Natural Resources Research, Vol. 34, No. 1, February 2025 (© 2024)
https://doi.org/10.1007/s11053-024-10423-4

Original Paper

Lateritic Ni–Co Prospectivity Modeling in Eastern Australia Using an Enhanced Generative Adversarial Network and Positive-Unlabeled Bagging

Nathan Wake,[1,2] **Ehsan Farahbakhsh**,[1] **and R. Dietmar Müller**[1]

Received 22 December 2023; accepted 19 October 2024
Published online: 18 November 2024

The surging demand for Ni and Co, driven by the acceleration of clean energy transitions, has sparked interest in the Lachlan Orogen of New South Wales for its potential lateritic Ni–Co resources. Despite recent discoveries, a substantial knowledge gap exists in understanding the full scope of these critical metals in this geological province. This study employed a machine learning-based framework, integrating multidimensional datasets to create prospectivity maps for lateritic Ni–Co deposits within a specific Lachlan Orogen segment. The framework generated a variety of data-driven models incorporating geological (rock units, metamorphic facies), structural, and geophysical (magnetics, gravity, radiometrics, and remote sensing spectroscopy) data layers. These models ranged from comprehensive models that use all available data layers to fine-tuned models restricted to high-ranking features. Additionally, two hybrid (knowledge-data-driven) models distinguished between hypogene and supergene components of the lateritic Ni–Co mineral systems. The study implemented data augmentation methods and tackled imbalances in training samples using the SMOTE–GAN method, addressing common machine learning challenges with sparse training data. The study overcame difficulties in defining negative training samples by translating geological and geophysical data into training proxy layers and employing a positive and unlabeled bagging technique. The prospectivity maps revealed a robust spatial correlation between high probabilities and known mineral occurrences, projecting extensions from these sites and identifying potential greenfield areas for future exploration in the Lachlan Orogen. The high-accuracy models developed in this study utilizing the Random Forest classifier enhanced the understanding of mineralization processes and exploration potential in this promising region.

KEY WORDS: Mineral exploration, prospectivity mapping, machine learning, random forest, generative adversarial network, lateritic Ni–Co.

[1]EarthByte Group, School of Geosciences, The University of Sydney, Sydney, Australia.
[2]To whom correspondence should be addressed; e-mail: nathan.wake07@gmail.com

INTRODUCTION

Critical metals[1] such as Li, Ni, Co, and REEs are naturally occurring components of the Earth's crust (Mudd & Jowitt, 2022). There has been a surge in demand for these critical metals over the past few decades driven by various factors, including (NSW Govt. Pamphlet, 2021): (i) innovations and emerging technologies in telecommunications supported by a rapidly growing consumer class worldwide; (ii) rapidly growing clean energy technologies driven by government climate change policies toward net-zero trends in carbon dioxide emissions; and (iii) the disruption of global supply chains due to geopolitical conflicts and pandemics. The critical metals featured in this study are Ni and Co. Nickel is fundamentally important for austenitic stainless steel, specialty alloys, electroplating, critical parts of jet engines, combustion turbines of electric power generation stations, and rechargeable battery electrodes in electronic devices and electric vehicles (Mudd & Jowitt, 2022). Cobalt is a critical and strategic metal in diverse commercial, industrial, and military applications. However, the leading use of Co is in rechargeable battery electrodes in electronic devices and electric vehicles (Mudd & Jowitt, 2022).

The known Ni resources of Australia are mainly found in Western Australia within komatiite-hosted Ni–Cu–PGE sulfide deposits, lateritic Ni–Co deposits, and to a lesser extent, within hydrothermal Ni-polymetallic sulfide deposits. The New South Wales (NSW) Government has developed a strategy to promote the growth of critical minerals and high-tech metal-related industries. Their vision is to develop a critical minerals industry hub, promote exploration and the development of critical minerals supply chains, and attract investment for downstream processing of critical minerals in the NSW central-west region (NSW Govt Pamphlet, 2021). Significant lateritic Ni–Co and Sc resources have been recognized in the NSW central-west region. Opportunities for additional discoveries and the development of these Ni–Co resources arise from the recognition of previously untested lateritic Ni–Co targets, the changing economics and technologies of Ni–Co mineral processing, the occurrence of prospective rocks located beneath shallow cover units, and the lack of previous exploration efforts focused on Ni and Co in NSW (Barnes & Downes, undated).

Despite this recognition and early exploration successes, there is an apparent knowledge gap in presenting a large-scale prospectivity analysis on the potential for additional lateritic Ni–Co resources in NSW. This study addressed this knowledge gap by applying a machine learning (ML)-based framework on publicly available geological and geophysical datasets. It aimed to produce prospectivity maps of lateritic Ni–Co within the Lachlan Orogen using quantitative models through the Random Forest classifier. Common ML challenges, such as training sample imbalances and the difficulty of defining negative training samples, were addressed using innovative techniques like synthetic minority over sampling technique–generative adversarial network (SMOTE–GAN) and positive-unlabeled learning, respectively.

Multiple spatial data processing approaches have been implemented to improve the effectiveness of quantitative mineral exploration during the past decades (Carranza & Laborte, 2015; Kreuzer et al., 2015; Farahbakhsh et al., 2023). The mineral system approach was used to transfer a conceptual mineral system into mappable exploration criteria (Zhou et al., 2023). This process requires evidential maps representing spatial proxies of the associated mineral system processes. These may include geological, geophysical, and geochemical data correlating spatially with known mineral occurrences. The evidential maps and their spatial features were incorporated into a mathematical function/model that assigned weighting/importance according to their spatial relationship to the known mineral deposits. This model was used in the approximation of the final prospectivity maps (Sun et al., 2020). The methods for weighting evidential maps may vary due to the goals of the project, the scale of prospectivity mapping, and the availability of data (Yousefi et al., 2021). Two common approaches include a data-driven approach in which there is an abundance of known mineral occurrences to make spatial associations with geographic information system (GIS) data and a knowledge-driven approach used in the absence of abundant mineral occurrences and

[1] The term 'critical metal' rather than 'critical mineral' is used to emphasize elemental metals that are important and scarce. A 'critical mineral,' in contrast to 'critical metal,' is a broad term that can include any mineral, element, substance, or material that is considered to have a high economic, technological, or strategic value and a high supply risk. It is acknowledged that both terms are sometimes used interchangeably and are not fixed or static but rather dynamic and evolving over time, depending on the changes in supply and demand, technology development, policy intervention, and market conditions.

training data, relying on expert knowledge of the mineral system (Ford et al., 2019). Knowledge-driven models are often applied for mineral targeting in underexplored areas with little information. In areas with more information and known mineral occurrences, data-driven models are favored for targeting (Sun et al., 2020).

Due to the complexity of mineral system processes and the scarcity of known mineral occurrences in underexplored areas, a knowledge-driven weights of evidence approach to mineral prospectivity mapping (MPM) is often implemented (Farahbakhsh et al., 2023). The weights of evidential features are assigned based on expert judgment. This approach may carry exploration bias and uncertainty due to the subjectiveness of understanding the complex mineral system processes (Yousefi et al., 2021). Data-driven approaches combat the subjectiveness by training a ML algorithm to learn the spatial associations with the mineral system proxies and known mineral occurrences. However, these models risk being biased toward the input data and could lead to ''garbage in, garbage out.'' They also lose effectiveness with a lack of training data (Yousefi et al., 2021). ML algorithms were chosen for this study because of their ability to integrate a large amount of spatial data and to identify hidden relationships between known mineral occurrences and exploration features. To address the issues associated with a data-driven approach, we used multiple algorithms and data augmentation methods in a multistep process to create a somewhat ''hybrid model'' with Random Forest as the main classifier.

This study focuses on ML-based mineral prospectivity mapping for lateritic Ni–Co in the East-Central Lachlan Orogen. ML utilizes data-driven algorithms and techniques that automate data clustering, classification, and prediction (Rad et al., 2018). We used supervised ML to train a model on known input and output data to predict outcomes and patterns. It aimed to provide prospectivity maps of lateritic Ni–Co by integrating various exploration data layers using ML algorithms and designing a workflow to process available public exploration datasets in NSW and turned them into features to determine the probability of finding the target mineralization type in potential regions. In addition to the prospectivity maps, the most important features in exploring the target mineralization type and its relationship with the mineral system model were investigated.

The main technical challenge was the relatively low number of known lateritic Ni–Co mineral occurrences (positive training samples) and choosing negative samples in barren regions. This paper presents a ML-based framework for generating prospectivity maps of lateritic Ni–Co in the East-Central Lachlan Orogen to address this problem. There were some limitations associated with available exploration datasets of the Lachlan Orogen, such as the irregularity of crucial basement outcrops, variations in the preservation and thickness of lateritic weathering profiles, and the presence of widespread Cenozoic sedimentary cover and complex regolith. This can impede understanding the geological setting and landform evolution, leading to large-scale mineralization events. However, employing geophysical data such as magnetic, radiometric, and spectral remote sensing along with other data types can aid in imaging the basement source rocks beneath the sedimentary cover and highly prospective and well-preserved lateritic weathering profiles.

GEOLOGICAL SETTING AND LATERITIC NI–CO OCCURRENCES

The Lachlan Orogen is part of a vast accretionary orogenic system that built the eastern one-third of the Australian continent over *c.* 300 myr, called the Tasmanides (Collins et al., 2019). The Lachlan Orogen is a richly endowed mineral province with a long gold, base metal, and tin mining history that commenced in the mid-late 1800s. It hosts a variety of mineral deposit styles, including world-class porphyry-related Au–Cu (e.g., Cadia and Parkes), orogenic Au veins (e.g., Hill End and Adelong), volcanic-associated massive sulfides (e.g., Woodlawn and Captains Flat), metasediment-associated massive sulfides (e.g., Elura and CSA), intrusion-related Sn–W–Mo (e.g., Ardlethan), and epithermal Au–Ag deposits (e.g., Gidginbung and Cowal) (Hough et al., 2007; Huston et al., 2012, 2016). This broad range of mineral systems reflects a complex history of plate convergence with associated orogenesis, extension, and magmatic events during the Paleozoic evolution of the Lachlan Orogen.

The study area is limited to approximately 160,000 km^2 within the Lachlan Orogen of New South Wales. It straddles a major crustal-scale fault system separating the Central and East Lachlan subprovinces. Cambro–Ordovician flysch sedimentary rocks of the Girilambone Group are intruded by Uralian-Alaskan-type mafic–ultramafic intrusive complexes of Late Ordovician–Silurian age, and Alpine-type serpentinized ophiolite belts of Late Cambrian age within the study area.

The mafic–ultramafic intrusions and serpentinized ophiolites are the primary sources of Ni and Co in the lateritic Ni–Co deposits of the Lachlan Orogen. These source rocks were uplifted and exposed during the multiple orogenic events of the Paleozoic. The current landforms of the study area have evolved over the last 300 Ma due to substantial variations in climate and concurrent tectonic events such as gentle uplift, warping, and the continental breakup of the southern Australian margin from Antarctica (Anand, 2005). The cumulative effects of Permian glaciation, Early Cretaceous subsidence and basin sedimentation, and the emergence and erosion of the landmass from the mid-Cretaceous resulted in peneplanation and low-rolling landforms across much of the western and central Lachlan Orogen subprovinces (Anand, 2005). Extended periods of deep chemical weathering under tropical to subtropical conditions during the Mesozoic and Cenozoic have produced supergene enrichment of Ni and Co in locally thick lateritic weathering profiles developed on the exposed mafic–ultramafic intrusions and serpentinized ophiolite source rocks (Elliot & Martin, 1991; Downes, 2021; Keays et al., 2022).

