

Artificial intelligence-based anomaly detection of the Assen iron deposit in South Africa using remote sensing data from the Landsat-8 Operational Land Imager



Glen T. Nwaila^{a,*}, Steven E. Zhang^{a,b}, Julie E. Bourdeau^{a,b}, Yousef Ghorbani^c, Emmanuel John M. Carranza^d

^a Wits Mining Institute, University of the Witwatersrand, 1 Jan Smuts Ave., Johannesburg, 2000, South Africa

^b Geological Survey of Canada, 601 Booth Street, Ottawa, Ontario, K1A 0E8, Canada

^c Department of Civil, Environmental and Natural Resources Engineering, Luleå University of Technology, SE-971 87, Luleå, Sweden

^d Department of Geology, University of the Free State, 205 Nelson Mandela Dr, Bloemfontein, 9301, South Africa

ARTICLE INFO

Keywords:

Anomaly detection
Iron deposit
Landsat-8
Remote sensing
Machine learning
Exploration
Prospecting

ABSTRACT

Most known mineral deposits were discovered by accident using expensive, time-consuming, and knowledge-based methods such as stream sediment geochemical data, diamond drilling, reconnaissance geochemical and geophysical surveys, and/or remote sensing. Recent years have seen a decrease in the number of newly discovered mineral deposits and a rise in demand for critical raw materials, prompting exploration geologists to seek more efficient and inventive ways for processing various data types at different phases of mineral exploration. Remote sensing is one of the most sought-after tools for early-phase mineral prospecting because of its broad coverage and low cost. Remote sensing images from satellites are publicly available and can be utilised for lithological mapping and mineral exploitation. In this study, we extend an artificial intelligence-based, unsupervised anomaly detection method to identify iron deposit occurrence using Landsat-8 Operational Land Imager (OLI) satellite imagery and machine learning. The novelty in our method includes: (1) knowledge-guided and unsupervised anomaly detection that does not assume any specific anomaly signatures; (2) detection of anomalies occurs only in the variable domain; and (3) a choice of a range of machine learning algorithms to balance between explainability and performance. Our new unsupervised method detects anomalies through three successive stages, namely (a) stage I – acquisition of satellite imagery, data processing and selection of bands, (b) stage II – predictive modelling and anomaly detection, and (c) stage III – construction of anomaly maps and analysis. In this study, the new method was tested over the Assen iron deposit in the Transvaal Supergroup (South Africa). It detected both the known areas of the Assen iron deposit and additional deposit occurrence features around the Assen iron mine that were not known. To summarise the anomalies in the area, principal component analysis was used on the reconstruction errors across all modelled bands. Our method enhanced the Assen deposit as an anomaly and attenuated the background, including anthropogenic structural anomalies, which resulted in substantially improved visual contrast and delineation of the iron deposit relative to the background. The results demonstrate the robustness of the proposed unsupervised anomaly detection method, and it could be useful for the delineation of mineral exploration targets. In particular, the method will be useful in areas where no data labels exist regarding the existence or specific spectral signatures of anomalies, such as mineral deposits under greenfield exploration.

1. Introduction

Satellite remote sensing is a major asset in geological mapping and/or mineral exploration, particularly in their early phases

(Diaz-Rodriguez et al., 2021). The tremendous increase in the diversity and quality of remote sensing data obtained from various platforms, as well as ground measurements, has enabled geoscientists to propose creative and effective data analysis approaches to assist exploration for a

* Corresponding author.

E-mail address: glen.nwaila@wits.ac.za (G.T. Nwaila).

variety of mineral deposit types (Rajesh, 2004; Song and Ying, 2015, 2020; Shirmard et al., 2022). Remote sensing is an important technique for the exploration of a variety of mineral deposit-types including those that are located in areas that are difficult to access or deposits that exhibit a surficial footprint but are located in regions that are politically and environmentally sensitive for ground exploration (Ciampalini et al., 2013; Ali et al., 2015). In addition, satellite-based remote sensing offers a spatial resolution that is several orders of magnitude higher than standard regional geochemical surveys (e.g., Zhang et al., 2022). Most of the commonly applied remote sensing techniques for mapping mineral deposits focus on the enhancement of satellite images using a variety of image processing techniques such as false colour composites, band ratios, and principal component analysis (PCA) (Ciampalini et al., 2013; Beiranvand Pour and Hashim, 2014). Band ratio analysis is commonly employed for the detection of mineral deposit occurrence from multispectral remote sensing images in the variable domain. It relies on mathematical combinations of band ratios to manually craft maps that proxy areal lithology and mineralogy (Agar and Coulter, 2007; Asokan et al., 2020). These methods leverage the differential spectral signature of minerals and land cover that is recorded in different bands. Band composites can be complemented with the use of PCA to explain variations in spectral information or for the identification of objects (Amer et al., 2010; Noori et al., 2019). For these manual methods, the ability to reliably extract information from spectral data depends on several factors, such as, among others, the experience of the user in selecting the best band combinations and signal processing (Shirmard et al., 2022). Banded iron formation (BIF) -hosted iron ore deposits are among the most studied deposits using remote sensing. This is in part due to the fact that most iron ore deposits exhibit a near-surface signature and occur in areas that are remote from civic infrastructure, which disfavours the use of manual survey methods (Ciampalini et al., 2013). Within the discipline of remote sensing, the portion of the spectrum most intensely reflected or absorbed by iron-bearing minerals is known (e.g., minerals containing ferric iron (Fe^{3+}) provide spectral responses that are broad and partly affected by intense atmospheric absorptions).

The general definition of data anomalousness is fluid and depends on the context of a task (e.g., known types of anomalies to target) and a background captured by data (Chandola et al., 2009). However, this can be constrained through an application of discipline-specific knowledge, which allows a reduction of the number of detected anomalies. In the unsupervised sense, descriptions of anomalies are not known a-priori and no data labels exist. An occurrence of a mineral deposit is comparatively rare in geological terms, and discovering one is certainly rare relative to the spatial coverage of satellite-based remote sensing data. However, if a mineral deposit occurrence is exposed on the surface, it constitutes a physical anomaly that could be captured and therefore targeted for detection in remote sensing data (or other types of data). Formulating the detection of a mineral deposit occurrence as an artificial intelligence task is clearly feasible. However, generalised anomalies that may be captured in remote sensing data may include not only mineral deposit occurrences but also anthropogenic structures, transient events (e.g., fires or floods) and data quality issues such as sensor failures (Barnett and Lewis, 1978; Hawkins, 1980). The latter types of non-geologically relevant anomalies may be located in any band, including bands of targeted geological relevance by coincidence, but generally, they are unlikely to contain the exact spectral signature or be of the same proportion covered by the data. This highlights the need for an approach to detect specific types of anomalies (in this case of a geological relevance) that leverages existing discipline-specific knowledge. The characteristics of differential spectral reflection of various minerals and the approximate bandwidth of interest are perhaps the most useful types of discipline-specific knowledge that could be exploited to target bands of interest to locate anomalies within a particular spectrum using artificial intelligence. This approach would essentially integrate discipline-specific knowledge with otherwise completely data-driven anomaly detection. However, as remote sensing

is often used as a reconnaissance tool that primarily supplements geophysics, geochemistry, and geology in the exploration of mineral deposits (e.g., for greenfield exploration and survey design), detected anomalies using any method require manual verification or targeted exploration to further ascertain their nature and usefulness.

In anomaly detection, using data that exhibit a spatial component, the application of an anomaly detection technique to such data can be in the spatial domain or the variable domain or a combination of both domains. The key difference between these two approaches lies in the construction of the background, which intends to approximate data normality. In the spatial domain approach, anomalies are assumed to present a spatially coherent pattern, and so digital image processing algorithms are applied to extract such patterns (e.g., rasterized objects). This has been a successful approach to spatial data, such as remote sensing data, and several types of methods have been developed that include: (1) matched filters (e.g., the Reed-Xiaoli algorithm, see Reed and Yu, 1990); (2) background refinement methods (e.g., BACON algorithm, see Billor et al., 2000; probabilistic detector, see Gao et al., 2014); and (3) bilateral filter-based methods (e.g., Yao and Zhao, 2018). Some of these spatial or digital image processing-based anomaly detection algorithms describe the image background with explicit linear representations (e.g., matrices), which are still being improved to accommodate multiple object representations (e.g., Tan et al., 2019). These techniques have limitations in their use among the geoscientific community for the detection of mineral deposits. For example, the Reed-Xiaoli algorithm is a general-purpose spatial domain anomaly detector that is functionally the opposite of PCA. In this manner, uncommon anomalous occurrences like mineral deposits can be detected in an image. However, anomalies that are often detected are not necessarily of geological relevance and thresholding of the anomalousness (to detect items that are more or less anomalous) is not driven by geoscientific intuition. In the case of bilateral filters, they attempt to preserve edges of images and can be viewed as noise reduction techniques. By reducing background noise, prominent features can be sharpened and therefore, some anomalies can be identified. However, it is naïve to expect that mineral deposits would exhibit well-defined edges or other spatial features that could be detected unambiguously using filtering methods. In general, spatial methods may not be useful for the detection of mineral deposits because their spatial footprint (e.g., geometry, continuity, and size) is usually not known a-priori and is likely to be highly inconsistent for deposits of the same type (unlike man-made structures, for which spatial anomaly detection seems to work quite well). In comparison, the variable-domain responses are essentially invariant for deposits of similar types (e.g., BIFs contain iron oxides-bearing minerals regardless of their physical footprint or location). This provides a strong basis to perform mineral deposit detection using anomaly detection in the variable domain instead in the spatial domain.

Although the analysis of remote sensing data seems to be generally comparatively more common in the spatial domain, the variable domain contains the majority of the encoded information. In the case of non-spatial data or for data that has excessively high dimensionality in the variable domain (relative to the number of spatial dimensions), anomaly detection can proceed either solely or efficaciously in the variable domain (e.g., detection of geochemical anomalies using geochemical data). Existing signal processing (not artificial intelligence-based) methods, such as spectral signature matching and spectral unmixing (Beiranvand Pour and Hashim, 2014) are effective for detecting specific anomalies where the exact signatures of the anomalies of interest are known. In the variable domain, spectral responses of spatially separated anomalies would be similar if the anomalies themselves were physically similar, which is clearly the case of mineral deposits of similar types. In the case of modern multispectral or hyperspectral data, the variable domain is dimensionally rich and typically its dimensionality far exceeds that of the spatial domain. However, detection of generalised anomalies in high-dimensional data space is computationally intensive

and requires a method to model the background in the variable domain, which, due to dimensionality, may be correspondingly more complicated than that in the spatial domain. For this task, machine learning algorithms are highly capable, and some of the more flexible types are already known to be successful in generalised anomaly detection in the variable domain (e.g., Morales-Forero and Bassetto, 2019; Zhou and Paffenroth, 2017; Zhang et al., 2021, 2022). In all such cases, the idea is that machine learning algorithms learn normal data behaviour within the dataset and replicate it through predictive modelling. In some of these cases (Zhang et al., 2021, 2022), the models capture the relationships within the data between a set of dimensions that are known to capture the type of anomaly being sought and the other dimensions. In the data-driven remote sensing context, the models serve as implicit backgrounds by reconstructing the most frequently observed behaviours in the training data under some performance constraint. Poor-reconstruction implies that the implicated data are anomalous against the background, and the reconstruction error (e.g., prediction residuals in Zhang et al., 2021, 2022) is a measure of the anomalousness of data points.

