Ti). The near universality of iron ore (determined both by present mineralogy and depositional setting)

Please cite this article as: Merdith, A.S., et al., Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier, Ore Geol. Rev. (2015), http://dx.doi.org/10.1016/j.oregeorev.2015.03.014

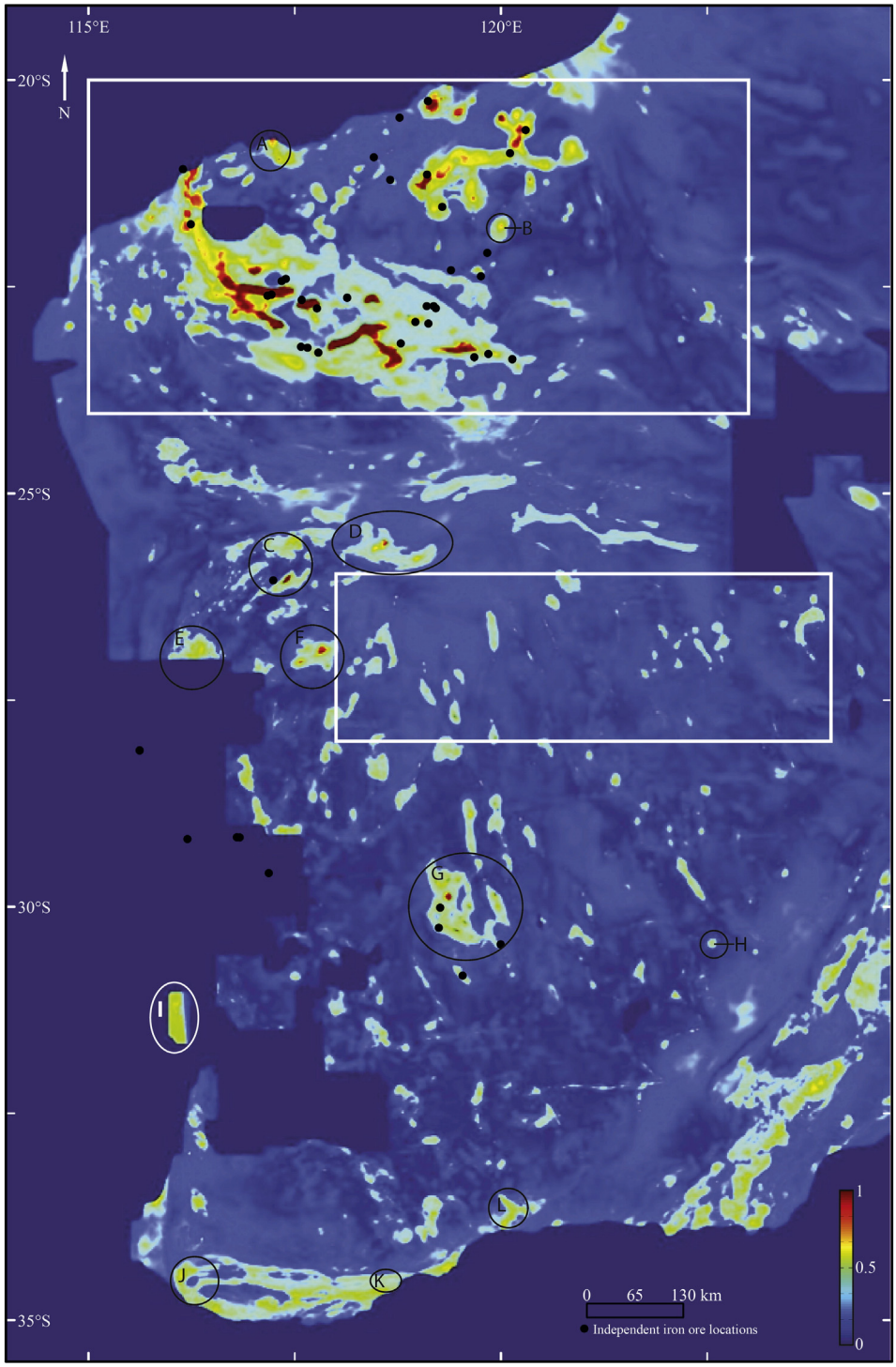

Fig. 10. Predictive confidence map for Western Australia overlayed with independent iron ore locations (Ewers et al., 2001). Open circles are mining areas throughout the Yilgarn Craton A, Cape Lambert; B, Nullagine GID; C, Dead Goat Hill, Taylor Range, Mt Gould; D, Mt Fraser, Jabiru, Valley Bore; E, Twin Peaks; F, Weld Range–Madoonga; G, Deception, Windarling, Mt Jackson; H, Steeple Hill; I, Coates (Fe and V–Ti), Crows Nest Hill, Wongamine North; J, Western Southdown; K, Southdown; and L, Ravensthorpe.

detected within Australia suggests that this methodology could be used globally for iron ore exploration.

6.4. Relationship of predictive map with regional geology

Comparing the predictive confidence map with the broad geological features of the Yilgarn Craton (Raymond and Retter, 2010), there is a notable similarity between the high prospective locations and the greenstone belts within the craton, especially towards the northwest where they follow the extent of the greenstone belts closely. Towards the southeast of the Yilgarn Craton, the areas of higher prospectivity still occur within greenstone belts though they occupy a considerably smaller extent of the belts than in the northwest. At a finer scale, comparing the predictive map with the geology of Western Australia, there is a correlation in the southwest Yilgarn Craton and in the Marble Bar area (e.g. between Abydos and Spinifex Ridge) of the Pilbara Craton between the higher areas of prospectivity and outcropping bands of mafic and ultramafic extrusive and intrusive rocks, and sedimentary–volcanic units.

Please cite this article as: Merdith, A.S., et al., Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier, Ore Geol. Rev. (2015), http://dx.doi.org/10.1016/j.oregeorev.2015.03.014