NSW has no historical record of Ni–Co production. Ni–Co-bearing Mn wads that developed in coarse quartz grits from the Tertiary period were reported in the Bungonia district in the early 1900s, and lateritic Ni mineralization was identified in the Colac serpentinite belt in the 1960s (Fitzpatrick, 1968). Lateritic Ni–Co occurrences in other parts of the Lachlan Orogen were recognized and investigated from the 1980s, and more intensive exploration has been conducted from the 2010s until now. There are 33 known lateritic Ni–Co mineral occurrences within the Lachlan Orogen and 28 of these occur within the study area (Fig. 1). Details on these mineral occurrences were downloaded from the NSW MinView3 database and summarized from Company websites (Tables 1 and 2). Only the Nyngan and Homeville mineral occurrences lie outside of this study area. Details on these mineral occurrences were downloaded from the NSW MinView[2] database and summarized from company websites (Tables 1 and 2).

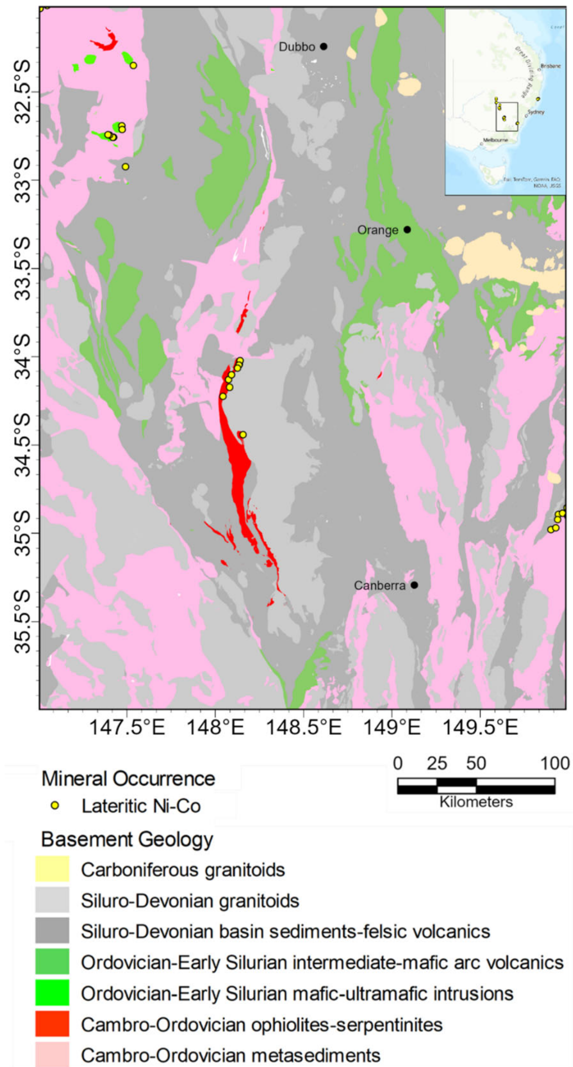


Figure 1. Lateritic Ni–Co occurrences of the Lachlan Orogen overlayed on the basement geology. The inset shows the study area in Eastern Australia.

MATERIALS AND METHODS

Genetic Model

Lateritic weathering profiles developed on mafic–ultramafic intrusions and serpentinites are important sources of supergene-enriched metals such as Ni, Co, Sc, PGM, and Au (Freyssinet et al., 2005). Lateritic Ni–Co refers to a lateritic

[2] MinView database: https://geonetwork.geoscience.nsw.gov.au/geonetwork/srv/eng/catalog.search#/metadata/03d7026374e323584f94f8018664230c84be716c

Table 1. Spatial attributes of lateritic Ni–Co Occurrences within the Lachlan Orogen (MinView NSW Geoscience metadata) (Geocentric Datum of Australia 1994 (GDA94))

Metallogenic district	Project	Name	Longitude (GDA94)	Latitude (GDA94)
Thuddungra	Nico Young	Ardnaree	− 34.222	148.043
		East Arm	− 34.172	148.080
		West Arm	− 34.129	148.074
		Tyagong	− 34.044	148.135
		West Arm (N)	− 34.103	148.092
		Tyagong (S)	− 34.064	148.126
		Tyagong	− 34.048	148.134
		Tyagong (N)	− 34.022	148.140
Syerston	Sunrise	Syerston	− 32.756	147.419
	Flemington	Flemington	− 32.744	147.401
			− 32.743	147.392
	Owendale	Owendale	− 32.692	147.470
	Malamute	Malamute	− 32.350	147.537
Homeville	Homeville	Collerina	− 32.010	147.048
		Gwinnear	− 32.028	147.007
Nyngan	Nyngan	Nyngan (Sc)	− 31.609	146.993
	West Lyn	West Lynn	− 31.498	147.039
		Summervale	− 31.422	147.013
Bungonia	Bungonia	Hillydale	− 34.893	150.000
		Para Grid	− 34.971	149.807
		Strauss	− 34.930	149.881
		Rose Dale	− 34.948	149.895
		Deposit No.5	− 34.983	149.897
		Dog Leg Grid	− 34.978	149.901
		Deposit No.2	− 34.966	149.923
		Deposit No.3	− 34.969	149.927
		Lumley Park	− 34.892	149.943
		Ardmore Park	− 34.899	149.938
		Jacqua Grid	− 34.923	149.938
		Bungonia	− 34.856	149.991
		Ossiers	− 34.905	149.939
		Inverary Jail	− 34.885	149.966
		Mt Inverary	− 34.903	149.995
		Windellama	− 35.014	149.818

weathering profile that contains economically exploitable concentrations of Ni and Co but not to a specific horizon or unit (Gleeson et al., 2003) (Fig. 2). The genetic model of this deposit involves the complex interplay of magmatic, metamorphic, and supergene weathering processes (Golightly, 2010; Marsh et al., 2013). The term ''laterite'' refers to the iron oxide-rich, silica-poor upper soil horizon of intensely weathered regolith. It is not a complete soil profile but specifically the hard iron oxide-rich, silica-poor duricrust that often forms at the top of a laterite weathering profile (Eggleton, 2001; Retallack, 2010). Laterite weathering profiles generally consist of some or all the following horizons (from the base): saprock, saprolite, plasmic zone, mottled zone, ferruginous and/or aluminous duricrust or gravels, and soil (Eggleton, 2001) (Fig. 2).

Lateritic Ni–Co deposits generally form under humid tropical to subtropical savannah conditions and have a multiphase development, evolving as their climatic and/or topographic environment change, which may influence the degree of enrichment of metals in the lateritic weathering profile (Golightly, 2010; Butt & Cluzel, 2013). Lateritic Ni–Co deposits form on ultramafic rocks of any kind and in tectonic environments. These generally contain elevated Ni and Co contents hosted by variably serpentinized forsterite olivine and Mg-rich pyroxenes. The metal contents of protoliths typically range 0.15–0.3% Ni and several hundreds to 1000 ppm Co (Golightly, 2010). Table 3 lists the

Table 2. Geological attributes of lateritic Ni–Co Occurrences within the Lachlan Orogen (MinView NSW Geoscience metadata)

Source rocks	Ore type	Mineral field/deposit name	Mineral resource(s)	Source
Alpine-type serpentinite	Clay-oxide	Thuddungra/Nico Young (1), Thuddungra/Nico Young (2)	53.6 Mt at 0.66% Ni and 0.05% Co 114.3 Mt at 0.56% Ni and 0.06% Co	Jervois (2017) https://jervoisglobal.com
Alaskan-type mafic–ultramafic intrusions	Clay-oxide	Syerston/Sunrise	160 Mt at 0.56% Ni, 0.09% Co and 71 ppm Sc	Sunrise Energy (2020) https://www.sunriseem.com
		Syerston/Flemington	2.7 Mt at 0.24% Ni, 0.1% Ci and 403 ppm Sc	Australian Mines (2022) https://australianmines.com.au
		Syerston/Owendale	35.6 Mt at 0.1% Ni, 0.06% Co, 0.28 g/t Pt and 405 ppm Sc	Platina (2023) https://platinaresources.com.au
		Homeville/Homeville	17.9 Mt at 0.89% Ni and 0.06% Co	Helix (2023) https://helixresources.com.au
		Nyngan/Nyngan	16.9 Mt at 295 ppm Sc	Scandium (2016) https://scandiummining.com
		Nyngan/West Lynn	14.7 Mt at 0.85% Ni and 0.05% Co	Alchemy (2022) https://alchemyresources.com.au
		Nyngan/Summervale	6.6 Mt at 0.82% Ni and 0.04% Co	
Unknown	Manganese wads		None defined	Cazaly (2016) https://cazalyresources.com.au

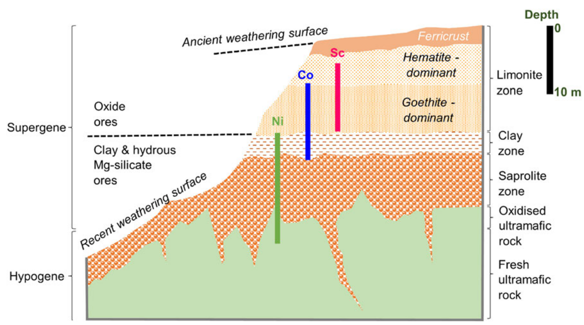


Figure 2. Idealized cross section through a lateritic Ni–Co weathered profile showing all the possible layers and relative enrichment zones of Ni, Co, and Sc. Note that the right-hand side shows a profile intact, and the left-hand side shows a profile partly eroded (modified from Marsh et al., 2013).

main effects of the chemical weathering of rocks in general and how these processes are manifested in the weathering of ultramafic rocks (Elias, 2002). The differing chemistry and mineralogy of the profile are influenced by climatic and geologic factors: climate, topography, drainage, tectonics, structure, and, importantly, parent rock lithology. Ni and Co become enriched to ore grade with stable minerals in the profile through various processes of weathering of ultramafic rocks (Elias, 2002, 2006):

- Leaching of Ni (and Co, Mg, and Mn) during weathering of olivine, pyroxene, and serpentine (leaching from alkalis);
- Absorption of Ni from solution during goethite and smectite formation (stable secondary mineral formation); and
- Absorption of Ni (and Co) from solution during precipitation of Mn oxides (redox).