Machine learning is commonly applied to analyse spectral data to detect mineral deposit occurrence (Cracknell and Reading, 2014; Harvey and Fotopoulos, 2016; Bachri et al., 2019; Chakouri et al., 2020). The combination of remote sensing and machine learning aids geoscientists in overcoming some drawbacks of knowledge-driven approaches, such as the subjectivity of the interpreter and the simplicity of manually-crafted models (e.g., band composites and ratios), as well as some issues that are specific to remote sensing (Beiranvand Pour et al., 2016, 2019; Lary et al., 2016; Dai et al., 2017), such as model subjectivity, in pursuit of more reliable, objective and explainable prospectivity maps, which lowers exploration cost and risk. Various algorithms for many different tasks have been used (Madhuanand et al., 2021). Several machine learning applications to remote sensing have focused on land use or cover, vegetation, and water classification (Mather and Tso, 2009; Sehgal, 2012; Al-doski et al., 2013; Madhuanand et al., 2021). Although in the past, generalised anomaly detection has been demonstrated using data replication through deep learning in other fields (e.g., An and Cho, 2015), no shallow learning outside of geochemical exploration (Zhang et al., 2021, 2022) has been attempted for prospectivity mapping. To the best of our knowledge, although machine learning is frequently used for anomaly detection (especially in the general sense, outside of remote sensing and exploration), the algorithms used are typically of the deep learning variety (e.g., autoencoders) and are not desirable in many strictly scientific applications. The more explainable shallow learning variety has not yet been used to perform unsupervised anomaly detection in the variable domain of remote sensing data to locate anomalies that may represent mineral deposits.

In this study, we demonstrate that targeted data reconstruction of bands of interest that are known to contain target anomalies is an effective method to enhance and therefore detect anomalies. This work seeks to bridge two divergent approaches in exploration using remote sensing – a traditional approach that targets known and specific bands of interest (in which an anomaly is expected) and a variable-domain and data-driven approach, which was first introduced to detect geochemical anomalies in Zhang et al. (2021, 2022). We model spectral data in a non-spatial manner by entirely considering the high-dimensional relationships between various spectra or bands within the data. For our particular task of mineral deposit targeting, the objective is to create effective anomaly maps to identify potential target locations for further investigation. Our method is unsupervised and does not require manual band manipulations beyond data processing, but it is minimally guided by an a-priori knowledge of the bands of interest. We demonstrate that our method is capable of enhancing and detecting the iron-enrichment signatures of the Assen iron ore deposit (South Africa) as the most significant anomaly of any type in the study area within the bands that are known to capture iron oxide spectra, despite the presence of many

anthropogenic structures.

2. Deposit geology

The Palaeoproterozoic-aged Assen iron ore deposit in the Transvaal Supergroup is located near the centre of the Crocodile River Dome, on the eastern limb of the Bushveld Complex, South Africa (Fig. 1). The heat and pressure from the intrusion of the Bushveld Complex into the Transvaal Supergroup sediments, combined with a density difference between the intrusion and the sedimentary rocks, resulted in the deformation of the sedimentary sequences, thus forming domes in the area (Figs. 1 and 2; Gerya et al., 2003). The Assen deposit comprises iron-rich rocks that are separated into three lithofacies (progressing up-stratigraphy): (i) calcitic hematite (23 m thick; 40–50% Fe), (ii) high-grade hematite (12 m thick; >60% Fe), and (iii) BIF (>50 m thick; 35% Fe). Orebodies, which contain >60% Fe appear as irregular, tabular bodies with thickness of ~80 m and strike length of 12 km (Fig. 2). A major ENE-striking syncline marks the center of the Crocodile River Dome, east of the Assen deposit. As a result, the stratigraphic sequence at the Assen mine strikes E-W and dips between 35 and 60°N. To the east and west, the dome is surrounded by 3–4 km long, SE-NW strike-slip faults, with the western fault overlapping the western-most area of the Assen mine. Near the western fault, rocks are heavily brecciated (fault zone up to 100 m across where exposed on the surface). In the mine lease area, a variety of minor folds and faults have been observed. Lithofacies at the Assen mine have been plastically deformed, partially metamorphosed, metasomatized, and recrystallized in addition to folding and faulting. Metamorphic minerals include amphibole (tremolite), talc, calcite, crystalline quartz, and dolomite, indicating contact metamorphism temperatures of 410–510 °C (Hartzler, 1987).

3. Data and methodology

3.1. Source of remote sensing data

The data used in this study originated from the Landsat-8 Operational Land Imager (OLI) satellite, which is an American (collaboration between USGS and NASA) Earth observation satellite, launched from Vandenberg Air Force Base in California on February 4th, 2013 (USGS, 2022). It is the eighth satellite in the Landsat program. In-orbit, Landsat-8 joins Landsat-7, providing expanded coverage of the Earth's surface. The satellite is a free-flying satellite, carrying OLI and the Thermal Infrared Sensor (TIRS) instruments. These two instruments capture data across nine visible, near-infrared, shortwave infrared, and longwave thermal bands. They offer a high signal-to-noise ratio (SNR) radiometer performance, allowing for 12-bit data quantization and additional bits for better land-cover characterisation. Landsat-8 offers moderate-resolution imagery of the Earth's surface and polar areas, ranging from 15 to 100 m (Irons et al., 2012; Roy et al., 2014). Table 1 shows the properties of the Landsat 8–9 OLI.

The Assen iron ore deposit was mapped using a cloud-free Landsat-8 level 1T (L1T, terrain-corrected) images obtained from the USGS Earth Resources Observation and Science Centre (USGS-EROS at <http://earthexplorer.usgs.gov/>). The Universal Transverse Mercator (UTM) map projection and the World Geodetic System 84 (WGS84) datum were used to define the L1T products. The Landsat-8 image used in this study was collected on the 30th of December 2021 (coordinate reference system (CRS) from European Petroleum Survey Group (EPSG) 32635, Transform: Affine [30.0, 0.0, 556875.0, 0.0, -30.0, -2776785.0]). The OLI bands employed in this investigation were bands 1 through 7. The panchromatic band (band 8) and cirrus cloud band (band 9) of Landsat-8 OLI, as well as the TIRS bands, were not used in this study due to their incoherent resolution compared to bands 1 to 7 and the information that they encode for iron deposit occurrence overlaps significantly with the other bands. The study area contains a number of anthropogenic structures associated with mining and farming activity, as well as

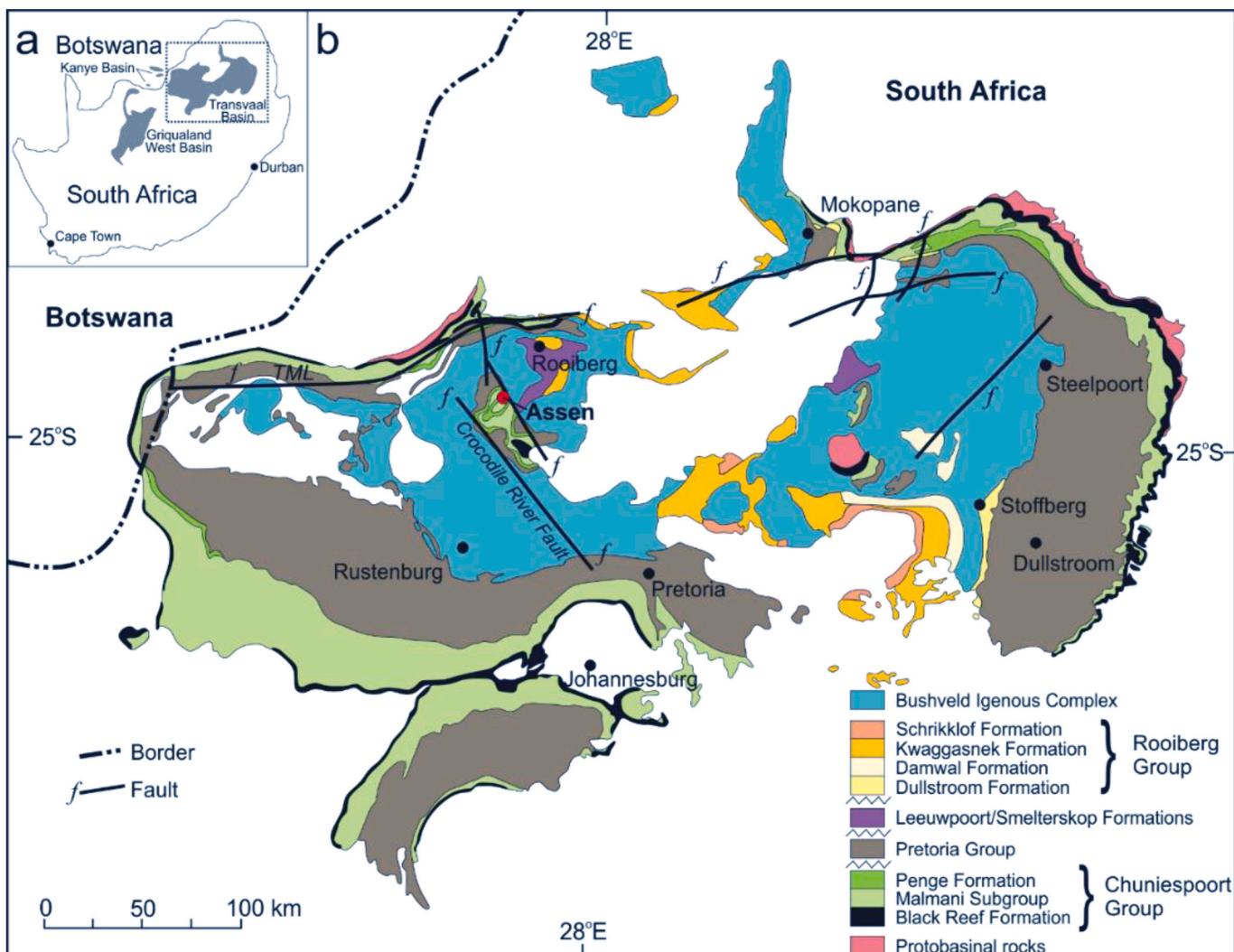


Fig. 1. Simplified geological maps of the Transvaal Basin, including the Bushveld Igneous Complex, showing the location of the Assen iron ore deposit (red dot). (a) Location of Transvaal Supergroup rocks and their basins within southern Africa. (b) Map of the Transvaal Basin. Figure modified and updated from Eriksson et al. (1995). Coordinates of the Assen Mine: 25° 7' 44.5692'' S, 27° 36' 23.292'' E. TML = Thabazimbi-Murchison Lineament.

outcrops and ore stockpiles associated with the Assen iron ore deposit (Fig. 2).

3.2. Landsat-8 image processing

For interactive mapping with Google Earth Engine (GEE), which is a cloud computing platform with a multi-petabyte database of satellite images and geospatial information, a Landsat-8 OLI image of the research area was pre-processed using the Rasterio and GEEMAP Python libraries (Wu et al., 2019; Wu, 2020). Pre-processing is concerned with the initial image treatments. Radiometric calibration, atmospheric correction, geometric correction, and noise removal are all addressed in the various pre-processing phases. Affine translation was applied to project the acquired satellite image into the WGS84 datum. One of the most significant processes applied to spectral remote sensing data for mineral exploration and lithological classification is radiometric correction (Rajendran and Nasir, 2014). The acquired image was optimised for best radiance, reflectance or brightness temperatures using radiometric calibration, which reduces errors in spectral pixels. The Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) approach was applied in this study, which is a first-principles atmospheric correction tool that corrects wavelengths up to 3 m in the visible, near-infrared, and shortwave infrared ranges (ENVI, 2009). Mineral

prospecting and geological mapping by remote sensing have both benefited from this pre-processing approach (Cooley et al., 2002; Salem et al., 2016). The tropical atmospheric and rural aerosol models were used to apply the FLAASH algorithm to the data (Manakos et al., 2011). Therefore, the raw imaging spectrometer radiance data were rescaled to reflectance data.