This is likely a reflection of their mode of formation being associated with submarine volcanic processes akin to Algoma type deposits, rather than sedimentary derived Superior type deposits. It is important to note that while the areas of high prospectivity occur overlapping these units, not all mafic and ultramafic units are correlated with the higher probability of iron ore. Furthermore, this correlation is consistent with the associated geology of already established Yilgarn deposits such as the Koolyanobbing deposit (Angerer and Hagemann, 2010) and Madoonga deposit (Duuring and Hagemann, 2013).

6.5. Application to economic geology

The primary focus of this paper is to present a new methodology for low expenditure, greenfield exploration. For iron ore in Australia, we have used simple and easily extractable features to demonstrate how machine learning techniques can be coupled to exploration problems in order to arrive at innovative solutions. The key to our methodology is the three-step process; pre-processing, feature extraction and classification, coupled with rigorous validation to ensure its generalisation. The preliminary data and feature selection phase is critical as it allows for a human to provide expert, knowledge-based input into the process by identifying features and parameters that are associated with different deposit types (e.g. edge and texture analysis for gold exploration). It also ensures the separation of organising the data to be analysed with the classification training, cutting down on computational cost and time, allowing for the user to easily switch between or compare different classification models and parameters. The second phase, classification, is general, allowing for the user to test and compare alternative classification models and parameters. It is expected that in exploring for other mineral types in other geographical areas, different algorithms for both dimensionality reduction and classification should be used in order to best determine the location of mineral deposits. Importantly, our method lends itself to continuous improvement over time, as more datasets are generated (e.g. ASTER multispectral data), other datasets improve in coverage and quality, and new machine learning algorithms are developed.