There are three main lateritic Ni–Co ore types characterized by their mineralogy, and these are based on the dominant minerals hosting Ni and Co (see Table 4) (Brand et al., 1998; Elias, 2006; Butt & Cluzel, 2013):

1. Oxide ores comprising primarily iron hydroxides and oxides in the upper part of the lateritic weathering profile with mean grades 1.0–1.6% Ni;
2. Clay-silicate ores comprising largely smectitic clays (e.g., nontronite) in the upper part of the lateritic weathering profile with 1.0–1.5% Ni;
3. Hydrous Mg-silicate ores comprising hydrated Mg–Ni silicates (serpentine and garnierite) occurring deep in the profile, which may be overlain by oxide and clay-silicate ore types, al-

Table 3. Main processes of chemical weathering and their effects in ultramafic rocks (Butt, 2016; Elias, 2006)

Geological processes	Effects on ultramafic rocks
1. Leaching of mobile constituents: alkalis, alkaline earths	Chemical breakdown of olivine, pyroxene, serpentine and leaching of Mg, Ni, Mn, and Co
2. Formation of stable secondary minerals–iron and aluminum oxides, clays	Goethite and smectite formation, adsorption of Ni from solution
3. Partial leaching of less mobile components–silica, alumina, titanium	Leaching of silica in rainforest and moist savanna climates
4. Mobilization and partial precipitation of redox-controlled constituents: iron and manganese	Precipitation of Mn oxides and adsorption of Ni and Co from solution
5. Retention and residual concentration of resistant minerals: zircon, chromite, quartz	Residual chromite concentration

Table 4. Lateritic Ni–Co ore types and minerals (Butt & Cluzel, 2013)

Laterite type	Mineral	Mineral group	Formula	Metal tenor
Oxide	Goethite	Oxide	α-$(Fe^{3+})O(OH)$	2% Ni, 0.2% Co,
	Absolane	Oxide	$(Ni^{2+},Co^{3+})_xMn^{4+}(O,OH)_4.nH2O$	16% Ni, > 4% Co,
	Lithiophorite	Oxide	$(Al,Li)Mn^{4+}O_2(OH)_2$	1% Ni, 7% Co
Clay silicate	Nontronite	Smectite	$Na_{0.3}Fe_2{}^{3+}(Si,Al)_4O_{10}(OH)_2.nH2O$	~ 4% Ni
	Saponite	Smectite	$Ca/2,Na)_{0.3}(Mg,Fe2+)_3(Si,Al)_4O_{10}(OH)_2.4H_2O$	~ 3% Ni
Hydrous Mg-silicate	Ni-lizardite, Nepouite	Serpentine	$(Mg,Ni)_3Si_2O_5(OH)_4$	6–33% Ni
	7A Garnierite	Serpentine	Variable, poorly defined	15% Ni
	Nimite	Chlorite	$(Ni_5Al)(Si_3Al)O_{10}(OH)_8$	17% Ni
	14A Garnierite	Chlorite	Variable, poorly defined	3% Ni
	Falcondoite	Sepiolite	$(Ni,Mg)_4Si_6O_{15}(OH)_2.6H_2O$	24% Ni
	Kerolite, Willemseite	Talc	$(Ni,Mg)_3Si_4O_{10}(OH)_2$	16–27% Ni
	10A Garnierite	Talc	Variable, poorly defined	20% Ni

Most minerals are poorly crystalline and variable in composition

though usually one ore type predominates in any one deposit.

The dominant ore types in the Lachlan Orogen are oxide and clay-silicate (e.g., Chassé et al., 2017; Downes, 2021).

Conceptual Model

This study applied a conceptual model, or mineral systems approach, linking the processes and features of the lateritic Ni–Co deposit model with available public exploration datasets to generate regional prospectivity maps that will support exploration targeting for new lateritic Ni–Co deposits in the Lachlan Orogen. The mineral systems approach requires understanding the geological processes needed to form and preserve ore deposits at all scales (McCuaig et al., 2010). In this approach, the critical geological processes acting together to form mineral deposits in general are (McCuaig et al., 2010; Porwal & Kreuzer, 2010; Kreuzer et al., 2015):

- Source: Geodynamic setting and associated magmas and fluids required to extract ore components (melts or fluids) from mantle and/or crustal sources
- Transport: Lithospheric structural architecture that provides pathways for fertile magmas and fluids, transferring ore components from source to trap
- Trap/Deposition: Geological processes required to concentrate ore components in the host rocks and/or structures
- Preservation: Geological processes required to preserve the accumulated ore components through time

Although the geological processes of a mineral system may be difficult to map directly, they can be recognized from other mappable features that are indicator elements or proxies that result from those geological processes. The targeting criteria for a mineral system are those features that should be present if the constituent geological processes responsible for the targeted mineralization were active in the area of interest (McCuaig et al., 2010). In this study, the selected targeting criteria manifest mappable features readily available in the exploration data sets provided by the Geological Survey of New South Wales through their MinView website.

Lateritic Ni–Co Minerals System

The lateritic Ni–Co mineral system features are based on detailed models by Marsh et al. (2013) and Downes (2021). The geological processes in this system involve both hypogene (primary) and supergene (secondary) components. The hypogene process forms Ni–Co-bearing mafic–ultramafic rocks through fertile Mg-rich magmas emplaced near the surface. In the Lachlan Orogen, these rocks are Cambro–Ordovician serpentinized ophiolites and Ordo–Silurian Alaskan-type intrusions. These magmas originate in the asthenosphere or lithospheric mantle, interacting with subducting slab melts and mantle wedge peridotites in back-arc and post-collisional extensional settings within accretionary orogens.

The supergene process involves meteoric waters circulating under humid tropical–subtropical conditions, breaking down primary minerals (olivine, pyroxene, serpentine), leaching Ni, Co, and other elements, and redepositing them in stable secondary minerals (Fe and Al oxides, clays, and silicates) within the lateritic weathering profile. Mg-rich magmas rise through the lithosphere at back-arc spreading ridges and along structures within the accretionary orogen. Faults enhance the penetration and flow of meteoric waters, facilitating the development of deep lateritic weathering profiles. The main energy drivers are deep-penetrating oxygenated groundwaters under humid conditions.

Hypogene trap and depositional processes include compositional layering within ophiolite bodies and mafic–ultramafic intrusions, with ultramafic compositions at the base. Ni and Co are partitioned into olivine and pyroxene during magma fractionation and crystallization. These rocks may undergo serpentinization, where olivine and pyroxene alter to serpentine, which retains Ni and Co. Ultramafic rocks have primary Ni content ranging from 0.15% to over 0.3% (Golightly, 2010).

Supergene trap and depositional processes involve chemical weathering and secondary metal enrichment in deep lateritic weathering profiles. Ni and Co are leached from primary minerals and carried downward by groundwater, depositing as secondary oxides, clays, and hydrous Mg silicates. This can result in metal enrichment up to 30 times the original content. The resulting laterite is influenced by climate, topography, tectonics, primary rock type, and structure (Elias, 2002).

Preservation of lateritic Ni–Co deposits requires formation in tectonically stable terrains with low relief, where chemical weathering exceeds erosion. Preservation is enhanced by low relief and protection from erosion by ferruginous or siliceous duricrusts (Butt & Cluzel, 2013). This model applies to deposits in the Lachlan Orogen, where source rock distribution, weathering depth, intensity, and preservation determine the size and continuity of the lateritic Ni–Co deposits.

Application

This study employed a variety of geological, geophysical, remote sensing, and topographic datasets to generate mappable targeting criteria or proxies for the prospectivity analysis of lateritic Ni–Co in the Lachlan Orogen. These datasets were obtained from open-file digital sources provided by the Geological Survey of New South Wales (MinView) and Geoscience Australia's portal.

Geological data layers included the age and distribution of lithological rock units and major regional fault architecture. Mafic–ultramafic source rocks, typically associated with specific geological settings in the Lachlan Orogen, include Alaskan-type zoned mafic–ultramafic intrusions and Alpine-type serpentinized ophiolite bodies (Webb, 2014; Belousova et al., 2022; Keays et al., 2022). Mapped areas of their distribution provide direct targeting criteria with high potential for the occurrence and preservation of lateritic Ni–Co deposits. Faults can be crucial as conduits for the emplacement of Mg-rich magmas or as terrane boundaries controlling the tectonic emplacement of the ophiolite crust. Their presence highlights the potential for deep lateritic weathering profiles and extends the known occurrence of source rocks into less exposed areas.

Various geophysical layers, including airborne magnetics, gravity, and radiometrics, are utilized to target mafic–ultramafic source rocks and develop deep lateritic weathering profiles in the Lachlan Orogen. Magnetic data layers are useful for identifying discrete magnetic high features associated with mafic–ultramafic complexes and serpentinized ophiolites, which contain significant magnetic minerals such as magnetite and chromite. When combined with geological data layers (lithology and structure), magnetics can assist in mapping extensions of mafic–ultramafic source rocks beneath shallow sedimentary and regolith cover areas (Farahbakhsh et al., 2023).

Gravity data layers provide information on density variations in rocks, potentially identifying mafic–ultramafic source rocks in the subsurface. Mafic–ultramafic complexes and serpentinized ophiolite bodies, composed of high specific gravity minerals, are typically denser than surrounding rocks, producing coincident gravity and magnetic highs if sufficiently large and distinct. However, while airborne magnetics and gravity are useful for identifying mafic–ultramafic source rocks, they are not effective for identifying weathering profiles for lateritic Ni–Co deposits.

Radiometric data, acquired through gamma-ray spectrometry, measure the concentrations of potassium (%K), equivalent-uranium (ppm eU), and equivalent-thorium (ppm eTh)[3] in soils and bedrock to a depth of approximately 30 cm. Under weathering conditions, U, Th, and K are redistributed through the weathering profile, with potassium being generally mobile, and U and Th being absorbed by goethite and hematite. High values of U and Th, along with low K values, can help detect surficial Fe-enrichment and deep lateritic weathering profiles developed in mafic–ultramafic source rocks (Barbosa et al., 2013; Iza et al., 2018; Downes, 2021).

Altimetry data, such as digital elevation models (DEM) from the Shuttle Radar Topography Mission, provide elevation and slope data, highlighting areas with low to moderate local relief that may preserve lateritic Ni–Co deposits and information on paleo-water table controls (Brand et al., 1998; Wilford, 2012; Motta & Junior, 2016; Iza et al., 2018; Downes, 2021).

Spectral analysis of multispectral remote sensing datasets and alteration maps can target lateritic weathering profiles on mafic–ultramafic source rocks. Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) satellite data products provide spectroscopy within the visible–near infrared, shortwave infrared, and thermal infrared sensor wavelength regions, useful for measuring specific minerals and mineral groups (Bedell, 2004; Shirmard et al., 2022a). Ferric oxide groups, hydroxyl-bearing alteration minerals, and other silicates display diagnostic signatures and absorption features within these wavelength regions, indicating mineral abundance or composition. However, remote sensing spectra need field spectrometry or other data for verification.

The integration of ASTER-derived alteration maps with DEMs and geological and geophysical data layers provides a robust tool for identifying areas likely to contain lateritic Ni–Co mineralization and targeting specific areas for further exploration. Enhancing geophysical data layers through textural features like correlation, dissimilarity, mean, and standard deviation maps can provide additional geological information and highlight new areas of interest. By combining these datasets and methodologies, this study aims to create accurate and comprehensive prospectivity maps, identifying potential zones for lateritic Ni–Co mineralization in the Lachlan Orogen.

Data Layers and Features

The source rocks for the lateritic Ni–Co mineral system in the Lachlan Orogen are Alaskan-type mafic–ultramafic intrusions and serpentinites formed from post-arc magmatism during the Early Silurian. These ultramafic rocks are identifiable through geological mapping of specific lithological units. The seamless geology dataset from NSW's MinView includes shapefiles of relevant geological ages with attributes such as dominant lithology. These shapefiles serve as polygon training data, enabling direct spatial correlation with known mineral occurrences and generating mineral system proxies/features (as-

[3] The leveled database has been used to produce the first 'Radiometric Map of Australia' – leveled and merged composite potassium (%K), equivalent-uranium (ppm eU), and equivalent-thorium (ppm eTh) grids over Australia at 100 m resolution. The terms equivalent-uranium (eU) and equivalent-thorium (eTh) are estimates of U and Th concentrations from gamma-ray spectrometric measurements based on the assumption of radioactive equilibrium in the U-238 and Th-232 radioactive decay series (Minty et al., 2009).

signed as the dominant lithology) ranked by importance using a Random Forest algorithm.