The image pre-processing stages was followed by image processing to extract spectral properties of bands with high Fe and clay minerals affinity. To perform band ratio analysis, bands were normalised using a standard scaler:

$$\text{Standard scaler (Band)} = \frac{(\text{data} - \text{mean of data})}{\text{data standard deviation}} \quad \text{Eq. 1}$$

The scaled bands were used to create false colour composite (FCC) images and were used for band ratio analysis. The FCC assigns to the red, green, and blue channels three different forms of information (e.g., image channels). Due to sheer differences in colours based on the values in the single channels, a colour composite makes it easier for humans to comprehend multichannel image data (Pohl and Van Genderen, 2014). The FCC colour composite image based on specified spectral features improves lithological and mineral discrimination (Salem et al., 2016). Band ratios are simple arithmetic combinations of different bands that allow semi-quantitative investigation of the distribution and strength of

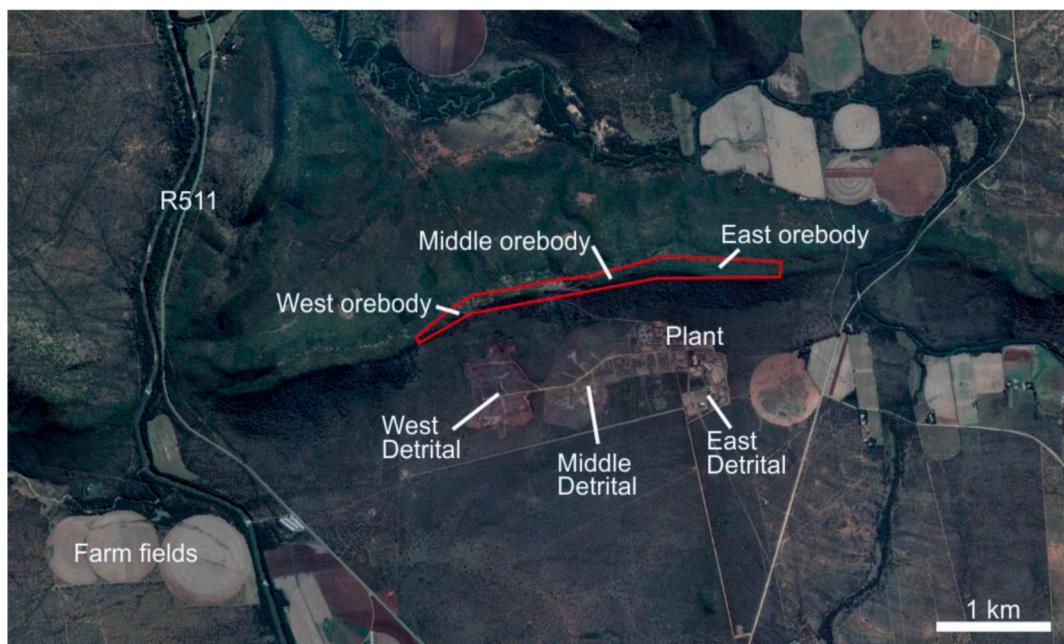


Fig. 2. Google Earth satellite image showing, agricultural structures, infrastructure, and major geological structures nearby or associated with the Assen iron ore deposit. Coordinates: West orebody = -25.127248°, 27.590647°; Middle orebody = -25.125509°, 27.597720°; East orebody = -25.124558°, 27.607325°.

Table 1

Properties of Landsat 8–9 OLI and thermal infrared sensor (<http://earthexplorer.usgs.gov/>).

Band Names and Descriptions	Wavelength (μm)	Resolution (m)
Band 1 - Coastal aerosol	0.43–0.45	30
Band 2 - Blue	0.43–0.45	30
Band 3 - Green	0.53–0.59	30
Band 4 - Red	0.64–0.67	30
Band 5 - Near Infrared (NIR)	0.85–0.88	30
Band 6 - Short-wave infrShort-wave1 (SWIR 1)	1.57–1.65	30
Band 7 - Short-wave infrShort-wave2 (SWIR 2)	2.11–2.29	30
Band 8 - Panchromatic	0.50–0.68	15
Band 9 - Cirrus	1.36–1.38	30
Band 10 - Thermal Infrared (TIRS) 1	10.6–11.19	100
Band 11 - Thermal Infrared (TIRS) 2	11.50–12.51	100

specific absorption or emission properties (Sabins, 1997). The numeric value of one band is divided by the numeric value of another band in this technique, for example. Numerous band ratios for lithological and mineralogical mapping have been extensively employed and developed depending on terrain cover and mineral deposit-type under consideration (Carranza and Hale, 2002; Gad and Kusky, 2006; Vicente and Filho, 2011; Pour et al., 2019; Shevyrev and Carranza, 2022). In this study, bands that encode information about ferrous/ferric ($\text{Fe}^{2+}/\text{Fe}^{3+}$) were used to assess anomalies of iron deposit occurrence (in the form of supra-background iron oxide enrichment) using both traditional band-ratio maps and our machine learning-based anomaly enhancement and detection method. For detection of ferrous and ferric oxide minerals, the relevant OLI bands are 4, 5 and 6, and the amplitudes of these bands were mathematically manipulated to create a knowledge-guided composite. Band ratios can mitigate illumination changes due to terrain. The ferric band ratio is sensitive to ferric iron even in low concentrations. High values of ferric minerals ratio are useful to detect the crystal-field absorption of ferric iron oxide (Rockwell, 2013). The ferrous and ferric OLI band ratios are, respectively:

$$\text{Ferrous } (\text{Fe}^{2+}) \text{ band ratio} = \frac{\text{Band 6}}{\text{Band 5}} \quad \text{Eq. 2}$$

$$\text{Ferric } (\text{Fe}^{3+}) \text{ band ratio} = \frac{\text{Band 4} + \text{Band 6}}{\text{Band 5}} \quad \text{Eq. 3}$$

Aside from iron oxide-related band ratios, there are other ones as well; for example, the clay band ratio (Eq. (4)), which leverages the fact that hydrous minerals such as clays and alunite absorb radiation in the 2.0–2.3-μm portion of the electromagnetic spectrum (Rowan et al., 1974; Goetz and Rowan 1981), thus:

$$\text{Clay band ratio} = \frac{\text{Band 6}}{\text{Band 7}} \quad \text{Eq. 4}$$

However, although OLI bands 6 and 7 are commonly used in the detection of clay minerals, the shape, and relative intensities of the reflectance curves of vegetation overlap with those of clay minerals.

3.3. Machine learning-based workflow for predictive modelling

The purpose of using machine learning algorithms in this study is to detect generalised anomalies across a set of bands in variable domain that are known to contain some information of interest (i.e., a known iron ore deposit). However, these anomalies are not known a-priori. Therefore, our task of anomaly detection is fully unsupervised. However, rather than trying all combinations of bands (which may yield a large amount of non-geologically relevant anomalies), we use the existing discipline-specific knowledge that the band ratios indicate that bands 4, 5 and 6 contain the most relevant information related to both ferric and ferrous oxides (Rockwell, 2013). This implies that, for the purpose of enhancing iron ore deposit anomalies, bands 4, 5 and 6 should be reconstructed, while the remainder of the bands can be used as features. Spectral information that is specific to bands 4, 5 and 6, which is reconstructed poorly, is therefore referred to as anomalies.

Machine learning-based predictive modelling in this application was used to model the data-driven background, which is a different approach compared to existing spatial methods (Maxwell et al., 2018). The spatial coordinates were used only to create maps but not for predictive modelling. The strength of the anomalies was measured by the

reconstruction error, for which we chose the prediction residuals, which corresponded to the predicted band amplitude minus the actual band amplitude. Once enhanced, the detection of anomalies can be accomplished through visual examination of the residuals map and/or the application of a post-processing algorithm, such as thresholding, pattern matching, continuity analysis, or other suitable spatial anomaly detection methods. In this sense, our method can serve both as a standalone method and as a pre-processing method if the results were followed by other methods, such as spatial anomaly detection through digital image processing. For this purpose, all bands were used after the image processing procedures with only an additional rescaling as required for machine learning. For data rescaling, all bands were rescaled such that their amplitudes span a uniform range between 0 and 1. This allows the bands, where used as machine learning features, to influence the algorithms equally (where the algorithms require rescaling; algorithms that do not require rescaling are unaffected). The application of the knowledge of the bands containing the most relevant information regarding iron oxides makes this approach a hybrid data- and knowledge-driven one. Therefore, our method assumes that the presence of this type of anomaly is differentially captured across the spectral bands, and this differential characteristic cannot be adequately accounted for by the general background in variable domain.

We intentionally employed shallow learning methods and included a range of algorithms to understand their feasibility. This is important for scientific purposes, where predictive modelling performance is not the most important outcome, but rather outcome explainability and method replicability. The algorithms that were explored in this study included the k-nearest neighbours (kNN), support vector machine (SVM), random forest, adaptive boosting (AdaBoost) and artificial neural networks (ANN). These algorithms represent a diversity of potentially useful approaches that include a simple non-parametric approach (kNN), a parametric approach (SVM), ensemble and boosted approaches (random forest and AdaBoost, respectively) and a neural network-based approach (ANN). Details of these algorithms, as well as their hyperparameters, are fully described in Zhang et al. (2021, 2022). Although the algorithms are supervised machine learning algorithms, the supervision was applied to build models that best capture relationships between the features and targets (to model the background in variable domain), but not between features and labelled anomalies (no labels are available within the data). The strength of relationships, as proxied by prediction residuals (predicted minus actual band amplitudes), was used to determine the anomalousness of data in target bands. In this manner, our application of machine learning is unsupervised. However, the building of predictive models requires explicit controls on model generalisability and accuracy, which were achieved through model tuning and selection using 5-fold cross-validation. We employed the coefficient of determination (CoD) metric for model tuning and selection. The parameter grid for the algorithms is given in Table 2.

Table 2

Parameters for all machine learning algorithms. For their definitions and their impact on algorithmic behaviours, see Zhang et al. (2021, 2022).

Algorithm	Parameter Grid
<i>kNN</i>	$k = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$
<i>SVM</i>	$C = \{10, 100, 200, 400, 600, 800, 1000\}$, $\epsilon = \{0.00001, 0.0001, 0.001, 0.01, 0.1, 0.5, 1.0\}$, kernel = {linear, radial basis function (RBF)}
<i>Random forest</i>	Ensemble size = {500, 1000, 1500}; maximum depth = {3, 1, unlimited}, maximum number of features = {1, 2, 3, 4}, minimum number of samples for a split = {1, 2, 3, 4}, minimum number of samples for a leaf = {1, 2, 3, 4}
<i>AdaBoost</i>	Number of classifiers = {100, 200, 500}, base algorithm = decision tree with the same parameter grid as the random forest algorithm
<i>ANN</i>	$\alpha = \{0.0001, 0.001, 0.01, 0.1, 1.0\}$, activation = {identity, logistic, tanh, relu}, learning rate = {constant, inverse scaling, adaptive}, maximum number of iterations = {1500, 2500}, hidden layer sizes = {100, 150, 200}

There are two types of anomaly maps that can be constructed from prediction residuals, i.e., single-band maps and multiple-band composites. For our purpose, we did not rely on manual band ratios to create multiple band composites for machine learning-based anomaly maps. Instead, we used PCA to re-coordinate the bands to create composites. The most significant anomalies were captured in the first principal component. In this manner, the creation of the most to least significant anomaly maps can be entirely automated. However, PCA is not the only algorithm suitable to summarise anomalies across several bands; other algorithms such as kernel PCA may be suitable as well (though not explored in this study). More complex but less explainable methods such as autoencoders can also be used to summarise anomalies, where appropriate. We consider our application of machine learning in this study to be a scientific one, and therefore demonstrate that our approach is compatible with the need for explainability. In other applications, this constraint need not apply, as concepts of uncertainty and explainability may not be applicable. To create maps, prediction residuals or combinations of prediction residuals were interpolated with the Lanczos algorithm, which can be used to smooth digital images and has been used to optimally interpolate images for resizing or rotation (Turkowski and Gabriel, 1990). The resulting images were colour-mapped using a colour map that sweeps through both luminous (from dim to bright) and chroma (from red to blue) domains to depict data contrast.