7. Conclusions

We present a versatile, low cost, generic methodology for greenfield mineral exploration using iron ore as a case study. Our selected classifier, a mixture of Gaussians with reject option preceded by principal component reduction, identified 88% of iron ore locations and eliminated 92% of non-iron ore locations during randomised hold-out validation and successfully fulfilled our ground-truthing hypotheses. The nature of our methodology is such that it lends itself to continual improvement over time as data increases in coverage and resolution, computational power increases and techniques develop further. It is also highly adaptable, allowing for users to target different commodities at different scales, with data and features that they believe are pertinent or have access to. The use of the reject option in the classifier (i.e. enclosing the domain in feature space around the target class) resulted in an overall ~4% increase in classification ability and a reduction in the misclassification rate for the outlier class, while suffering only a small loss in the performance for the target class. The use of reject option classifiers for greenfield mineral prospectivity is promising for protecting a small target class against a large heterogeneous outlier class. We test the value of radiometric data for iron ore exploration and show that it is of minimal use for exploration, though our results demonstrate that, with effective processing of data and algorithm choice, potentially useful data can be included without risk of penalising classification ability. The near universality of iron deposits detected throughout both the Pilbara and Yilgarn Craton by our classifier indicates that the algorithm could be easily applied for iron ore exploration in other regions that are well covered by geophysical data (e.g. Canada).

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.oregeorev.2015.03.014.

Conflict of interest

There is no conflict of interest in the work completed in this study or its submission to *Ore Geology Reviews* that the authors are aware of.

Acknowledgements

This research was supported by the Australian Research Council (FL099224) and the Science Industry Endowment Fund (RP 04-174) Big Data Knowledge Discovery project. The authors thank two anonymous reviewers and John M. Carranza for their comments that greatly improved the manuscript.