Regional geochemistry can also be used to highlight target lithological units via stream sediment and rock analysis. However, due to limited data, this dataset primarily validates highlighted prospective areas. The basement source rocks originated from the evolving stages of the subduction system, generating mafic–ultramafic magmas within the upper mantle, which then ascended through the crust due to heat. These processes are effectively highlighted by geophysical data, which provide deeper signatures. Geoscience Australia's geophysical dataset includes useful basement-reaching features such as magnetics and gravity data. These raster files are grid layers that display a spectrum of values for a particular measurement type, useful for highlighting domains of source material based on their spatial association with known mineral occurrences.

Aeromagnetic data are effective for visualizing underlying magnetic rock bodies and improving understanding of geological structures. The total magnetic intensity (TMI) is obtained from aeromagnetic surveys of the Earth's magnetic field and measured in nanotesla (nT). Variations in TMI can be interpreted to approximate geological formations (Isles & Rankin, 2013). The original resolution of the magnetic layers ranges between 80 and 88 m (8.33×10^{-4} degrees). TMI maps can undergo transformations to enhance detail and underlying features. Filters applied include reduction to the pole (RTP), first vertical derivative (1st VD), half-vertical derivative (0.5 VD), vertical gradient, total horizontal derivative, analytic signal, pseudo gravity (P_{grav}), pseudo susceptibility, upward continuation, and phase.

Gravity maps are useful for detecting underlying source rock through density anomalies, though they show less detail than magnetic anomaly maps. High gravity anomaly values indicate high-density source rock (Dentith & Mudge, 2014). Different types of gravity maps include observed gravity presented as free-air or Bouguer anomaly. The free-air correction accounts for height differences between the gravity meter and the datum level, while the Bouguer correction accounts for gravity contributions from topographic variations. Terrain correction, used with the Bouguer correction, accounts for mass above the measuring station. The original resolution of gravity layers is 435 m (4.17×10^{-3} degrees). Gravity map transformations include Complete Spherical Cap Bouguer Anomaly (CSCBA), isostatic residual, 1st VD, 0.5 VD, tilt angle, de-trended isostatic residual anomaly, and free air.

The source mafic–ultramafic magmas travel through underlying pathway structures, suture zones, and faults along margins and within cratons and subduction/collisional zones. Fault layers from geological provinces and ages in MinView's Seamless Geology package are treated as polylines, with the proximity of each mineral occurrence to the closest line used as a feature to proxy mineralization areas. These polylines include intrusion boundaries, metamorphic boundaries, and metamorphic isograds. Geological mapping can identify lithological and intrusion boundaries that act as pathway structures for the source material.

The next stage of the mineral system involves the Ni–Co laterite system within the weathering profile at the surface. The geophysical dataset from Geoscience Australia includes layers suited for identifying surficial mineral system processes. For instance, raster data such as geoid DEMs represent terrain/topography elevations more accurately, while ellipsoid DEMs offer globally consistent elevation levels. The original resolution of these DEM layers is 435 m (4.17×10^{-3} degrees).

The ASTER geoscience map of Australia, compiled by Commonwealth Scientific and Industrial Research Organization (CSIRO) in collaboration with Geoscience Australia and other agencies, is a remote sensing dataset containing calibrated, processed, and standardized mosaic ASTER products. The spectral bands of ASTER span wavelengths responsive to key rock-forming minerals. These final products cover a range of mineral group content, including green vegetation, ferric oxide, ferrous iron, opaque minerals, AlOH, kaolin, FeOH, MgOH, silica, quartz, and gypsum (Cudahy et al., 1997). ASTER maps are treated as raster files, with pixel values representing band ratio magnitudes (detected minerals), ranging from low abundance (blue) to high abundance (red). The original resolution of ASTER layers is 30 m (2.78×10^{-4} degrees).

Similar to ASTER data, airborne radiometric data can identify laterite zonation by spotlighting areas with high concentrations of elements associated with lateritic Ni–Co deposits. This approach is enhanced by additional geophysical mapping techniques to mitigate distortions caused by weathering processes and surface cover (Isles & Rankin, 2013).

The original resolution of radiometric layers is 100 m (0.001 degrees). Numerical enhancements, such as channel/element ratios, amplify channel variations and reduce the negative effects of water content, providing more useful correlations with different lithotypes due to higher resolution (Dentith & Mudge, 2014).

The spatial distribution of weathering units and profiles is valuable when correlating with known lateritic Ni–Co occurrences. This is represented by regolith mapping and inferred landscape evolution from the Seamless Geology package, specifically the Cenozoic Sedimentary Province polygon data. While the model is trained to correlate mineral occurrences with basement source rock polygons (mafic–ultramafic), these occurrences may also correlate with overlying sedimentary polygons representing weathering profiles.

Geochemical data provide useful sample composition data from rock, soil, and drill hole samples, identifying Ni and Co trace elements. However, due to the limited number of samples across the study area, geochemical data are more useful for assessing the accuracy and reliability of predicted prospective areas. Many data points, such as stream sediment samples, represent elements regionally rather than locally, which could reduce their reliability as training data for the ML algorithm.

FRAMEWORK

Data Processing Steps

The ML workflow is inspired by the work of Farahbakhsh et al. (2023), as shown in Figure 3. The preprocessing steps involve collecting and preparing a geodatabase of public datasets. This step is important for generating reliable mineral prospectivity maps since the data are usually available in various forms from multiple sources. The minerals system knowledge of the targeted critical metals guides the data layers selected. Feature layers were created from the datasets, and the ML model was trained and tested using the training set of known mineral occurrences to predict the testing set. Performance metrics and plots were calculated by comparing the predictions to the known testing set. If the performance was not satisfactory, the model was adjusted, and additional data or different modeling techniques may be required. Incorporating particular ML algorithms addresses common problems, such as the lack of a sufficient number of positive samples and a reliable set of negative samples.

The geological and geophysical data layers were translated into training proxy layers through mathematical equations and textural features, addressing the difficulty of defining negative training samples using positive and unlabeled bagging. Bagging or bootstrap aggregating is a technique that improves performance by averaging multiple predictions (Sun et al., 2020). Imbalances in the number of positive and negative training samples were addressed using the recently developed SMOTE–GAN method (Farahbakhsh et al., 2023), a technique to balance data by creating realistic and diverse synthetic examples for the minority class, producing a high-accuracy model using the Random Forest classifier and the bagging process. These methods can handle complex and nonlinear exploration data and improve the recognition of hidden patterns. After modifying the model, the next step was to create prospectivity maps using the most accurate model.

The workflow started with determining the research questions and then searching for relevant information on the target commodity's mineral system. This provides insight into collecting the appropriate data from public exploration datasets to form a geodatabase, which is later processed as data layers. It is important to document and clean these datasets as they may come in different forms from different sources. The downloaded data were processed to 0.01 degrees ($\sim$ 1110 m) resolution and clipped to NSW in GDA94 projection within the workflow. New data files were generated after processing, incorporating the clean NSW database. Then, the ML model split the data into a training set and a testing set to train and evaluate, respectively, the model's performance. All processing was carried out within a Python notebook utilizing a variety of libraries.

The particular data types used within the workflow included point, vector, and raster data, which were further elaborated in the Data Layers and Features section. Features were then generated using the initial datasets and treated as the training data files associated with the positive and negative samples. We calculated the distance to known deposit coordinates for the faults and contacts. We assigned binary numbers for the geological rock unit categories based on their association with known deposit coordinates. We calculated the mean, standard deviation, dissimilarity, and correlation for the

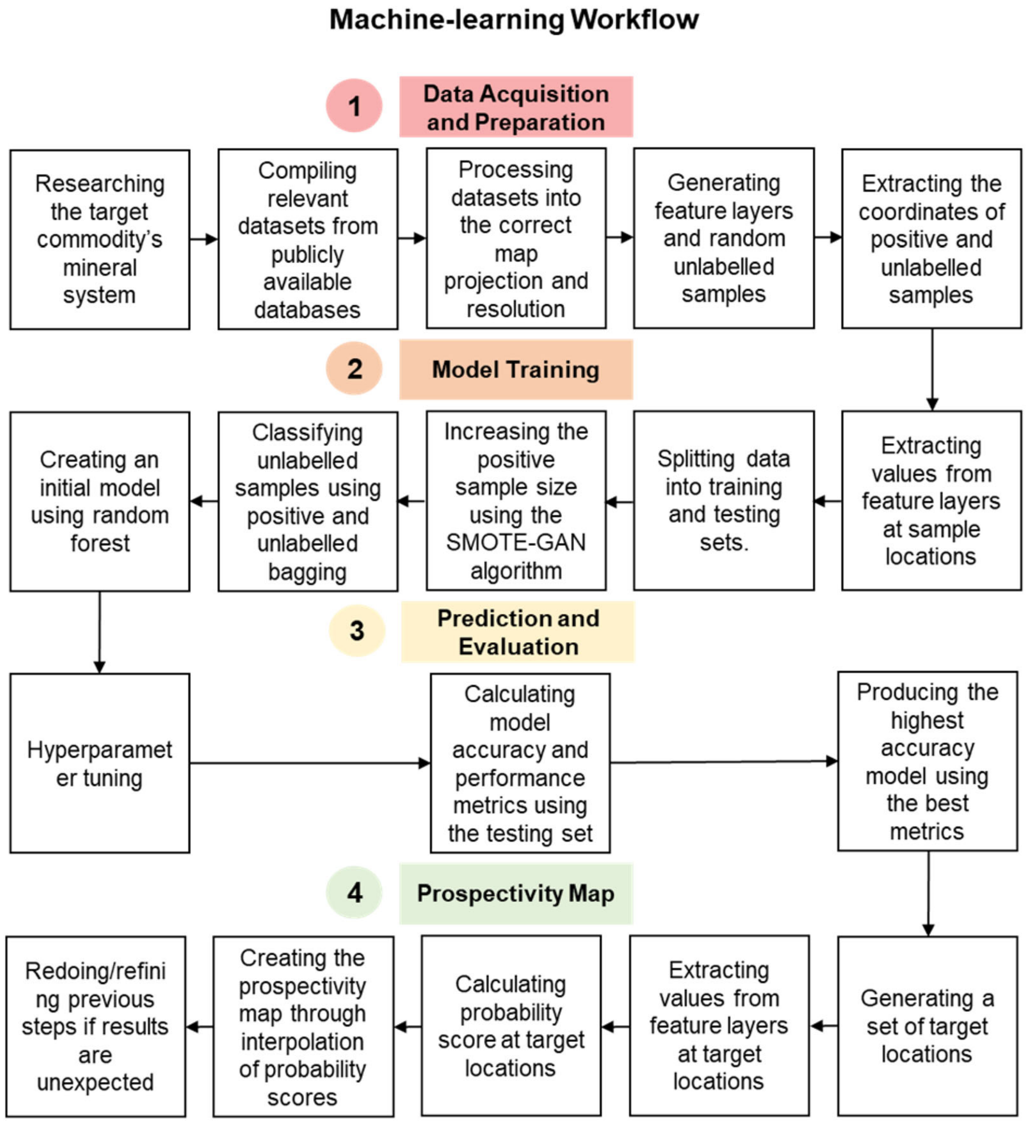


Figure 3. Machine learning workflow for producing the Ni–Co laterite prospectivity map (modified from Farahbakhsh et al., 2023).

grid data. Certain processes were used to enhance the detail and to create feature layers with the raster data. Firstly, the mean and standard deviation of pixel values in a circular buffer zone surrounding each target point were calculated, providing information on the surrounding area of each point. Additionally, we calculated the textural features, such as the correlation and dissimilarity of pixels in a square buffer zone, to enhance the spatial detail/information of the geology and structures around each point (Haralick et al., 1973). The DEM data were processed differently by calculating the horizontal gradients of the layers, which involved the mean value in a circular buffer zone surrounding each target point.