4. Results

4.1. Band visualisation and band ratio images

After image processing, bands 4, 5 and 6 demonstrate that the highest pixel intensities correspond to a number of anthropogenic structures, mainly agricultural fields (circularly shaped structures in Figs. 3–5). Band 4 captures some structures of the Assen iron ore mine, which appear as skeletal connections that are characteristic of mining infrastructure (centre of Fig. 3). The band ratio images (ferrous and ferric iron oxides, and clay) show that the anthropogenic structures exhibit high reflectance over the study area (Figs. 6–8). In addition, none of the band ratio images was able to clearly isolate the occurrence of the Assen iron ore deposit. However, the ferrous and ferric minerals ratio maps (Figs. 6 and 7) clearly show high reflectance values in the vicinity of the Assen iron ore deposit, whereas the clay minerals ratio map (Fig. 8) show the opposite behaviour. This implies that the choice of bands 4, 5 and 6 as prediction targets using machine learning was justified.

4.2. Machine learning-based anomaly enhancement and detection

Algorithm selection indicates that, on average, the tree-based methods (random forest and AdaBoost) and kNN model the backgrounds the best across bands 4, 5 and 6 (Fig. 9). The performance differential between these algorithms is often negligibly small (on the order of a few percent), which seems to suggest that complex machine learning models that are difficult to explain are unnecessary to produce high-performance and practical background models. In addition, the execution time for the kNN algorithm to perform hyperparameter optimisation was about three orders of magnitude shorter than for the tree-based methods, which implies that it would be possible to sacrifice minute amounts of absolute performance in exchange for massive reductions in computation cost. In order to best reconstruct the data, we rank-ordered the algorithms in terms of their CoD metric (R^2) scores through exhaustive cross-validation and then used the best algorithm/model for each band. Since each band was modelled using its own workflow (of algorithm and model selection), there was no particular rationale to restricting all bands to a single algorithm/model, which would lead to less optimal performance than specific algorithms/models per band. This approach was first used in Zhang et al. (2022), in which the authors noted that despite the performance differential across a

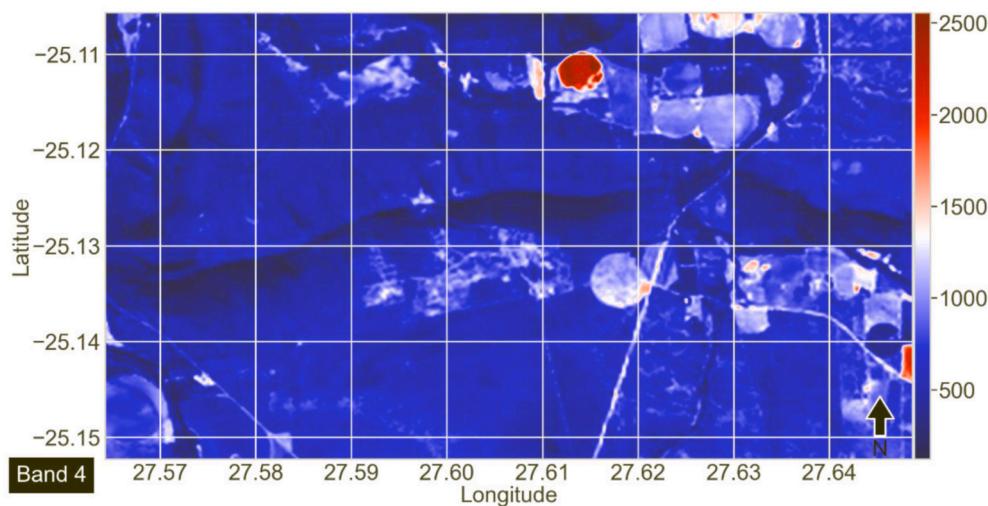


Fig. 3. Image of band 4 over the study area. The spectral response of the area is colour-mapped to the band intensity. Circular objects are agricultural structures. The skeletal and patchy structures in the centre of the image are structures associated with the Assen iron ore deposit.

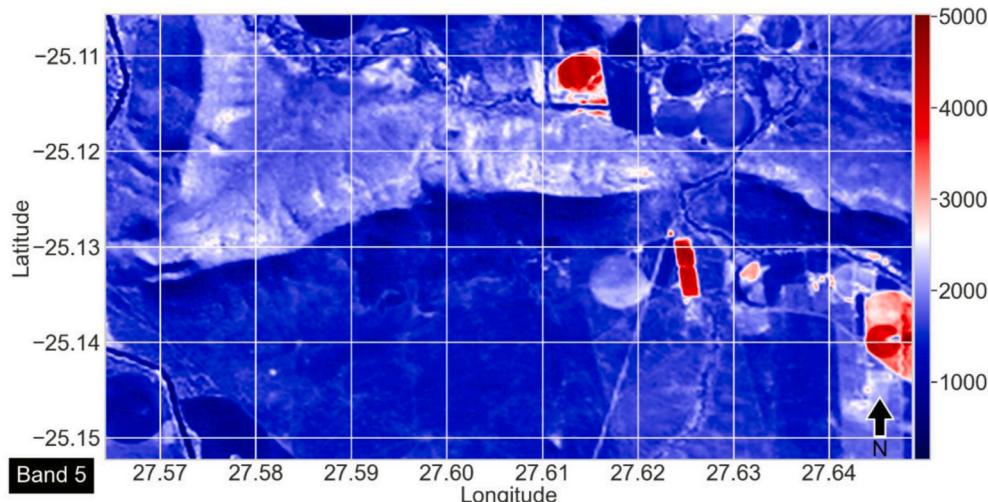


Fig. 4. Image of band 5 over the study area. The spectral response of the area is colour-mapped to the band intensity. Circular objects are agricultural structures. The Assen iron ore deposit (centre) is not visible.

variety of algorithms, the relative sizes of the anomalies both in terms of their reconstruction error and spatial footprints were qualitatively similar. The optimised hyperparameters for the best performing algorithms per band are shown in Table 3. It is worthy to note that AdaBoost came in at a close second place for band 6 (which is not obvious due to rounding to 3 decimal places of the CoD metric score in Fig. 9). The anomaly maps of band 6 produced through either random forest or AdaBoost was visually challenging to distinguish. In general, the second best and third best combinations produced visually similar maps, which were expected based on the models' CoD values (not shown). Visual examination of the predicted versus actual values for all bands (Fig. 10) demonstrates that the best prediction performance occurred for band 4, while the worst was for band 5. This indicates that the planar size of the anomalies must be the largest in band 5 and the smallest in band 4. This was indeed observed on maps of the prediction residuals. For bands 4, 5 and 6, the prediction residual maps revealed faint and smaller anomalies in band 4 (Fig. 11), a prominent anomaly in the vicinity of the iron ore deposit in band 5 (centre of Fig. 12) and even more selectively in band 6 (Fig. 13). This corresponds with the excavated footprint of the Assen deposit and ore stockpiling, both due to local mining activity. To further summarise the key anomalies within bands 4, 5 and 6, we used a

PCA-based approach to generate three maps of the largest to smallest anomalies in the area (Figs. 14–16). The principal component 1 (PC1) map of the anomalies captured the Assen iron ore deposit (centre of Fig. 14). In comparison, the PC2 and PC3 maps did not exhibit major anomalies that are known to be associated with the mineral deposits in the area at this time. In all band-specific and principal component-based maps, the anthropogenic structures were substantially suppressed compared to raw band-specific maps (Figs. 3–5) and band-ratio maps (Figs. 6–8).

5. Discussion

This section highlights the main points that distinguish our proposed method from conventional band ratio/band combination methods that are typically used to process spectral data for detecting mineral deposit occurrence. The discussion also includes comments on implications of the proposed approach for exploration targeting of mineral deposit occurrences with similar properties in other geological terranes.

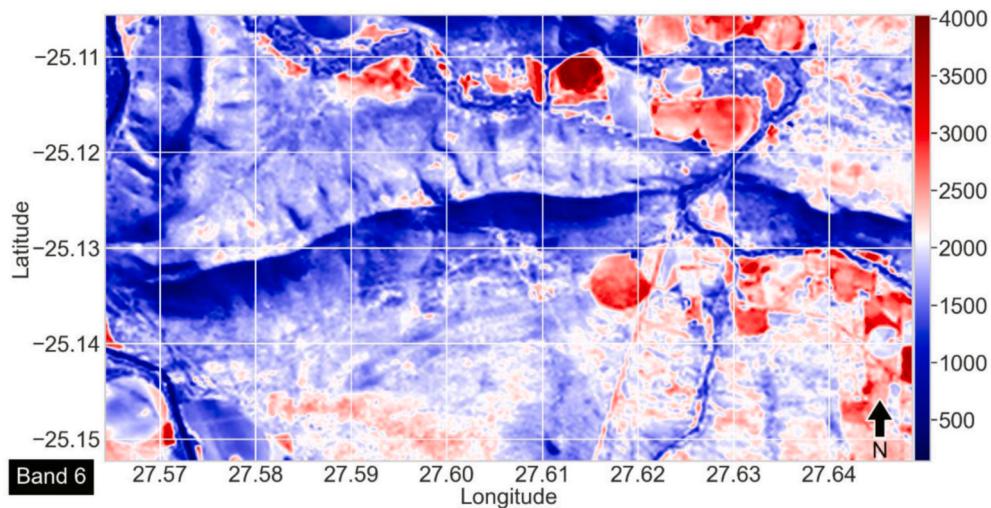


Fig. 5. Image of band 6 over the study area. The spectral response of the area is colour-mapped to the band intensity. Circular objects are agricultural structures. The Assen iron ore deposit (centre) is not well captured.

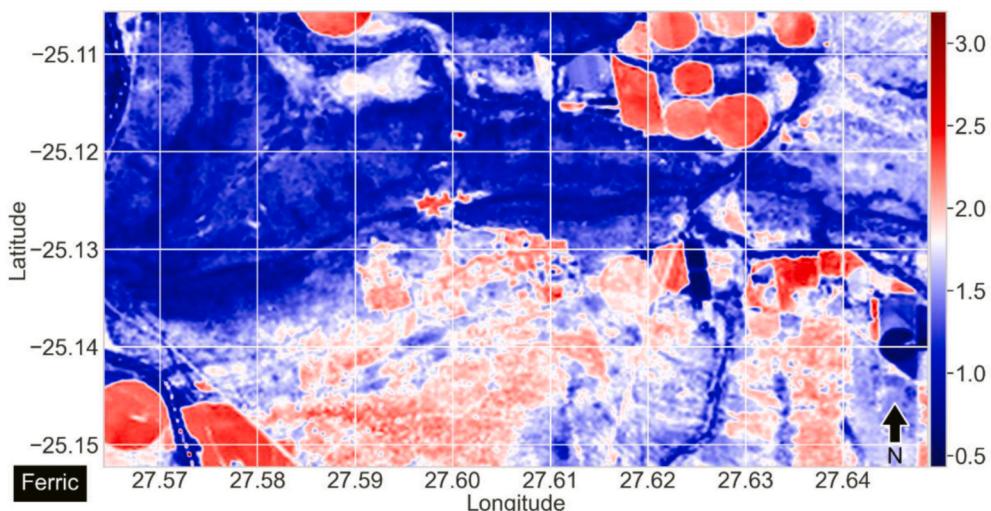


Fig. 6. Image of ferric iron oxides (warm colours) using band ratios over the study area. The spectral response of the area is colour-mapped to the ratio.