References

Abedi, M., Norouzi, G.H., Bahroudi, A., 2012. Support vector machine for multi-classification of mineral prospectivity areas. Comput. Geosci. 46, 272–283.
Alpaydin, E., 2010. Introduction to machine learning. 2nd ed. Methods in Molecular Biology vol. 1107. Massachusetts Institute of Technology, Cambridge.
Anand, R.R., Paine, M., 2002. Regolith geology of the Yilgarn Craton, Western Australia: implications for exploration. Aust. J. Earth Sci. 49 (1), 3–162.
Andrada de Palomera, P., van Ruitenbeek, F.J., Carranza, E.J.M., 2014. Prospectivity for epithermal gold–silver deposits in the Deseado Massif, Argentina. Ore Geol. Rev. http://dx.doi.org/10.1016/j.oregeorev.2014.12.007.
Angerer, T., Hagemann, S.G., 2010. The BIF-hosted high-grade iron ore deposits in the Archean Koolyanobbing Greenstone Belt, Western Australia: structural control on synorogenic- and weathering-related magnetite-, hematite-, and goethite-rich iron ore. Econ. Geol. 105, 917–945.
Bacchin, M., Milligan, P.R., Wynne, P., Tracey, R., 2008. Gravity Anomaly Grid of the Australian Region. Geoscience Australia, Canberra.
Barley, M.E., Pickard, a.L., Hagemann, S.G., Folkert, S.L., 1999. Hydrothermal origin for the 2 billion year old Mount Tom Price giant iron ore deposit, Hamersley Province, Western Australia. Mineral. Deposita 34, 784–789.
Bekker, A., Slack, J.F., Planavsky, N.J., Krapež, B., Hofmann, A., Konhauser, K.O., Rouxel, O.J., 2010. Iron formation: the sedimentary product of a complex interplay among mantle, tectonic, oceanic, and biospheric processes. Econ. Geol. 105, 467–508.
Bishop, C.M., 2006. In: Jordan, M., Kleinberg, J., Schölkopf, B. (Eds.), Pattern Recognition and Machine Learning. Springer Science, Singapore.
Blake, T.S., Barley, M.E., 1992. Tectonic evolution of the Late Archaean to Early Proterozoic Mount Bruce Megasequence set, Western Australia. Tectonics 11 (6), 1415–1425.
Bonham-Carter, G.F., 1994. Geographic Information Systems for Geoscientists. Pergamon, New York.
Brown, W.M., Gedeon, T.D., Groves, D.I., Barnes, R.G., 2000. Artificial neural networks: a new method for mineral prospectivity mapping. Aust. J. Earth Sci. 47 (4), 37–41.
Canfield, D.E., 2005. The early history of atmospheric oxygen: homage to Robert M. Garrels. Annu. Rev. Earth Planet. Sci. 33, 1–36.
Carranza, E.J.M., 2011. From predictive mapping of mineral prospectivity to quantitative estimation of number of undiscovered prospects. Resour. Geol. 61 (1), 30–51.
Carranza, E.J.M., Hale, M., Faassen, C., 2008. Selection of coherent deposit-type locations and their application in data-driven mineral prospectivity mapping. Ore Geol. Rev. 33 (3–4), 536–558.
Carranza, E.J.M., Sadehgi, M., Billay, A., 2014. Predictive mapping of prospectivity for orogenic gold, Giyani greenstone belt (South Africa). Ore Geol. Rev. http://dx.doi.org/10.1016/j.oregeorev.2014.10.030.
Chow, C.K., 1970. On optimum recognition error and reject tradeoff. IEEE Trans. Inf. Theory 16 (1), 41–46.
Chung, C.F., Agterberg, F.P., 1980. Regression models for estimating mineral resources from geological map data. Math. Geol. 12 (5), 473–488.
Cooper, R. 2013a. Iron deposits of the Pilbara Region — 2013 (1:750000 scale). Geological Survey of Western Australia.
Cooper, R. 2013b. Iron ore deposits of the Yilgarn Craton — 2013 (1:1500000 scale). Geological Survey of Western Australia.
Cracknell, M.J., Reading, A.M., 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. Comput. Geosci. 63, 22–33.
Cracknell, M.J., Reading, A.M., McNeill, A.W., 2013. Mapping geology and volcanic-hosted massive sulfide alteration in the Hellyer–Mt Charter region, Tasmania, using Random Forests™ and self-organising maps. Aust. J. Earth Sci. 61 (2), 287–304.
Dauth, C., 1997. Airborne magnetic radiometric and sat imagery for regolith mapping in the Yilgarn Craton of Western Australia. Explor. Geophys. 28 (2), 199–203.
Dentith, M.C., Frankcombe, K.F., Trench, A., 1994. Geophysical signatures of Western Australian mineral deposits: an overview. Explor. Geophys. 25, 103–160.
Dubuisson, B., Masson, M., 1993. A statistical decision rule with incomplete knowledge about classes. Pattern Recogn. 26 (1), 155–165.
Duda, R., Hart, P., Stork, D., 2000. Pattern Classification. 2nd ed. Wiley, New York.

Duin, R.P.W., Juszzak, P., Paclik, P., Pekalska, E., de Ridder, D., Tax, D.M.J., Verzakov, S., 2007. PRTools4.1, A Matlab Toolbox for Pattern Recognition. Delft University of Technology.

Duuring, P., Hagemann, S., 2013. Genesis of superimposed hypogene and supergene Fe orebodies in BIF at the Madoonga deposit, Yilgarn Craton, Western Australia. Mineral. Deposita 48, 371–395.

Evans, K.A., McCuaig, T.C., Leach, D., Angerer, T., Hagemann, S.G., 2013. Banded iron formation to iron ore: a record of the evolution of Earth environments. Geology 41 (2), 99–102.

Ewers, G.R., Evans, N., Kilgour, B., 2001. MINLOC Mineral Localities Database. Geoscience Australia, Canberra.

Ewers, G.R., Evans, N., Hazell, M., Kilgour, B., 2002. OZMIN Mineral Deposits Database. Geoscience Australia, Canberra.

Findlay, D., 1994. Diagenetic boudinage, an analogue model for the control on hematite enrichment iron ores of the Hamersley Iron Province of Western Australia, and a comparison with Krivoi Rog of Ukraine, and Nimba Range, Liberia. Ore Geol. Rev. 9, 311–324.