These processes allowed us to transfer the conceptual mineral system into the final feature data matrix of numerical values used for training the model and into mappable exploration criteria. To achieve a balanced training dataset, it was essential to significantly increase our sample size—targeting a ratio of 10 times more samples than features (Alwosheel et al., 2018). To accomplish this, we implemented the SMOTE–GAN algorithm. It enabled us to make predictions in the absence of a large amount of known mineral occurrences, which is a common issue in underexplored areas. Around 308 features, along with the set of known mineral occurrences, were input into the SMOTE algorithm, and the newly generated set was used to train the GAN algorithm to produce more positive samples.

The SMOTE–GAN algorithm only works with numerical values. Thus, the binary numbers assigned for the categorical data were used to calculate the mode for a particular feature, which was then used to assign to the synthetic positive samples. The main classifier is Random Forest, predicting mineralization probability. Positive and unlabeled learning (PUL) reduces false positive/negative results and prevents over-fitted data associated with training on only positive or negative samples. Another common issue is the difficulty in identifying negative samples for training, as it is challenging to pinpoint locations with a zero percent chance of mineralization. The PUL algorithm learned from a training set consisting of positive samples and a set of unlabeled samples and then assigns the probability of a sample being labeled positive. The positive and unlabeled bagging (PUB) as a subset of PUL improved the performance and accuracy of the model by training multiple models with different sets of training data and combining and averaging the predictions.

The performance metrics of the predictions were then evaluated, leading to adjustments to the model when necessary. This may involve adding/removing data, adjusting the hyperparameters of the algorithms and/or using other algorithms. The goal was to produce an accurate model—low numbers of false-positive and false-negative predictions—while having a balanced amount of positive and negative samples. The Random Forest classifier was then run multiple times, and the results were averaged to create an accurate model along with a list of important features.

Almost the entire NSW dataset was used in the initial run of the notebook to generate a large feature importance list, which we call 'Model 1.' Features that did not reach a certain cutoff threshold were removed, and only important features were used in the final prospectivity mapping. The first cutoff importance threshold is 0.001 for 'Model 2,' leaving 82 features. The second threshold is 0.01 for 'Model 3,' leaving 27 features. The next stage involved feature engineering to better understand the machine algorithm outputs and address unexpected results against common sense. Models 1–3 involved the entire dataset, while a hybrid approach applied more knowledge about the mineral system by selecting specific features as they fit within the different mineral system stages. High-ranking features from Models 1–3 were selected to highlight the importance of primary mineralization (source, pathways) in 'Model 4' and secondary mineralization (trap, preservation) in 'Model 5.' We implemented a tenfold cross-validation to determine the most accurate model and avoid overfitting. This involved dividing the training dataset into ten subsets, where nine were used for training and the remaining one was used for testing. This was repeated ten times (Sun et al., 2020).

Determining the optimal cell size for a ML-based mineral prospectivity map required balancing the highest and lowest resolutions of the integrated data layers to ensure the retention of detail without losing information or creating false precision. Additionally, the scale of analysis, computational efficiency, and the specific requirements of the map's purpose and end users must be considered. We conducted various experiments, including sensitivity analysis and model testing, to identify a cell size that provided accurate, robust, and manageable results. In evaluating the best cell size for our mineral prospectivity map at a scale of 1:3,000,000, a sensitivity analysis revealed how different cell sizes affect the balance between detail and readability. If we were to use a 0.01-degree cell size, this would offer very high detail, matching the resolution of input raster files to the workflow. However, at a scale of 1:3,000,000, each cell would be extremely small, making the map overly detailed and difficult to interpret. The fine resolution would result in a map that is cluttered and impractical, as individual cells would be too small to distinguish easily.

A 0.02-degree cell size provided slightly less detail but still maintained a high level of complexity. While this size was more manageable than 0.01 degrees, it would still result in a very dense map where each cell is only slightly more distinguishable. This could still overwhelm the viewer, making it challenging to extract meaningful patterns or insights from the map. Conversely, a 0.05-degree cell size strokes a better balance. It simplified the map enough to ensure readability while retaining important spatial patterns. At this scale, the cells were large enough to be easily distinguishable, making the map clearer and more practical for both analysis and presentation. The 0.05-degree cell size allowed the map to be detailed enough to convey essential information without becoming cluttered or difficult to interpret.

If the cell size were increased to 0.1 degrees, the map would be much simpler to read, with clearly visible cells. However, this larger size could oversimplify the data, potentially losing important details and local variations that were crucial for accurate prospectivity analysis. While a 0.1-degree cell size might be useful for a broad overview, it would not be suitable for more detailed geological exploration. Therefore, for a map at a scale of 1:3,000,000, a 0.05-degree cell size emerged as the optimal choice. It maintained a balance between preserving necessary detail and ensuring the map remains clear, readable, and effective for its intended purpose. This made the 0.05-degree cell size the most suitable option for our mineral prospectivity map.

GLCM Feature Map Texture

We produced textural feature maps for the grid data including geophysical layers to enhance the spatial detail or provide new ways to visualize the data. We achieved this using the gray level co-occurrence matrix (GLCM), which tabulates how often pixel brightness values (gray levels) occur within a specified area. The spatial relationships we wanted to visualize had different co-occurrence matrices (Hall-Beyer, 2017). The correlation texture measured the linear dependency of gray levels on the surrounding pixel-gray levels. According to the below Eq. (1), a high correlation value suggests that the pixels have a strong linear relationship within the specified window (Hall-Beyer, 2017). Therefore, based on the specified window size on the grids, high correlation values indicate that all the magnitudes in that window represent one feature. If the correlation value starts to decline upon decreasing the window size, the size of the feature could be indicated. The dissimilarity texture looks for a contrast between adjacent pixels and measures the difference in intensity between the pixel pairs. The sum of these values in the GLCM represents the dissimilarity for a particular window (Hall-Beyer, 2017). In terms of geophysical map interpretation, high dissimilarity values indicate a wide range of pixel magnitudes in that window/area and could represent a complex texture or a nonhomogeneous geophysical feature. The scikit-image package implemented in the workflow utilized the following equations to produce the correlation and dissimilarity maps (Van der Walt et al., 2014; Scikit-image Team, 2023):

$$\text{Correlation:} \quad \sum_{ij=0}^{\text{levels}-1} P_{ij}\left[\frac{(i-\mu_i)\left(j-\mu_j\right)}{\sqrt{\left(\sigma_i^2\right)\left(\sigma_j^2\right)}}\right] \tag{1}$$

where P_{ij} represents the pixel intensity at position (i, j) in a matrix P, being summated over all combinations of indices i and j; μ_i and μ_j are the mean values (central tendency) of rows (μ_i) and columns (μ_j) in the matrix P; and σ_i^2 and σ_j^2 represent variance (spread or dispersion) of rows (σ_i^2) and columns (σ_j^2) in the matrix P. Together, we summed the product of the standardized deviations of i and j from their respective means, divided by the product of the square roots of their variances. It quantifies how each position (i, j) deviates from the mean in terms of standard deviations (a normalized measure).

$$\text{Dissimilarity:} \quad \sum_{ij=0}^{\text{levels}-1} P_{ij}|i-j| \tag{2}$$

where $|i - j|$ is the measure of dissimilarity between positions i and j (higher value when i and j are far). Together, we summed the dissimilarity values ($|i$ $j|$) for all pairs of positions (i, j) in the texture; each term in the sum is weighted by the value of P_{ij}.

SMOTE–GAN

There is an insufficient number of positive samples for training, which would lead to inaccurate mineralization prediction. To create a balanced training dataset, we used the SMOTE–GAN algorithm. GAN employs neural networks and it is described as a game to deceive the discriminator with generated synthetic samples that are not in the training set and are indistinguishable from real samples. The generator function trains on the mineral occurrences and mineral system features, while the discriminator function tests the generated outputs against the known positive data. A third function combines the two functions and updates the generator model based on its performance against the discriminator (Zeiler & Fergus, 2014). The uniqueness of incorporating the SMOTE algorithm assists with addressing imbalanced datasets. SMOTE can increase the number of positive samples by interpolating between an original sample and its nearest neighbors (Fig. 4). This also helps to reduce the processing time for GAN by the SMOTE output, which is then input into the GAN algorithm to train on a larger set initially.

Positive and Unlabeled Learning

Different methods exist to create negative samples, each with its pros and cons (Farahbakhsh et al., 2023). Training on negative samples that are hard to identify can lead to overfitting and cause false positives or negatives. Positive and unlabeled learning (PUL) was employed to address this issue.

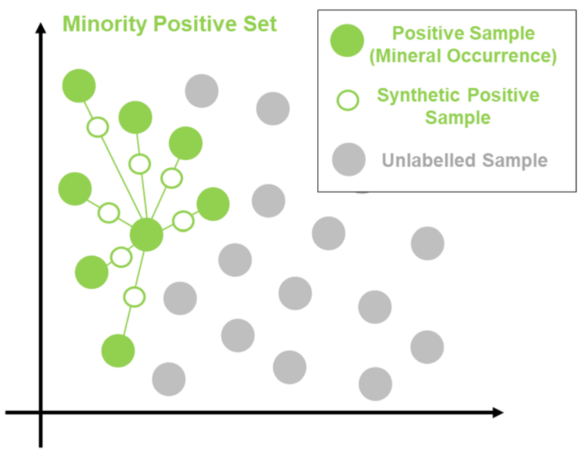


Figure 4. SMOTE technique, synthesizing new points by interpolating between samples and nearest neighbors.

This semi-supervised binary classification approach recovers labels from unknown samples by learning from positive samples and assigning labels to unknown samples (Mordelet & Vert, 2014; Farahbakhsh et al., 2023) (Fig. 5). The goal was to increase the likelihood of accurately labeling samples as positive by identifying common features in the unlabeled samples from the known positive set. As a result, the positive set grew, enabling the model to classify unlabeled samples and distinguish negative samples accurately.