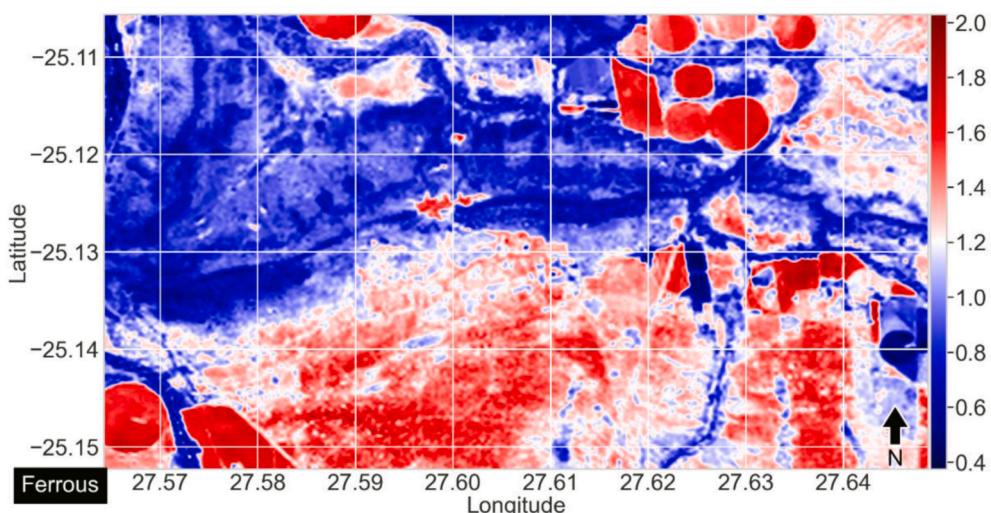


Fig. 7. Image of ferrous iron oxides (warm colours) over the study area. Image was produced using band ratios. The spectral response of the area is colour-mapped to the ratio.

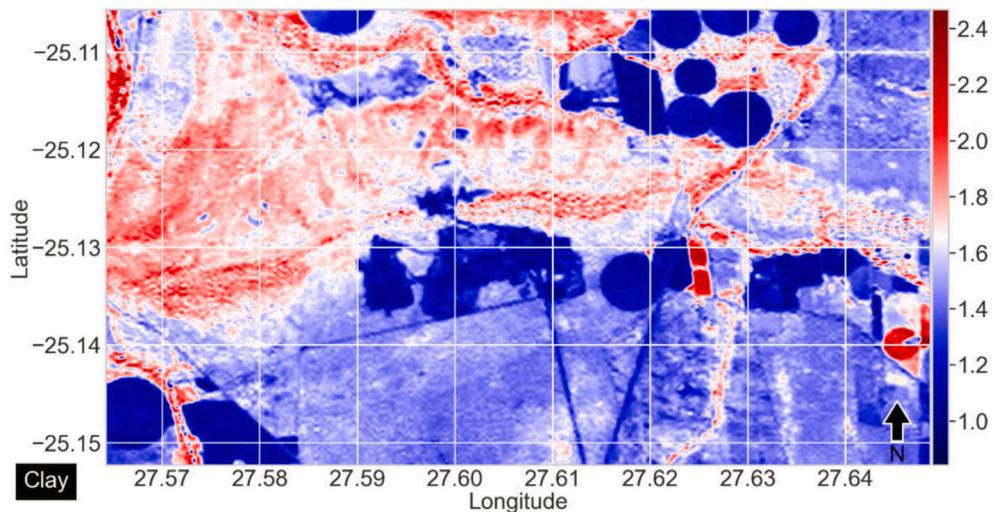


Fig. 8. Image of the clay content (warm colours) over the study area. Image was produced using band ratios. The spectral response of the area is colour-mapped to the ratio.

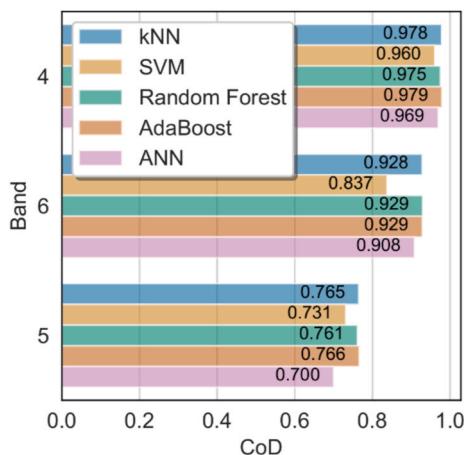


Fig. 9. Results of the machine learning algorithm selection as measured through the CoD (coefficient of determination) metric for all tested bands. Here kNN is k-nearest neighbours, SVM is support vector machine, AdaBoost is adaptive boosting and ANN is artificial neural networks.

Table 3
Best prediction results from the optimum grid.

BAND	METHOD	COD	PARAMETERS
BAND4	AdaBoost	0.978775	maximum depth = None, maximum number of features = 1, minimum number of samples for a leaf = 3, minimum number of samples for a split = 4, Ensemble size = 500
BAND6	Random Forest	0.929147	maximum depth = None, maximum number of features = 2, minimum number of samples for a leaf = 4, minimum number of samples for a split = 3, Ensemble size = 1500
BAND5	AdaBoost	0.766057	maximum depth = None, maximum number of features = 1, minimum number of samples for a leaf = 3, minimum number of samples for a split = 4, Ensemble size = 200

5.1. Detection of mineral deposit occurrence anomaly using machine learning

For the Assen iron ore deposit and the remote sensing data that we

employed, the a-priori knowledge of the bands that capture ferrous and ferric iron oxide reflections was key in our ability to subset the entire dataset into features and targets. Without this knowledge, it may be more difficult to or time-consuming to systematically subset all combinations of bands and identify relevant anomalies. The anthropogenic structures that are prominent in the raw band maps (Figs. 6–8) are well suppressed in the prediction residual maps (Figs. 11–13). However, this is because of two factors: (1) that such structures occur with sufficient frequency, such that the predictive models were able to incorporate their signatures into the background; and (2) that such structures contain broad band information that was also captured in some of the bands that were used as features. For (1), with increasing occurrence of anthropogenic structures, their relationships between bands were increasingly incorporated into the baseline and therefore, their signatures became better constructed by the machine learning models. For (2), as the relationships between the features and targets became stronger for any structure, they increasingly became part of the background. Rare objects with signatures that are distinctly different from the background cannot be adequately suppressed and therefore become detected anomalies (e.g., the anomaly in Fig. 14). An implication of this is that the sizing of the data's spatial coverage should be made large enough such that the occurrence of uninteresting phenomenon is frequent enough so that they cannot be regarded as anomalous. It is not possible to directly control factor (2) using existing remote sensing data. In our unsupervised approach to anomaly detection of mineral deposits, the definition of an anomaly can be summarised as: spectral signatures that are rare but relevant within the context of a-priori knowledge, within a region of interest. In this sense, for similar applications where remote sensing is used to detect mineral deposit occurrence that contains any signature that is differential across the bands (e.g., known litho-cap, dispersal train or surficial alteration compositions), we expect that our method is able to enhance and detect anomalies, and thus contribute to the detection of mineral deposits. As our technique is not specific to anomalies of a mineral deposit type, we further expect that our method may be used to enhance or detect other forms of surficial anomalies. However, this remains to be positively demonstrated on a per-purpose and per-application basis.

5.2. Anomaly signal enhancement using PCA

The application of a-priori knowledge of the bands of interest provides a means to delineate subsets of data as targets. Combining anomalies across these bands using PCA seems to be an effective method

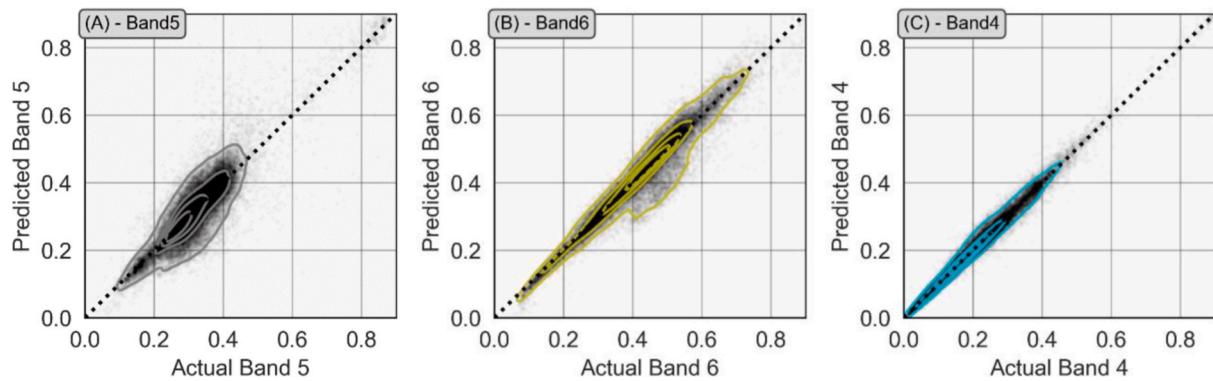


Fig. 10. Prediction versus actual for bands, in order of increasing CoD for band 5 ($\text{CoD} = 0.766$), band 6 ($\text{CoD} = 0.929$) and band 4 ($\text{CoD} = 0.979$) using algorithms and parameters in Table 3. Notice that band 4 is reconstructed the best, which implies that it contains the least amount of anomalies.

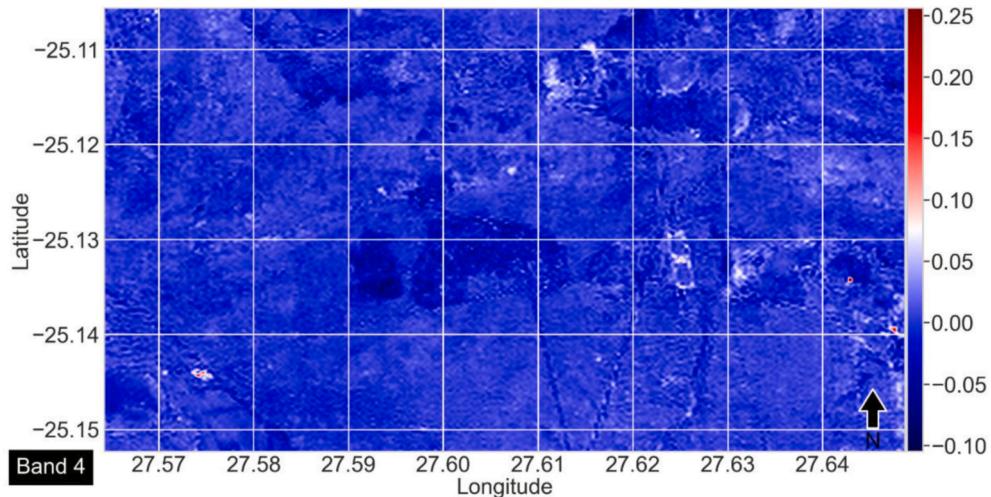


Fig. 11. Prediction residual map using the AdaBoost (adaptive boosting) algorithm. The prediction residuals are colour-mapped in this figure. A faint central object in dark blue is roughly the footprint of the Assen iron ore deposit.