Gardoll, S.J., Groves, D.I., Knox-Robinson, C.M., Yun, G.Y., Elliott, N., 2000. Developing the tools for geological shape analysis, with regional- to local-scale examples from the Kalgoorlie Terrane of Western Australia. Aust. J. Earth Sci. 47 (5), 943–953.

Gee, R.E., Baxter, J.L., Wilde, S.A., Williams, I.R., 1981. Crustal development in the Archaean Yilgarn Block, Western Australia. In: Glover, J.E., Groves, D.I. (Eds.), Archaean Geology. Geological Society of Australia Special Publication, pp. 43–57.

Gole, M.J., Klein, C., 1981. Banded iron-formations through much of Precambrian time. J. Geol. 89 (2), 169–183.

Goodwin, A.M., 1973. Archean iron-formations and tectonic basins of the Canadian Shield. Econ. Geol. 68, 915–933.

Greentree, M.R., Lord, D., 2007. Iron Mineralisation in the Yilgarn Craton and Future Potential. Geoconferences (WA) Inc., Kalgoorlie, pp. 70–73.

Griffin, W.L., Belousova, E.A., Shee, S.R., Pearson, N.J., Reilly, S.Y., 2004. Archean crustal evolution in the northern Yilgarn Craton: U–Pb and Hf-isotope evidence from detrital zircons. Precambrian Res. 131, 231–282.

Gross, G.A., 1980. On classification of iron formations based on depositional environments. Can. Mineral. 18, 215–222.

Groves, D.I., Goldfarb, R.J., Knox-Robinson, C.M., Ojala, J., Gardoll, S., Yun, G.Y., Holyland, P., 2000. Late-kinematic timing of orogenic gold deposits and significance for computer-based exploration techniques with emphasis on the Yilgarn Block, Western Australia. Ore Geol. Rev. 17, 1–38.

Hand, D., 1998. Consumer credit and statistics. In: Hand, D., Jacka, S. (Eds.), Statistics in Finance. Arnold, London, pp. 69–81.

Hickman, A.H., 2004. Two contrasting granite–greenstone terranes in the Pilbara Craton, Australia: evidence for vertical and horizontal tectonic regimes prior to 2900 Ma. Precambrian Res. 131, 153–172.

Holden, E.-J., Dentith, M., Kovesi, P., 2008. Towards the automated analysis of regional aeromagnetic data to identify regions prospective for gold deposits. Comput. Geosci. 34, 1505–1513.

Holden, E.-J., Wong, J.C., Kovesi, P., Wedge, D., Dentith, M., Bagas, L., 2012. Identifying structural complexity in aeromagnetic data: an image analysis approach to greenfields gold exploration. Ore Geol. Rev. 46, 47–59.

Hronsky, J.M.A., Groves, D.I., 2008. Science of targeting: definition, strategies, targeting and performance measurement. Aust. J. Earth Sci. 55 (1), 3–12.

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 (4), 725–749.

Huston, D.L., Logan, G.A., 2004. Barite, BIFs and bugs: evidence for the evolution of the Earth's early hydrosphere. Earth Planet. Sci. Lett. 220, 41–55.

Hutchinson, M.F., Stein, J.A., Stein, J.L., 2006. GEODATA 9 second digital elevation model (DEM-9S) version 3. Geoscience Australia, Canberra.

Isley, A.E., 1995. Hydrothermal plumes and the delivery of iron to banded iron formation. J. Geol. 103 (2), 169–185.

Isley, A.E., Abbott, D.H., 1999. Plume-related mafic volcanism and the deposition of banded iron formation. J. Geophys. Res. 104, 15461.

James, H.L., 1954. Sedimentary facies of iron-formation. Econ. Geol. 49 (3), 235–293.

James, H.L., Trendall, A., 1982. Banded iron formation: distribution in time and paleoenvironment significance. In: Holland, H., Schidlowski, M. (Eds.), Mineral Deposits and the Evolution of the Biosphere. Springer-Verlag, New York, pp. 199–218.