This study employed a PUL method called positive and unlabeled bagging (PUB) to label unknown samples (Mordelet & Vert, 2014). PUB is a parallelized bootstrap approach where the algorithm trains multiple binary classifiers to discriminate known positive examples from random subsamples of the unlabeled set and averages their predictions (Fig. 5) (Mordelet & Vert, 2014; Gan et al., 2022). By averaging multiple-run predictions, PUB improves performance, leading to more accurate predictive models (Farahbakhsh et al., 2023). The objective function in PUB integrates loss components for both positive and unlabeled samples, balancing initial assumptions about unlabeled data with iterative refinements. The goal was to minimize the misclassification of positive instances while accurately identifying negatives within the unlabeled set, leveraging ensemble methods to enhance robustness and performance. It was necessary to have at least ten times more training samples than features to create a meaningful model (Figueroa et al., 2012; Farahbakhsh et al., 2023). Therefore, we generated

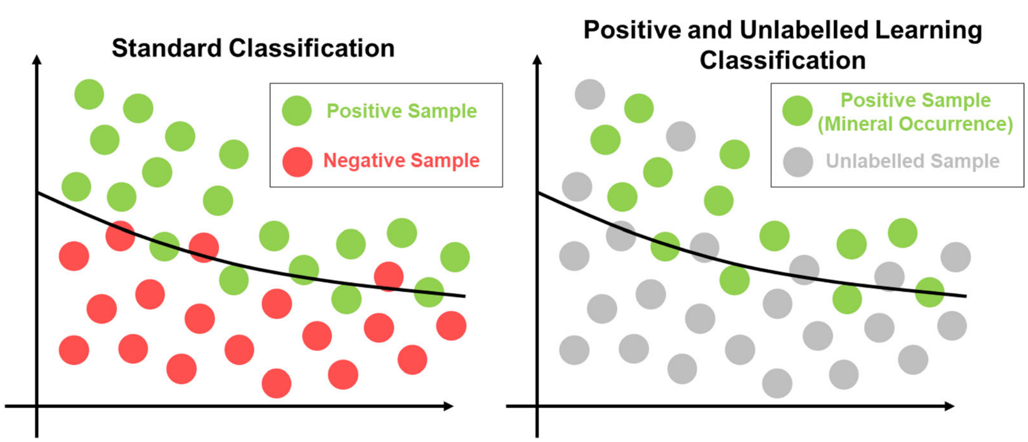


Figure 5. Positive-unlabeled learning for training the algorithm in the absence of negative samples (Jaskie & Spanias, 2019).

equal numbers of random and synthetic positive samples in the area. We can apply PUB to any supervised algorithm to obtain reliable positive and negative samples. In this study, we use Random Forest as the main classifier.

Random Forest

Random Forest is a classification algorithm used to make predictions of the phenomenon represented by the training data. It consists of multiple decision trees based on a set of restrictions or conditions, where a decision is made between a parent node and its child nodes (Breiman, 2001; Carranza & Laborte, 2015; Ford et al., 2019). It utilizes training subsets of random samples that can be reselected for generating other trees in a ''bootstrap aggregating'' procedure, also known as bagging (Sun et al., 2020). Approximately 2/3 of the set was employed for prediction, while the other 1/3 was used for validation and prediction accuracy (Carranza & Laborte, 2015). Trees grow by splitting into binary leaf nodes until a stop condition is reached. The optimal decision tree and node were determined by the maximum purity of the tree, which measures the probability of a random sample from the input dataset being labeled correctly (Sun et al., 2020). In terms of use in MPM, the trees determine whether a mineral system proxy/spatial feature value predicts a known mineral occurrence or not based on a portion of the training subset. The accuracy of the model was validated with the remaining portion of the training subset, based on the amount of true positive and negative (correctly labeled) samples. The features with the highest accuracy/purity were considered prospective or unprospective and are used to generate final mineral prospectivity maps (Carranza & Laborte, 2015; Ford et al., 2019). Figure 6 outlines the detailed steps taken to generate the necessary positive and negative samples using the framework employed in this study.

Hyperparameter Tuning

Hyperparameters are like settings of a ML algorithm that the user must set, affecting the performance and accuracy of the results. These were adjusted through the Scikit-Learn package (Pedregosa et al., 2011) by trialing multiple sensible parameters and tuning/validating the model through cross-validation until the optimal combination of hyperparameters is selected for the highest model

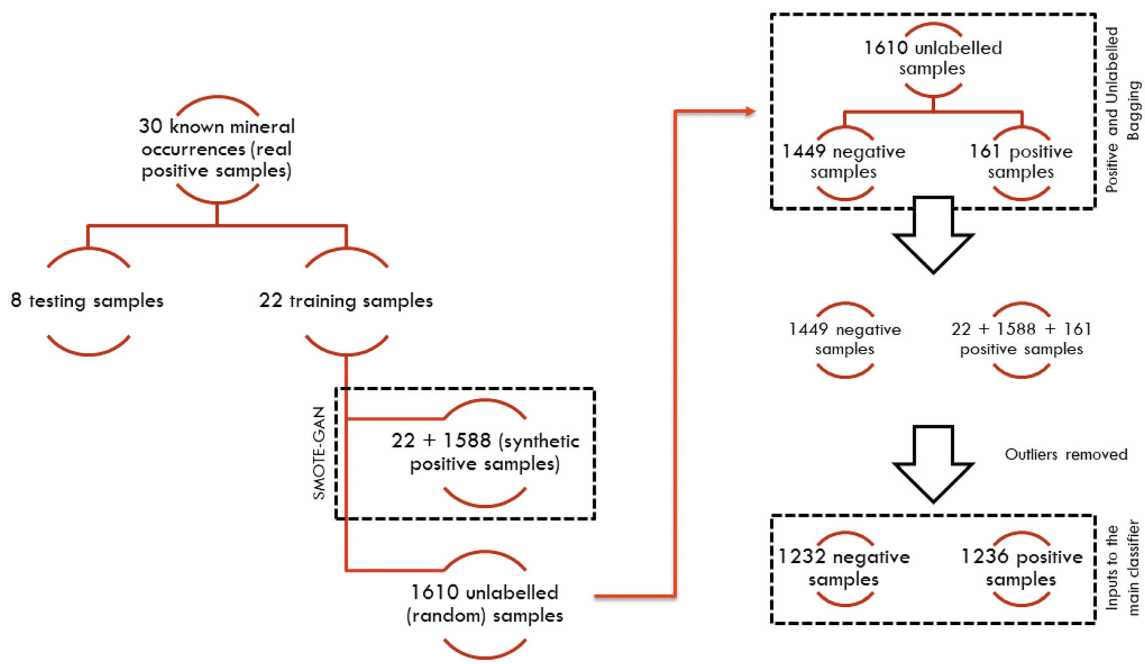


Figure 6. The process of generating the necessary positive and negative samples for training the Random Forest classifier.

accuracy (Koehrsen, 2018a). The cross-validation technique helps prevent the model from overfitting training data by evaluating the performance of each subset/fold of a training set, where each subset was tested by training on the other subsets. The average of the subset performances was used to validate the model, and the higher the number of subsets/folds, the lower the chances of overfitting (Koehrsen, 2018a).

Bayesian optimization was used to find the optimal hyperparameters, which builds a probability model that aims to be ''less wrong'' through continuous updates and trials with the true function. The steps involved include (Koehrsen, 2018b):

1. Creation of a surrogate probability model;
2. Proposing optimal parameters for the surrogate model (highest probability score);
3. Application of these hyperparameters on the true function;
4. Modification of the surrogate model by incorporating the new results;
5. Repeat steps 2–4 once specified iterations/time is reached.

The main hyperparameters in the Random Forest algorithm were automatically adjusted in the workflow for the best results (as described in Koehrsen, 2018b), and they are presented in Table 5.

Hyperparameter tuning was also applied to the SMOTE–GAN algorithm to refine the accuracy and outputs of the Random Forest model through experimentation. The main hyperparameter adjusted was the learning rate, which produced the highest accuracy at the rate of 0.01.

RESULTS

The ML models were trained using various geological and geophysical data layers spanning New South Wales (Table 6). These models produced prospectivity maps that identified potential zones for lateritic Ni–Co mineralization, with performance metrics provided in Table 7. While the study was based on real positive samples, it lacked negative samples, limiting the ability to calculate comprehensive performance metrics. However, by incorporating real data along with synthetic data generated via the SMOTE–GAN and PUB methods, we can evaluate the models using a broader range of metrics. Accuracy, the most straightforward metric, measures the proportion of correct predictions (true positives and true negatives) relative to the total number of cases. The F1-score, a harmonic mean of precision and recall, serves as a balanced metric, especially useful for model comparison. Precision indicates the proportion of true positive predictions among all positive predictions, which is critical when false positives are costly. Recall reflects the proportion of true positives identified out of all actual positives, which is essential when minimizing false negatives is important. Since the F1-score incorporates both precision and recall, we only reported the F1-score in Table 7. In the framework, Model 1 utilized the full feature set, Model 2 used features with importance scores above 0.001, and Model 3 used features with importance scores above 0.01. Model 4 included hypogene (primary) features, while Model 5 incorporated supergene (secondary) features. All five models achieved an accuracy rate of 75%, correctly predicting 6 out of 8 known mineral occurrences. Models 1–3 maintained consistent accuracy despite removing feature layers, and Models 4 and 5 retained accuracy despite focusing on hypogene and supergene mineralization features.

The Bayesian optimization progress plots in Figure 7 show the evolution of the mean accuracy during 50 iterations of hyperparameter tuning for the five prospectivity models, each using a different set of features. Figure 7a represents the model using the entire feature set. This model generally shows

Table 5. Main hyperparameters used in the Random Forest algorithm

Hyperparameters	Use	Optimal value
n_estimators	The number of trees in the forest	180
max_features	The maximum number of features considered for splitting a node	sqrt
max_depth	The maximum number of levels in each decision tree	15
min_samples_split	The minimum number of data points placed in a node before it splits	2
min_samples_leaf	The minimum number of data points allowed in a leaf node	2
bootstrap	The method for sampling data points (with or without replacement)	False

sqrt square root of the number of features

Table 6. List of data layers sourced from NSW databases used to create feature layers as input for training

Data type		Data layer	
Vector	Polyline	Rock Unit bounding faults	Lachlan Orogen (LAO)
			Cenozoic Sedimentary Province (CSP)
		Intrusions boundaries	
		Metamorphic isograds	
		Metamorphic boundaries	
	Polygon	Rock Units	Lachlan Orogen
			Cenozoic Sedimentary Province
		Intrusions	Tabberabberan
		Metamorphic Facies	Benambran
			Tabberabberan
			Kanimblan
Raster	Magnetic	Total magnetic intensity (TMI) with 40 m resolution 2019	
		Total magnetic intensity (TMI) with 80 m resolution 2019	
		TMI Reduced-to-pole (TMI RTP)	
		First vertical derivative of TMI	
		First vertical derivative of TMI RTP	
		Half-vertical derivative of TMI RTP	
		Analytic signal of TMI RTP	
		Pseudo gravity (Pgrav) of TMI RTP	
		Pseudo Susceptibility of TMI RTP	
		Vertical gradient of TMI RTP	
		Upward continued (up to 50 km) TMI RTP	
		Total Horizontal Derivative of PGrav TMI RTP	
		Phase of TMI RTP	
	Gravity	Complete spherical cap Bouguer anomaly (CSCBA) gravity 2016	
		Isostatic Residual gravity 2016	
		Spherical cap Bouguer anomaly gravity 2016	
		Complete spherical cap Bouguer anomaly (CSCBA) gravity 2019	
		First vertical derivative of CSCBA 2019	
		First vertical derivative of CSCBA including airborne survey (IA) 2019	
		Half-vertical derivative of CSCBA 2019	
		Half-vertical derivative of CSCBA IA 2019	
		Tilt angle of CSCBA 2019	
		Tilt angle of CSCBA IA 2019	
		CSCBA IA 2019	
		De-trended isostatic residual anomaly (DTGIR) gravity 2019	
		First vertical derivative of DTGIR 2019	
		First vertical derivative of DTGIR IA 2019	
		Half-vertical derivative of DTGIR 2019	
		Half-vertical derivative of DTGIR IA 2019	
		Tilt angle of DTGIR 2019	
		Tilt angle of DTGIR IA 2019	
		DTGIR IA 2019	
		Free-air gravity 2019	
		Free-air gravity IA 2019	
	Radiometric	Radiation dose rate unfiltered	
		Radiation dose rate filtered	
		Potassium (K) concentration unfiltered	
		Potassium concentration filtered	
		Thorium (Th) concentration unfiltered	
		Thorium concentration filtered	
		Uranium (U) concentration unfiltered	
		Uranium (U) concentration filtered	
		Ratio of Th to K	
		Ratio of U to Th	
		Ratio of U to K	
		Ratio of U squared to Th	

Table 6. continued

Data type		Data layer
	Remote Sensing	AlOH group composition
		AlOH group content
		FeOH group content
		Ferric oxide composition
		Ferric oxide content
		Ferrous iron content in MgOH
		Ferrous iron index
		Green vegetation content
		Gypsum index
		Kaolin group index
		MgOH group composition
		MgOH group content
		Opaque index
		Quartz index
		Silica index
	Miscellaneous (Elevation)	Ellipsoid digital elevation model (DEM)
		Geoid DEM
		Airborne survey drape ellipsoid (Ausdrape elevation)
		Airborne survey drape geoid (Ausdrape elevation)

Resolution for all layers (except vector data—not applicable) is 0.01 degrees

high accuracy, fluctuating between 0.96 and 0.985 across iterations, with some variability but a consistent upward trend over time. This was likely due to including all potential variables, which allowed the model to leverage a broad range of information despite some noise from irrelevant features. The overall accuracy remained high, likely because of the rich feature set covering relevant and partially irrelevant data points. By iteration 50, the model reached its peak performance. Figure 7b displays the model using only features with importance greater than 0.001. This plot shows a more stable performance than Figure 7a, with accuracy primarily above 0.97. Eliminating less significant features has likely reduced the noise in the model, making it more robust. Although there were still fluctuations, the tighter accuracy range, and consistently high values suggest that the retained features are more informative. By the final iterations, the accuracy stabilized around 0.99, indicating that the model had successfully identified the optimal parameters.