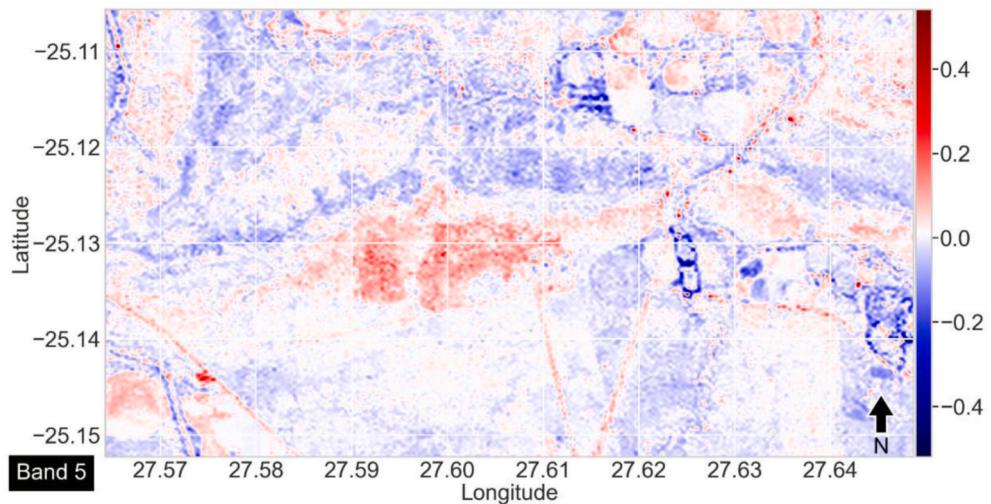


Fig. 12. Prediction residual image using the AdaBoost (adaptive boosting) algorithm. The prediction residuals are colour-mapped in this figure. The central object in red is roughly the footprint of the Assen iron ore deposit.

to sequentially extract the largest to the smallest anomalies in the area, with the largest anomalies being captured in the first principal component. In this sense, PCA replaces highly specific mathematical models of

band amplitudes, such as band ratios. In our application, the PC1 map clearly captured and delineated the Assen iron ore deposit (Fig. 14), which is a feat unmatched by either the ferrous or ferric iron oxide maps

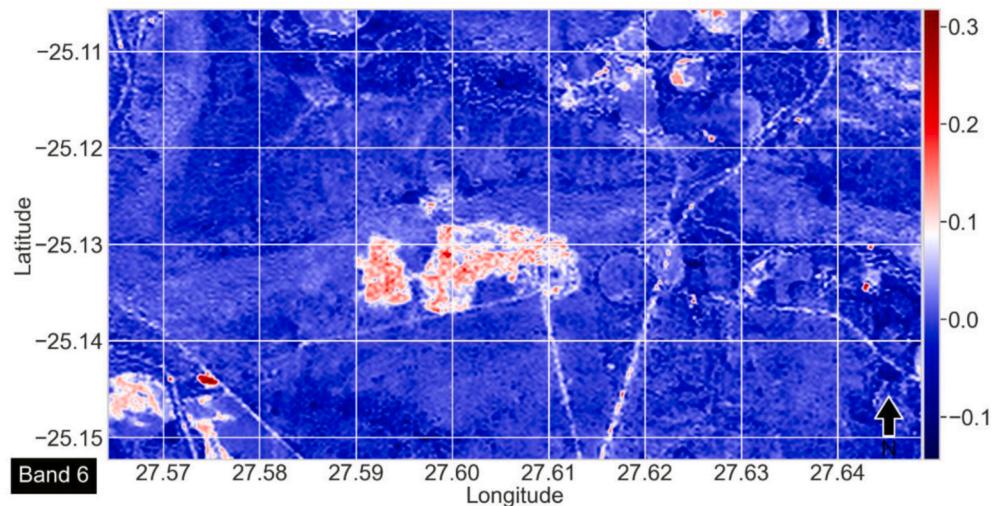


Fig. 13. Prediction residuals image using the random forest algorithm. The prediction residuals are colour-mapped in this figure. The central object in red is roughly the footprint of the Assen iron ore deposit.

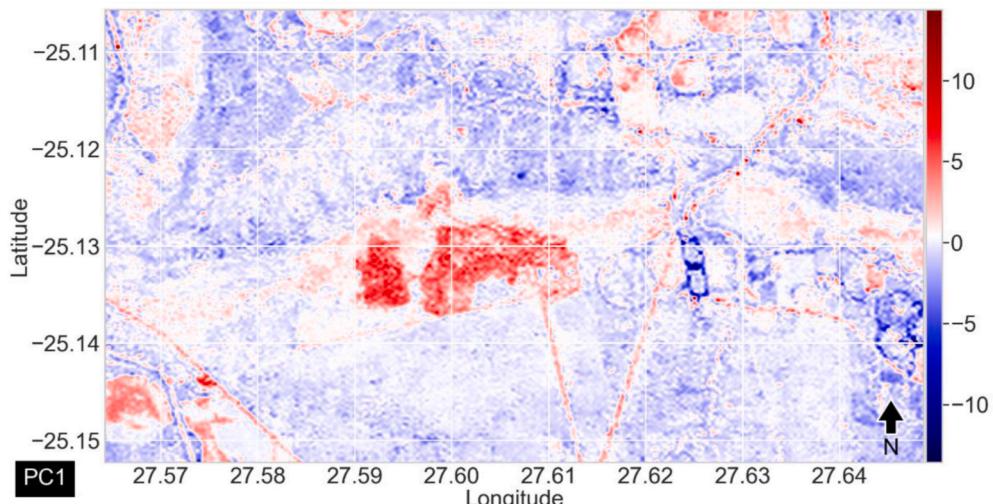


Fig. 14. Principal component 1 (PC1) map of anomalies within the study area. The prediction residuals are colour-mapped in this figure. The bright red anomaly in the centre of the map mainly corresponds with the excavated footprint of the Assen iron deposit.

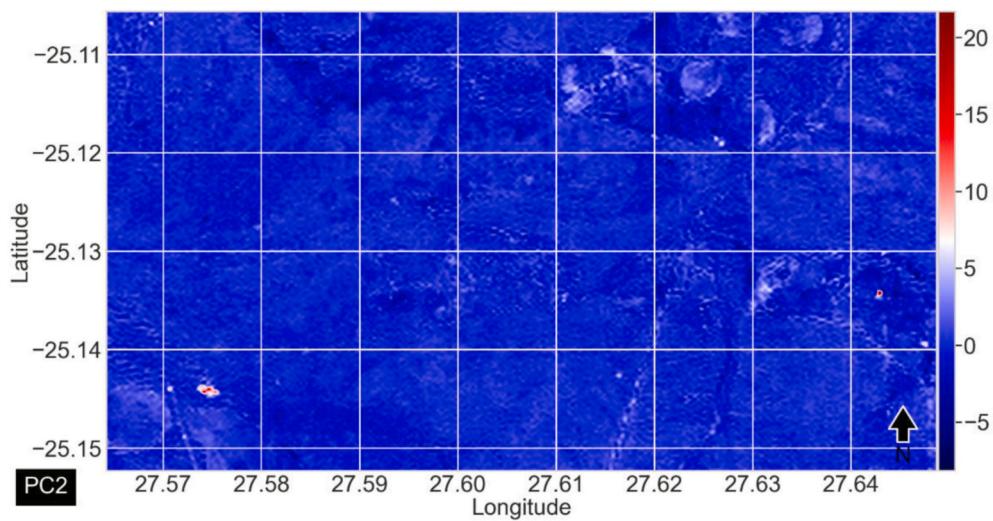


Fig. 15. Principal component 2 (PC2) map of anomalies within the study area. The most prominent anomaly is at the bottom left of the map.

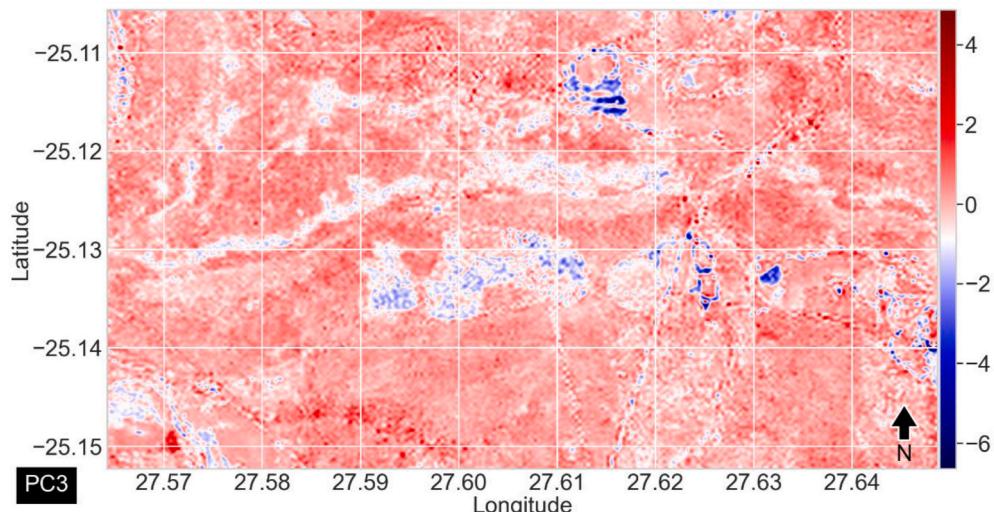


Fig. 16. Principal component 3 (PC3) map of anomalies in the study area. Aside from some footprint of the Assen iron ore mine (centre of map), no other significant anomalies are visible.

(Figs. 6 and 7), both of which were much less selective and contained unsuppressed reflections of anthropogenic structures. A key limitation of PCA is that it can only create linear combinations of anomalies, which is not always the case with remote sensing band composites, such as band ratios. Aside from PCA, there are other techniques for summarising anomalies in a nonlinear manner, the most obvious of which are kernel PCA and autoencoders, and the former uses a nonlinear kernel to create nonlinear combinations of the input. However, the use of kernel PCA or other nonlinear techniques in an unsupervised manner requires substantial additional investigation, as the degrees of freedom generally increases with increasing nonlinearity of models. This implies that the choice to recombine prediction residual maps into multi-band combinations with the use of nonlinear techniques may be more subjective, and therefore, less data-driven. Autoencoders are a better choice in this regard with a characteristic caveat of such methods being their outputs are not easily explainable.

5.3. Implications for exploration of near-surface ore deposits

Classification and clustering machine learning methods are commonly used in data-driven approaches for mapping exploration targets (Bolouki et al., 2020; Gupta and Venkatesan, 2020). The application of well-established techniques such as fractals, PCA, and geostatistics for the construction of mineral prospectivity maps has gained a lot of attention and success in various geological terranes (Gabr et al., 2010; Taravat et al., 2015; Shirmard et al., 2020). Remote sensing has been applied in several exploration geology campaigns, but much of its application to mineral deposit occurrence mapping has always been based on band ratios or expert knowledge combination of spectral bands (Van der Meer et al., 2012; Beiranvand Pour et al., 2019). One challenge with knowledge-driven approaches is that multiple indicators of mineral deposit occurrence are derived, resulting in multiple exploration targets of which some may be false positives. Yet another challenge with knowledge-driven approaches is that the models must be human comprehensible and, regardless of operator expertise, some subjectivity in the translation of knowledge into useable models is unavoidable. Both challenges are relevant for a range of disciplines but, for geosciences, the latter is particularly so, as geosciences tend to be more abductive rather than inductive and deductive (Kasens, 2013; Kleinhans, 2021). Therefore, given the variety of models (e.g., formation models of mineral deposits) available and the operator's bias and exposure to them, it is not necessary that indicators and band composites are consistently produced in an identical manner, even for the same area and deposit types.

We chose an unsupervised but knowledge-guided data-driven approach over the more established supervised classification approaches for mineral deposit detection because the unsupervised approach is more objective, does not require training data (that contain labelled mineral deposits or spectral signatures) and is more general (e.g., detects an unlimited range of types of anomalies). Therefore, we avoided the issue of crafting consistent and objective indicators, band composites, as well as the issue of class imbalance in training data (e.g., there is always more area that is devoid of mineral deposits than the opposite). In addition, our approach is formulated in a manner that specifically favours scientific uses by adopting both discipline-specific knowledge and shallow learning (for non-scientific uses, deep learning may be simpler to implement and higher performing, although additional investigations are needed). The Assen iron ore deposit area created an opportunity to test the proposed approach, which we believe could be successful for the targeting of other deposits that are located near the surface or exhibits a surficial signature. In conventional approaches, an ore deposit, such as the Assen deposit, which exhibits mixed characteristics of highly weathered and pristine ore that is partially covered by a thin overburden may be difficult to constrain using multispectral data. However, we did not encounter such a difficulty using our approach. Conventional spectral data processing techniques and those based on supervised classification require a number of existing deposits to train algorithms and the process may be computationally intensive. Because our proposed approach is unsupervised, it does not require training with a set of a prescribed minimum number of existing mineral deposits as it only searches for anomalous signals in target spectral bands and validation is done using field ground-truthing or through expert knowledge. In parallel with recent developments in self-organised maps for interpreting spectral bands (Kristollari and Karathanassi, 2020), our new approach has also demonstrated the value of an unsupervised, highly explainable, and data-driven approach for enhancing and detecting anomalous spectral signals.