Jolliffe, I., 2002. Principal component analysis. Springer Series in Statistics, 2nd ed. Springer-Verlag, New York.

Klein, C., 2005. Some Precambrian banded iron-formations (BIFs) from around the world: their age, geologic setting, mineralogy, metamorphism, geochemistry, and origins. Am. Mineral. 90 (10), 1473–1499.

Landgrebe, T.C.W., Tax, D., Duin, R.P.W., 2006. The interaction between classification and reject performance for distance-based reject-option classifiers. Pattern Recogn. Lett. 27 (8), 908–917.

Landgrebe, T.C.W., Merdith, A., Dutkiewicz, A., Müller, R.D., 2013. Relationships between palaeogeography and opal occurrence in Australia: a data-mining approach. Comput. Geosci. 56, 76–82.

Lascelles, D.F., 2006. The Mount Gibson banded iron formation-hosted magnetite deposit: two distinct processes for the origin of high-grade iron ore. Econ. Geol. 101, 651–666.

Lascelles, D.F., 2007. Black smokers and density currents: a uniformitarian model for genesis of banded iron-formations. Ore Geol. Rev. 32, 381–411.

Leith, C.K., 1903. The Mesabi iron-bearing district of Minnesota. Monogr. U. S. Geol. Surv. 43, 316.

Lyons, T.W., Anbar, A.D., Severmann, S., Scott, C.T., Gill, B.C., 2009. Tracking Euxinia in the Ancient Ocean: a multiproxy perspective and Proterozoic case study. Annu. Rev. Earth Planet. Sci. 37, 507–534.

Macleod, W.N., 1966. The geology and iron deposits of the Hamersley Range area, Western Australia. West. Aust. Geol. Surv. Bull. 117, 170.

Mahanta, A.M., 2015. +40 years of geophysics in Pilbara and beyond. International Geophysical Conference and Exhibition. ASEG-PESA, Perth.

McLachlan, G.J., Basford, K.E., 1987. Mixture Models. Marcel Dekker, New York.

Merdith, A.S., Landgrebe, T.C.W., Dutkiewicz, A., Müller, R.D., 2013. Towards a predictive model for opal exploration using a spatio-temporal data mining approach. Aust. J. Earth Sci. 60 (2), 217–229.

Metz, C.E., 1978. Basic principles of ROC analysis. Semin. Nucl. Med. 8 (4), 283–298.

Milligan, P.R., Franklin, R., 2010. Magnetic Anomaly Map of Australia. 5th ed. Geoscience Australia, Canberra.

Minty, B.R.S., Franklin, R., Milligan, P.R., Richardson, L.M., Wilford, J., 2010. Radiometric Map of Australia. 2nd ed. Geoscience Australia, Canberra.

Moon, C.J., Whateley, M.K.G., 2006. Reconnaissance exploration. In: Moon, C.J., Whateley, M.K.G., Evans, A.M. (Eds.), Introduction to Mineral Exploration, 2nd ed. Blackwell Publishing, Oxford, p. 482.

Morris, R.C., 1980. A textural and mineralogical study of the relationship of iron ore to banded iron-formation in the Hamersley Iron Province of Western Australia. Econ. Geol. 75, 184–209.

Morris, R.C., 1985. Genesis of iron ore in banded iron-formation by supergene and supergene-metamorphic processes — a conceptual model. In: Wolf, K.H. (Ed.), Handbook of Strata-bound and Stratiform Ore Deposits. Elsevier, Oxford.

Morris, R.C., Kneeshaw, M., 2011. Genesis modelling for the Hamersley BIF-hosted iron ores of Western Australia: a critical review. Aust. J. Earth Sci. 58 (5), 417–451.

Myers, J.S., 1993. Precambrian history of the West Australian Craton and adjacent orogens. Annu. Rev. Earth Planet. Sci. 21, 453–481.

Nykänen, V., Groves, D.I., Ojala, V.J., Eilu, P., Gardoll, S.J., 2008. Reconnaissance-scale conceptual fuzzy-logic prospectivity modelling for iron oxide copper–gold deposits in the northern Fennoscandian Shield, Finland. Aust. J. Earth Sci. 55 (1), 25–38.