Figure 7c shows the model using features with importance greater than 0.01. This model experienced greater variability in performance compared to Figure 7b, with mean accuracy fluctuating between 0.95 and 0.985. The more stringent feature selection criterion may have excluded some useful features, causing a wider range of performance during cross-validation. However, the overall accuracy remained high, implying that the model's most critical features are still present. As the optimization progressed, the model converged toward an accuracy just above 0.98, slightly lower than the previous models. Figure 7d represents the model using only hypogene features. The accuracy in this case was also relatively high, with mean accuracy fluctuating between approximately 0.95 and 0.99. Some notable drops in accuracy, especially around iterations 20 and 40, indicate that the model was more sensitive to certain parameter settings when restricted to hypogene features. These fluctuations could be due to the nature of hypogene features being more limited or less predictive on their own for lateritic Ni–Co deposits. Despite this, the model managed to converge toward a stable accuracy of close to 0.99 by the end of the optimization.

Table 7. Performance metrics of the five models created using (1) entire feature set, (2) features showing importance greater than 0.001, (3) features showing an importance greater than 0.01, (4) hypogene features, and (5) supergene features

Model	Accuracy (real positive samples)	Accuracy	F1-score
1	0.750	0.993	0.992
2	0.750	0.994	0.992
3	0.750	0.994	0.992
4	0.750	0.964	0.967
5	0.750	0.944	0.958

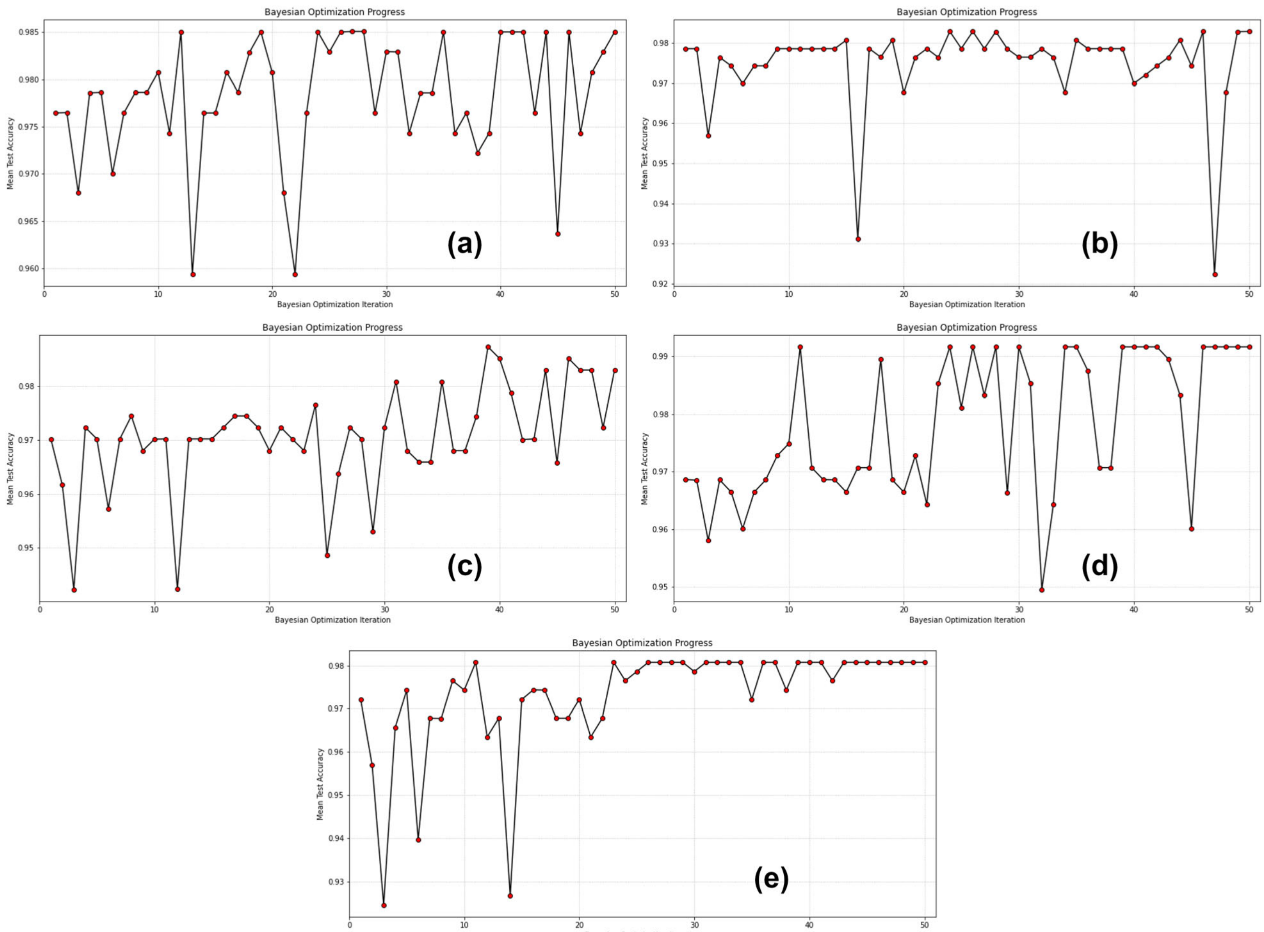


Figure 7. Bayesian optimization progress plots for the five models, each created using different feature sets: (**a**) the entire feature set; (**b**) features with an importance greater than 0.001; (**c**) features with an importance greater than 0.01; (**d**) hypogene features; and (**e**) supergene features. These plots illustrate the gradual improvement in the mean accuracy obtained by tenfold cross-validation.

Figure 7e shows the model using only supergene features. The most striking feature of this plot was its steady improvement, with accuracy rising rapidly and stabilizing at around 0.98 after approximately 40 iterations. This suggests that supergene features alone were highly predictive for this type of mineralization. The relatively smooth and stable progression of accuracy implies that these features provided a strong signal with less noise, leading to faster convergence and higher accuracy during the optimization process. The high accuracies across all models can be attributed to the strong predictive capacity of both hypogene and supergene features in the context of lateritic Ni–Co mineralization, along with the sufficient number of training samples generated using the SMOTE–GAN method. The models using the most significant features (greater than 0.001 or 0.01 importance) maintained high performance due to their ability to retain the most relevant information while discarding less impactful features. Additionally, the model using supergene features demonstrated particularly strong performance, likely because these features are closely related to surface geochemical processes that dominate lateritic mineralization systems.

The top 15 features, their importance scores, and cumulative importance from Models 1–3 for targeting lateritic Ni–Co deposits in the Lachlan Orogen are presented in Table 8. These tables show that removing features from the initial data-driven model has minimal impact on the prospectivity map outputs, with only minor shifts in feature rankings when low-importance features are excluded. Among the data layers, magnetic intensity, onshore gravity, remote sensing, Cenozoic clastic and regolith cover, Cambro–Ordovician sandstone/serpentinite, and

Table 8. Feature importance and cumulative importance of the top 15 features in Model 1, Model 2, and Model 3 for lateritic Ni–Co mineralization

Feature	Importance	Cumulative importance
Model 1		
Correlation of AlOH group content	0.099	0.099
Mean of the phase of TMI RTP	0.084	0.183
Benambran subgreenschist facies	0.067	0.250
Standard deviation of CSCBA gravity	0.055	0.305
CSP clastic sediments	0.041	0.346
Faulted boundaries in the Lachlan Orogen	0.039	0.385
Non-CSP units	0.038	0.423
Mean of CSCBA gravity	0.034	0.457
Unconformities of metamorphic boundaries	0.028	0.486
Dissimilarity of ferric oxide content	0.026	0.512
Mean of radiation dose rate	0.024	0.535
Mean of the pseudo gravity of TMI RTP	0.023	0.558
Horizontal derivative of the ellipsoid DEM along the *x*-axis	0.022	0.580
Correlation of the half-vertical derivative of TMI RTP	0.022	0.602
Mean of FeOH group content	0.021	0.623
Model 2		
Mean of the phase of TMI RTP	0.093	0.093
Correlation of AlOH group content	0.081	0.174
Benambran subgreenschist facies	0.078	0.252
Standard deviation of CSCBA gravity	0.073	0.325
CSP clastic sediments	0.046	0.371
Mean of CSCBA gravity	0.037	0.408
Non-CSP units	0.034	0.442
Unconformities of metamorphic boundaries	0.033	0.475
Dissimilarity of ferric oxide content	0.030	0.504
Faulted boundaries in the Lachlan Orogen	0.029	0.533
Correlation of the half-vertical derivative of TMI RTP	0.027	0.560
Horizontal derivative of the ellipsoid DEM along the x-axis	0.025	0.585
Mean of radiation dose rate	0.024	0.609
Mean of the pseudo gravity of TMI RTP	0.022	0.631
Dissimilarity of free-air gravity	0.019	0.649
Model 3		
Correlation of AlOH group content	0.134	0.134
Benambran subgreenschist facies	0.127	0.260
Mean of the phase of TMI RTP	0.115	0.375
CSP clastic sediments	0.080	0.456
Standard deviation of CSCBA gravity	0.076	0.532
Non-CSP units	0.057	0.589
Unconformities of metamorphic boundaries	0.045	0.634
Faulted boundaries in the Lachlan Orogen	0.040	0.674
Mean of radiation dose rate	0.030	0.704
Mean CSCBA gravity	0.027	0.732
Standard deviation of first vertical derivative of CSCBA gravity	0.022	0.754
Mean of the pseudo gravity of TMI RTP	0.021	0.775
Mean of ferrous hydroxide (FeOH) group content	0.021	0.796
Dissimilarity of silica index	0.021	0.817
Dissimilarity of the first vertical derivative of CSCBA gravity	0.020	0.837