6. Conclusion

There is a vast amount of remote sensing data available, in that they either are or approach big data. For reconnaissance mineral exploration, the key research question is how to maximize the value of data and the value of insights that are derived from data. We attempted to address this question in a manner that we thought is the most efficacious, which is by adopting data-driven techniques to perform anomaly detection, guided by discipline-specific knowledge. This is one significant

contribution of this study. On the one hand, the data-driven approach makes use of the abundance and quality of remote sensing data. On the other hand, the discipline-specific knowledge ensures that the resulting extracted anomalies are the most likely to be relevant to a specific exploration purpose (e.g., a deposit type being sought). These vastly reduce the solution space (and thereby maximize the value of insights extracted) from completely unconstrained data-driven anomalies to a subset that most plausibly contain the results being sought, and thus also facilitate downstream use of the results. From the algorithmic perspective, although general purpose anomaly detectors exist (e.g., autoencoders), they seem to be entirely of the deep learning variety. Our data-driven approach focuses on the use of multiple shallow learning algorithms and demonstrates that it is possible to balance model explainability and performance, which is important for scientific purposes (alongside the use of existing scientific knowledge, which also facilitates result interpretability). Another contribution of this study is the recognition that remote sensing data are highly suitable for anomaly detection in the variable domain, which to the best of our knowledge, has only consistently been performed using manual data analysis (e.g., band ratio analysis) or in the spatial domain using spatial/digital image processing techniques. This may be very important for mineral deposits, because their spatial footprint is likely to be highly variable and generally incomparable, depending on the exposure, geological context, and other alterations (natural or man-made). In comparison, similar mineral deposits consistently demonstrate a comparable assemblage of minerals across space and time. This is also a form of discipline-specific knowledge that effectively guides the adoption of artificial intelligence in geosciences. In this study, we leveraged this fact, along with the knowledge of the bands of interest for a particular type of deposit (here, BIF) to create an artificial intelligence-based anomaly detection method.

To demonstrate a proof-of-concept of our method, we used the Assen iron ore deposit in the Transvaal Supergroup (South Africa) as a target, as captured in Landsat-8 OLI satellite imagery. The approach presented in this study relies on machine learning algorithms to model the relationships between target spectral bands that are known to have high affinity for iron oxides and the feature bands. This created machine learning-based models of the background in variable domain, through which the reconstruction error in the target bands can be calculated as a measure of the anomalousness of each data point. For the detection of possible iron deposit occurrence, prediction residuals were utilised to build spectral anomaly maps that feature enhanced pixel content for iron deposit occurrence. Without manually crafting band composites that proxy iron deposit occurrence, our proposed approach was able to detect existing and probable mineral deposit occurrence hotspots. The proposed mineral deposit occurrence anomaly detection approach is the study's most significant contribution, as it uses a general machine learning workflow. Its applicability is broad and not limited to remote sensing data or for the targeting of mineral deposits; it can also be used with geochemical data or a mixture of datasets. To summarise the anomalies in the study area, we demonstrated the effectiveness of PCA; however, this is a post-processing step to aid targeting and interpretation and is not a key step in our proposed method.

There are numerous directions for future research. Within the applied science and exploration community, we reasonably expect that our approach can be applied to iron ore deposits in various geological terranes and can be potentially extended to other types of metal deposits (e.g., manganese and Fe-Ti-V magnetite ores) in sedimentary and mafic-ultramafic complexes. Within the methodology community, it is possible to further incorporate spatial anomaly detection as a post-processing step to further delineate anomalies of a particular spatial footprint to further reduce the solution space to aid downstream uses. Where our approach is not required to produce explainable results, it is possible to also consider other deep learning methods for anomaly detection (e.g., autoencoders).

Funding

This study was supported by a Department of Science and Innovation (DSI)-National Research Foundation (NRF) Thuthuka Grant (Grant UID: 121973) and DSI-NRF CIMERA.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank Mr. Elekanyani Negwangwatini for providing geological details about the Assen iron ore deposit. We would also like to thank an anonymous reviewer for constructive comments which have greatly improved the quality of this article. Prof. Hua Wang is also thanked for editorial handling.

References

- Agar, B., Coulter, D., 2007. Remote sensing for mineral exploration – a decade perspective 1997–2007. In: Milkereit, B. (Ed.), Fifth Decennial International Conference on Mineral Exploration, pp. 109–136.
- Al-doski, J., Mansor, S.B., Zulhaidi, H., Shafri, M., 2013. Image classification in remote sensing. *Environ. Earth Sci.* 3 (10), 141–148.
- Ali, I., Greifeneder, F., Stamenkovic, J., Neumann, M., Notarnicola, C., 2015. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Rem. Sens.* 7, 16398–16421. <https://doi.org/10.3390/rs71215841>.
- Amer, R., Kusky, T., Ghulam, A., 2010. Lithological mapping in the central eastern desert of Egypt using ASTER data. *J. Afr. Earth Sci.* 56, 75–82. <https://doi.org/10.1016/j.jafrearsci.2009.06.004>.
- An, J., Cho, S., 2015. Variational autoencoder based anomaly detection using reconstruction probability. SNU Data Centre – Special Lecture on IE 2015 (2) on the 25 April 2022 from. <http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf>.
- Asokan, A., Anitha, J., Ciobanu, M., Gabor, A., Naaji, A., Hemanth, D.J., 2020. Image processing techniques for analysis of satellite images for historical maps classification—an overview. *Appl. Sci.* 10, 4207. <https://doi.org/10.3390/app10124207>.
- Bachri, I., Hakodaoui, M., Raji, M., Teodoro, A.C., Benbouziane, A., 2019. Machine learning algorithms for automatic lithological mapping using remote sensing data: a case study from Souk Arbaa Sahel, Sidi Ifni Inlier, Western Anti-Atlas, Morocco. *ISPRS Int. J. Geo-Inf.* 8 (6), 248. <https://doi.org/10.3390/ijgi8060248>.
- Beiranvand Pour, A., Hashim, M., 2014. ASTER, ALI and Hyperion sensors data for lithological mapping and ore minerals exploration. *SpringerPlus* 3, 130. <https://doi.org/10.1186/2193-1801-3-130>.
- Beiranvand Pour, A., Hashim, M., Hong, J.K., Park, Y., 2019. Lithological and alteration mineral mapping in poorly exposed lithologies using Landsat-8 and ASTER satellite data: north-eastern Graham Land, Antarctic Peninsula. *Ore Geol. Rev.* 108, 112–133. <https://doi.org/10.1016/j.oregeorev.2017.07.018>.
- Beiranvand Pour, A., Hashim, M., Makoundi, C., Zaw, K., 2016. Structural mapping of the Bentong-Raub Suture Zone using PALSAR remote sensing data, Peninsular Malaysia: implications for sediment-hosted/orogenic gold mineral systems exploration. *Resour. Geol.* 66 (4), 368–385. <https://doi.org/10.1111/rge.12105>.
- Barnett, V., Lewis, L., 1978. *Outliers in Statistical Data*, third ed. John Wiley & Sons Ltd, p. 608.
- Billor, N., Hadi, A.S., Velleman, P.F., 2000. BACON: blocked adaptive computationally efficient outlier nominators. *Comput. Stat. Data Anal.* 34 (3), 279–298. [https://doi.org/10.1016/S0167-9473\(99\)00101-2](https://doi.org/10.1016/S0167-9473(99)00101-2).
- Bolouri, S.M., Ramazi, H.R., Maghsoudi, A., Beiranvand Pour, A., Sohrabi, G., 2020. A remote sensing-based application of Bayesian networks for epithermal gold potential mapping in Ahar-Arasbaran area, NW Iran. *Rem. Sens.* 12 (1), 105. <https://doi.org/10.3390/rs12010105>.
- Carranza, E.J.M., Hale, M., 2002. Mineral mapping with Landsat Thematic Mapper data for hydrothermal alteration mapping in heavily vegetated terrain. *Int. J. Rem. Sens.* 23 (22), 4827–4852. <https://doi.org/10.1080/01431160110115014>.
- Chakouri, M., El Harti, A., Lhissou, R., El Hachimi, J., Jelloul, A., 2020. Geological and mineralogical mapping in moroccan central jebilet using multispectral and hyperspectral satellite data and machine learning. *Int. J. Adv. Trends Comput. Sci. Eng.* 9, 5772–5783. <https://doi.org/10.30534/ijatcse/2020/234942020>.
- Chandola, V., Banerjee, A., Kumar, V., 2009. Anomaly detection: a survey. *ACM Comput. Surv.* 41 (3), 1–58. <https://doi.org/10.1145/1541880.1541882>.
- Ciampalini, A., Garfagnoli, F., Del Ventisette, C., Moretti, S., 2013. Potential use of remote sensing techniques for exploration of iron deposits in Western Sahara and southwest of Algeria. *Nat. Resour. Res.* 22, 179–190. <https://doi.org/10.1007/s11053-013-9209-5>.
- Cooley, T., Anderson, G.P., Felde, G.W., Hoke, M.L., Ratkowski, A.J., Chetwynd, J.H., Gardner, J.A., Adler-Golden, S.M., Matthew, M.W., Berk, A., Bernstein, L.S.,