Planavsky, N.J., McGoldrick, P., Scott, C.T., Li, C., Reinhard, C.T., Kelly, A.E., Lyons, T.W., 2011. Widespread iron-rich conditions in the mid-Proterozoic ocean. Nature 477 (7365), 448–451.

Pohl, W.L., 2011. Economic Geology Principles and Practice: Metals, Minerals, Coal and Hydrocarbons — Introduction to Formation and Sustainable Exploitation of Mineral Deposits. Wiley-Blackwell, West Sussex.

Porwal, A., Carranza, E.J.M., Hale, M., 2003. Artificial neural networks for mineral-potential mapping: a case study from Aravalli Province, Western India. Nat. Resour. Res. 12 (3).

Porwal, A., González-Álvarez, I., Markwitz, V., McCuaig, T.C., Mamuse, A., 2010. Weights-of-evidence and logistic regression modeling of magmatic nickel sulfide prospectivity in the Yilgarn Craton, Western Australia. Ore Geol. Rev. 38 (3), 184–196.

Porwal, A., Das, R.D., Chaudhary, B., González-Álvarez, I., Kreuzer, O., 2014. Fuzzy inference systems for prospectivity modeling of mineral systems and a case-study for prospectivity mapping of surficial Uranium in Yeelirrie Area, Western Australia. Ore Geol. Rev. http://dx.doi.org/10.1016/j.oregeorev.2014.10.016.

Raymond, O. L., and Retter, A. J. 2010. Surface geology of Australia 1:1000000 scale. Canberra: Geoscience Australia.

Rogan, J., Franklin, J., Stow, D., Miller, J., Woodcock, C., Roberts, D., 2008. Mapping land-cover modifications over large areas: a comparison of machine learning algorithms. Remote Sens. Environ. 112, 2272–2283.

Simonson, B.M., Schubel, K.A., Hassler, S.W., 1993. Carbonate sedimentology of the early Precambrian Hamersley Group of Western Australia. Precambrian Res. 60, 287–335.

Singer, D.A., Kouda, R., 1999. A comparison of the weights-of-evidence method and probabilistic neural networks. Nat. Resour. Res. 8 (4), 287–298.

Tan, P., Steinbach, M., Kumar, V., 2006. Introduction to Data Mining. Addison-Wesley, Boston.

Tax, D., 2014. DDTools, the Data Description Toolbox for Matlab.

Taylor, D., Dalstra, H.J., Harding, A.E., Broadbent, G.C., Barley, M.E., 2001. Genesis of high-grade hematite orebodies of the Hamersley Province, Western Australia. Econ. Geol. 96, 837–873.

Tyler, I. M., and Hocking, R. M. 2007. Tectonic units of Western Australia, 1:500000. Geological survey of Western Australia.

Tyler, I. M., and Hocking, R. M. 2008. Interpreted bedrock geology of Western Australia, 1:500000. Geological survey of Western Australia.

Van Kranendonk, M.J., Hickman, A.H., Smithies, R.H., Nelson, D.R., Pike, G., 2002. Geology and tectonic evolution of the Archean North Pilbara Terrain, Pilbara Craton, Western Australia. Econ. Geol. 97 (4), 695–732.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer, New York.

Whitaker, A.J., 2001. Components and structure of the Yilgarn Craton, as interpreted from aeromagnetic data. Proceedings of Fourth International Archaean Symposium Ext. Abstracts. AGSO-Geoscience Australia Record vol. 37, pp. 536–538 (Perth).

Wulder, M.A., Kurz, W.A., Gillis, M., 2004. National level forest monitoring and modeling in Canada. Prog. Plan. 61 (4), 365–381.

Please cite this article as: Merdith, A.S., et al., Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier, Ore Geol. Rev. (2015), http://dx.doi.org/10.1016/j.oregeorev.2015.03.014