- Acharya, P.K., Miller, D., Lewis, P., 2002. FLAASH, a MODTRAN4-based atmospheric correction algorithm, its application and validation. *Proc. Geosci. Rem. Sens. Symp.* 3, 1414–1418. <https://doi.org/10.1109/IGARSS.2002.1026134>. IEEE International.
- 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. <https://doi.org/10.1016/j.cageo.2013.10.008>.
- Dai, J., Wang, D., Dai, H., Liu, L., Wu, Y., 2017. Geological mapping and ore prospecting study using remote sensing technology in Jiajika area of western Sichuan Province. *Chin. Geol.* 44 (2), 389–398.
- Diaz-Rodriguez, J., Dietmar Muller, R., Chandra, R., 2021. Predicting the emplacement of cordilleran porphyry copper systems using a spatio-temporal machine learning model. *Ore Geol. Rev.* 137, 104300 <https://doi.org/10.1016/j.oregeorev.2021.104300>.
- ENVI, 2009. Atmospheric correction module: QUAC and FLAASH user's guide. http://www.exelisvis.com/portals/0/pdfs/envi/Flaash_Module.pdf (2009). (Accessed 31 March 2022).
- Eriksson, P.G., Hattingh, P.J., Altermann, W., 1995. An overview of the geology of the transvaal sequence and Bushveld complex, South Africa. *Miner. Deposita* 30 (2), 98–111. <https://doi.org/10.1007/BF00189339>.
- Gabr, S., Ghulam, A., Kusky, T., 2010. Detecting areas of high-potential gold mineralization using ASTER data. *Ore Geol. Rev.* 38, 59–69. <https://doi.org/10.1016/j.oregeorev.2010.05.007>, 1–2.
- Gad, S., Kusky, T., 2006. Lithological mapping in the Eastern Desert of Egypt, the Barramiya area, using Landsat thematic mapper (TM). *J. Afr. Earth Sci.* 44, 196–202.
- Gao, L., Guo, Q., Plaza, A., Li, J., Zhang, B., 2014. Probabilistic anomaly detector for remotely sensed hyperspectral data. *J. Appl. Remote Sens.* 8 (1), 083538 <https://doi.org/10.1117/1.JRS.8.083538>.
- Gerya, T.V., Uken, R., Reinhardt, J., Watkeys, M.K., Maresch, W.V., Clarke, B.M., 2003. Cold fingers in hot magma: numerical modeling of country-rock diapirs in the Bushveld Complex, South Africa. *Geol.* 31, 753–756. <https://doi.org/10.1130/G19566.1>.
- Goetz, A.F.H., Rowan, L.C., 1981. Geologic remote sensing. *Science* 211, 781–791. <https://doi.org/10.1126/science.211.4484.781>.
- Gupta, P., Venkatesan, M., 2020. Mineral identification using unsupervised classification from hyperspectral data. In: Venkata Krishna, P., Obaidat, M. (Eds.), Emerging Research in Data Engineering Systems and Computer Communications, Advances in Intelligent Systems and Computing, vol. 1054. Springer. https://doi.org/10.1007/978-981-15-0135-7_25.
- Hartzer, F.J., 1987. *Die Geologie van die Krokkodil rivier fragment*. Transvaal. MSc thesis, Rand Afrikaans University, Johannesburg, South Africa.
- Harvey, A., Fotopoulos, G., 2016. Geological mapping using machine learning algorithms. *Int. Arch. Photogram. Rem. Sens. Spatial Inf. Sci.* 41-B8, 423–430. <https://doi.org/10.5194/isprsarchives-XLI-B8-423-2016>.
- Hawkins, D.M., 1980. *Identification of Outliers*. Chapman and Hall London, New York, p. 183.
- Irons, J.R., Dwyer, J.L., Barsi, J.A., 2012. The next Landsat satellite: the Landsat data continuity mission. *Remote Sens. Environ.* 122, 11–21. <https://doi.org/10.1016/j.rse.2011.08.026>.
- Kastens, K.A., 2013. Earth science, philosophy of. In: Gunstone, R. (Ed.), Encyclopedia of Science Education. Springer. https://doi.org/10.1007/978-94-007-6165-0_243-1.
- Kleinhanz, M.G., 2021. Down to Earth: history and philosophy of geoscience in practice for undergraduate education. *Euro. J. Phil. Sci.* 11, 81. <https://doi.org/10.1007/s13194-021-00402-4>.
- Kristollari, V., Karathanassi, V., 2020. Fine-tuning self-organizing maps for Sentinel-2 Imagery: separating clouds from bright surfaces. *Rem. Sens.* 12 (12), 1923. <https://doi.org/10.3390/rs12121923>.
- Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. *Geosci. Front.* 7 (1), 3–10. <https://doi.org/10.1016/j.gsf.2015.07.003>.
- Madhuwanand, L., Sadavarte, P., Visschedijk, A.J.H., Denier Van Der Gon, H.A.C., Aben, I., Osei, F.B., 2021. Deep convolutional neural networks for surface coal mines determination from sentinel-2 images. *Eur. J. Remote Sens.* 54 (1), 296–309. <https://doi.org/10.1080/22797254.2021.1920341>.
- Manakos, I., Manevski, K., Kalaitzidis, C., Edler, D., 2011. Comparison between atmospheric correction modules on the basis of Worldview-2 imagery and in situ spectroradiometric measurements. In: 7th EARSeL SIG Imaging Spectroscopy Workshop, pp. 1–14. Edinburgh.
- Mather, P.M., Tso, B., 2009. *Classification Methods for Remotely Sensed Data*, second ed. Taylor & Francis.
- Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. *Int. J. Rem. Sens.* 39 (9), 2784–2817. <https://doi.org/10.1080/01431161.2018.1433343>.
- Morales-Forero, A., Bassetto, S., 2019. Case study: a semi-supervised methodology for anomaly detection and diagnosis. In: 2019 IEEE International Conference on Industrial Engineering and Engineering Management, pp. 1031–1037.
- Noori, L., Beiranvand Pour, A., Askari, G., Taghipour, N., Pradhan, B., Lee, C.W., Honarmand, M., 2019. Comparison of different algorithms to map hydrothermal alteration zones using ASTER remote sensing data for polymetallic vein-type ore exploration: toroud-Chahshirin magmatic belt (TCMB), North Iran. *Rem. Sens.* 11, 495. <https://doi.org/10.3390/rs11050495>.
- Pohl, C., Van Genderen, J., 2014. Remote sensing image fusion: an update in the context of Digital Earth. *Int. J. Digit. Earth* 7 (2), 158–172. <https://doi.org/10.1080/17538947.2013.869266>.
- Pour, A.B., Hashim, M., Park, Y., Hong, J.K., 2019. Lithological and alteration mineral mapping in poorly exposed lithologies using Landsat-8 and ASTER satellite data: north-eastern Graham Land, Antarctic Peninsula. *Ore Geol. Rev.* 108, 112–133. <https://doi.org/10.1016/j.oregeorev.2017.07.018>.
- Rajendran, S., Nasir, S., 2014. Hydrothermal altered serpentised zone and a study of Ni magnesioferrite–magnetite–awaruite occurrences in Wadi Hibi, Northern Oman. Mountain discrimination through ASTER mapping. *Ore Geol. Rev.* 62, 211–226. <https://doi.org/10.1016/j.oregeorev.2014.03.016>.
- Rajesh, H., 2004. Application of remote sensing and GIS in mineral resource mapping—An overview. *J. Mineral. Petrol. Sci.* 99, 83–103. <https://doi.org/10.2465/jmps.99.83>.
- Reed, I.S., Yu, X., 1990. Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. *IEEE Trans. Acoust. Speech Signal Process.* 38, 1760–1770. <https://doi.org/10.1109/29.60107>.
- Rockwell, B.W., 2013. *Automated Mapping of Mineral Groups and Green Vegetation from Landsat Thematic Mapper Imagery with an Example from the San Juan Mountains, Colorado*. U.S. Geological Survey Scientific Investigations, p. 25. Map 3252, scale 1:325,000.
- Rowan, L., Wetlaufer, P., Goetz, A., Billingsley, F., Stewart, J., 1974. Discrimination of Rock Types and Detection of Hydrothermally Altered Areas in South-Central Nevada by the Use of Computer-Enhanced ERTS Images, vol. 883. U.S. Geological Survey Professional Paper, p. 35.
- Roy, D.P., Wulder, M.A., Loveland, T.R., Woodcock, C.E., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., et al., 2014. Landsat-8: science and product vision for terrestrial global change research. *Remote Sens. Environ.* 145, 154–172. <https://doi.org/10.1016/j.rse.2014.02.001>.
- Sabins, F., 1997. *Remote Sensing: Principles and Interpretation*, third ed. Waveland Pr. Inc.
- Salem, S.M., El Sharkawi, M., El-Alfy, Z., Soliman, N.M., Ahmed, S.E., 2016. Exploration of gold occurrences in alteration zones at Dungash district, Southeastern Desert of Egypt using ASTER data and geochemical analyses. *J. Afr. Earth Sci.* 117, 389–400.
- Sehgal, S., 2012. Remotely Sensed landsat image classification using neural network approaches. *Comput. Sci.* 2 (5), 43–46.
- Shirmard, H., Farahbakhsh, E., Beiranvand Pour, A., Muslim, A.M., Müller, R.D., Chandra, R., 2020. Integration of selective dimensionality reduction techniques for mineral exploration using ASTER satellite data. *Rem. Sens.* 12, 1261. <https://doi.org/10.3390/rs12081261>.
- Shevyrev, S., Carranza, E.J.M., 2022. Application of maximum entropy for mineral prospectivity mapping in heavily vegetated areas of Greater Kurile Chain with Landsat 8 data. *Ore Geol. Rev.*, 104758 <https://doi.org/10.1016/j.oregeorev.2022.104758>.
- Shirmard, H., Farahbakhsh, E., Müller, R.D., Chandra, R., 2022. A review of machine learning in processing remote sensing data for mineral exploration. *Remote Sens. Environ.* 268, 112750 <https://doi.org/10.1016/j.rse.2021.112750>.
- Song, K., Wang, E., Yao, Y., Fu, J., Hao, D., You, X., 2020. Spectral alteration zonation based on close-range hypersc-320 m imaging spectroscopy: a case study in the Gongchangling high-grade iron ore deposit, Liaoning Province, NE China. *Appl. Sci.* 10, 8369. <https://doi.org/10.3390/app10238369>.
- Song, Y.Y., Ying, L., 2015. Decision tree methods: applications for classification and prediction. *Shanghai Arch. Psychiatry* 27 (2), 130–135. <https://doi.org/10.11919/j.issn.1002-0829.215044>.
- Tan, K., Hou, Z., Ma, D., Chen, Y., Du, Q., 2019. Anomaly detection in hyperspectral imagery based on low-rank representation incorporating a spatial constraint. *Rem. Sens.* 11 (13), 1578. <https://doi.org/10.3390/rs11131578>.
- Taravat, A., Peronaci, S., Sist, M., Del Frate, F., Oppelt, N., 2015. The combination of band ratioing techniques and neural networks algorithms for MSG SEVIRI and Landsat ETM+ cloud masking. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Italy, pp. 2315–2318.
- Turkowski, K., Gabriel, S., 1990. Filters for common resampling tasks. In: Glassner, A.S. (Ed.), *Graphics Gems*. Academic Press.
- USGS, 2022. Landsat 8. <https://www.usgs.gov/landsat-missions/landsat-8>. (Accessed 1 April 2022).
- Van der Meer, F.D., van der Werff, H.M.A., van Ruitenbeek, F.J.A., Hecker, C.A., Bakker, W.H., Noomen, M.F., van der Meijde, M., Carranza, E.J.M., de Smeth, J.B., Woldai, T., 2012. Multi- and hyperspectral geologic remote sensing: a review. *Int. J. Appl. Earth Obs. Geoinf.* 14 (1), 112–128. <https://doi.org/10.1016/j.jag.2011.08.002>.
- Vicente, L.E., Filho, C.R.S., 2011. Identification of mineral components in tropical soils using reflectance spectroscopy and advanced spaceborne thermal emission and reflection radiometer (ASTER) data. *Remote Sens. Environ.* 115, 1824–1836. <https://doi.org/10.1016/j.rse.2011.02.023>.
- Wu, H., 2020. Geemap: a Python package for interactive mapping with Google Earth engine. *J. Open Source Softw.* 5 (51), 2305. <https://doi.org/10.21105/joss.02305>.
- Wu, Q., Lane, C.R., Li, X., Zhao, K., Zhou, Y., Clinton, N., DeVries, B., Golden, H.E., Lang, M.W., 2019. Integrating LiDAR data and multi-temporal aerial imagery to map wetland inundation dynamics using Google Earth Engine. *Remote Sens. Environ.* 228, 1–13. <https://doi.org/10.1016/j.rse.2019.04.015>.
- Yao, X., Zhao, C., 2018. Hyperspectral anomaly detection based on the bilateral filter. *Infrared Phys. Technol.* 92, 144–153. <https://doi.org/10.1016/j.infrared.2018.05.028>.
- Zhang, S.E., Nwaila, G.T., Bourdeau, J.E., Ashwal, L.D., 2021. Machine learning-based prediction of trace element concentrations using data from the Karoo large igneous

- province and its application in prospectivity mapping. *Artif. Intell. Geosci.* 2, 60–75. <https://doi.org/10.1016/j.aiig.2021.11.002>.
- Zhang, S.E., Bourdeau, J.E., Nwaila, G.T., Corrigan, D., 2022. Towards a fully data-driven prospectivity mapping methodology: a case study of the Southeastern Churchill Province, Québec and Labrador. *Artif. Intell. Geosci.* 2, 128–147. <https://doi.org/10.1016/j.aiig.2022.02.002>.
- Zhou, C., Paffenroth, R.C., 2017. Anomaly detection with robust deep autoencoders. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 665–674.