Ordo–Silurian ultramafic basement units were the most influential, receiving the highest feature scores. Additionally, Table 9 present similar attributes obtained from Models 4 and 5, highlighting the most important hypogene and supergene features. Figure 8 shows probability maps for lateritic Ni–Co, revealing that regions with high prospectivity largely overlap with known mineral occurrences. The models not only identified areas with existing mineral occurrences but also revealed new regions

Table 9. Feature importance and cumulative importance of the top 15 hypogene and supergene features in Model 4 and Model 5 for lateritic Ni–Co mineralization

Feature	Importance	Cumulative importance
Model 4		
Correlation of AlOH group content	0.133	0.133
Mean of the phase of TMI RTP	0.125	0.258
Benambran subgreenschist facies	0.117	0.375
Faulted boundaries of the Lachlan Orogen	0.073	0.448
Standard deviation of CSCBA gravity	0.065	0.513
Unconformities of metamorphic boundaries	0.065	0.578
Mean of MgOH group content	0.063	0.641
Mean of the pseudo gravity of TMI RTP	0.051	0.691
Mean of CSCBA gravity	0.048	0.740
Unconformable boundaries of intrusion boundaries	0.041	0.781
Sandstones in the Lachlan Orogen	0.029	0.810
Dissimilarity of the pseudo gravity of TMI RTP	0.022	0.832
Correlation of opaque index	0.021	0.853
Benambran greenschist facies	0.018	0.870
Dissimilarity of the phase of TMI RTP	0.017	0.888
Model 5		
Correlation of AlOH group content	0.135	0.135
CSP clastic sediments	0.080	0.215
Correlation of green vegetation index	0.069	0.283
Faulted boundaries of the Lachlan Orogen	0.060	0.343
Mean of the radiometric dose rate	0.054	0.397
Dissimilarity of ferric oxide content	0.053	0.450
Mean of FeOH group content	0.047	0.496
Mean of Th/K ratio	0.045	0.541
Dissimilarity of silica index	0.043	0.584
Non-CSP units	0.043	0.627
Correlation of the opaque index	0.034	0.661
Correlation of the half-vertical derivative of TMI RTP	0.032	0.693
Correlation of silica index	0.023	0.716
Mean of the half-vertical derivative of TMI RTP	0.023	0.739
Standard deviation of the radiometric dose	0.022	0.760

extending from and surrounding these known sites. The ML models were validated by comparing the top-ranked features with published geological maps and known occurrences, enhancing the understanding of mineral systems and formation processes. This approach provides a valuable guide for future exploration and data collection efforts.

The initial data-driven Model 1 incorporated all the data layers and features listed in Table 6. The resulting prospectivity map (Fig. 8a) indicated a high likelihood of additional lateritic Ni–Co mineralization in areas surrounding and extending from known mineral occurrence clusters. Notable patterns in this model included two near-parallel linear belts trending roughly north–northwest to south–southeast on the western side (the Coolac–Thuddungra serpentinite belts) and a large lobate feature in the northwest (Syerston–Homeville–Nyngan mafic–ultramafic intrusion field), which contained the majority of known lateritic Ni–Co occurrences. Additionally, a curvilinear feature on the eastern side of the study area (Byng Volcanics—shoshonitic ultramafic lavas) showed no known lateritic Ni–Co occurrences, while a linear-lobate feature on the far eastern side (Bungonia manganiferous grits) hosts a cluster of exotic supergene Co occurrences in Cenozoic sandstone, for which no source has yet been identified. All of these features were considered potential exploration targets.

The prospectivity maps for Model 2 (82 features, with importance cutoff of 0.001) and Model 3 (27 features, with importance cutoff of 0.01) showed minimal changes in the overall distribution of prospective zones compared to Model 1, maintaining similar performance metrics (Figs. 8b, c). However, Model 3 may offer slightly tighter constraints

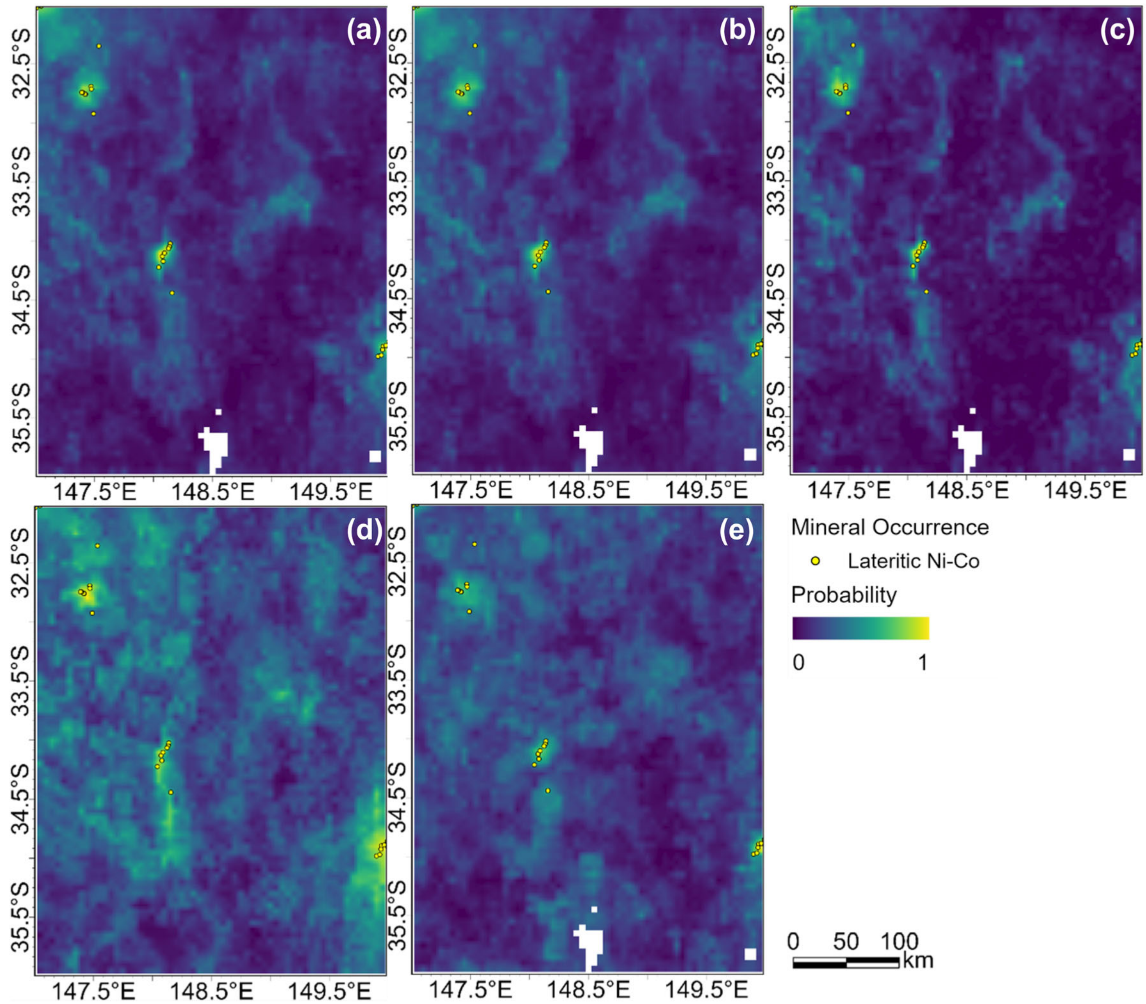


Figure 8. Prospectivity maps for lateritic Ni–Co mineralization, where the heat bar highlights areas with moderate to high likelihood (from blue to yellow) for mineralization: (**a**) Model 1 includes all features; (**b**) Model 2 includes features with importance above 0.001; (**c**) Model 3 includes features with importance above 0.01; (**d**) Model 4 uses hypogene features; and (**e**) Model 5 uses supergene features. White NaN values indicate missing radiometric data, except in Model 4, which excludes radiometric data.

on the probability zones, as indicated by a reduction in noise due to the fewer features used. The other two models (Models 4 and 5) were developed using features selected for their relevance based on mineral system knowledge. These models represent the Ni–Co primary source rocks (hypogene model) and the supergene-enriched lateritic Ni–Co deposits (supergene model), with low-importance features determined by Model 1 excluded. Figures 8d and e display the prospectivity maps for these hypogene and supergene models, respectively.

The prospectivity map generated from the hypogene model (Fig. 8d) aligns with the previous models and exhibited similar performance metrics but with better-defined (or sharper) boundaries around zones of higher prospectivity. The high-probability areas appeared slightly larger and more coherent within the areas of interest, suggesting the primary source rocks' influence extends further than the effects of weathering and laterite development. In contrast, the prospectivity map from the supergene model (Fig. 8e) appeared noisier, with more diffuse boundaries and more restricted high-prospectivity zones. This implies that the influence of deep weathering profiles was less extensive over the source rocks and that deep weathering impacts extend beyond the source rock units. Despite the noise, areas of interest remained apparent, and the overlap between the hypogene and supergene models may help in prioritizing exploration target areas.

In addition to the metrics reported in Table 7, we used prediction–area (P–A) plots (Fig. 9) to further evaluate the performance of the models. These plots display the relationship between the predicted area of mineral potential and the actual occurrence of mineral deposits (Yousefi & Carranza, 2015). They help assess how well a predictive model identifies areas likely to contain mineral resources. They are preferred over common metrics like accuracy and F1-score because they incorporate the spatial distribution of predictions, which is crucial in geological contexts. Additionally, they highlight how much area needs to be explored to find a certain percentage of the deposits, aiding in efficient resource allocation. P–A plots provide a clearer picture of the trade-off between covering large areas and the probability of finding deposits, offering a more practical and insightful evaluation for mineral prospectivity mapping.

In Figure 9a, which represents the model using the entire feature set, we observed that, at a probability threshold of 0.3, 92% of the known mineral occurrences were captured within just 8% of the study area. This model demonstrated high-performance level, effectively targeting most occurrences in a relatively small area. The model shown in Figure 9b performed similarly but slightly less effectively using features with importance greater than 0.001. At a probability threshold of 0.33, 93% of occurrences were captured, but the corresponding area was reduced to 7%, showing that excluding less important features did not negatively impact performance, and the area required to capture the occurrences decreased. In Figure 9c, the performance decreased somewhat using features with importance greater than 0.01. At a probability threshold of 0.24, 89% of occurrences were captured within 11% of the area. This model required a larger area to capture a slightly lower percentage of occurrences compared to the previous models, suggesting that increasing the threshold for feature importance led to the exclusion of useful information.

Figure 9d represents the model using only hypogene features. At a probability threshold of 0.285, 91% of the occurrences were captured within 9% of the area. This performance is comparable to the model shown in Figure 9b, showing that the hypogene features alone effectively capture a large percentage of the occurrences. Finally, Figure 9e represents the model using only supergene features. This model captured 88% of occurrences at a probability threshold of 0.345 but required 12% of the area, indicating a lower performance than the other models. The supergene features seem less effective in isolating mineral occurrences in a compact area than the entire feature set or hypogene features. The model using the entire feature set (Fig. 9a) and the one using features with importance greater than 0.001 (Fig. 9b) performed the best, capturing a high percentage of known occurrences in a smaller percentage of the area. The hypogene feature model (Fig. 9d) also performed well, while the models using higher feature importance thresholds (Fig. 9c) or just supergene features (Fig. 9e) showed reduced efficiency.

Figure 7 focuses on model accuracy during hyperparameter tuning, showing that all models generally achieved high performance, with the supergene feature model showing a smooth improvement and the models using more significant features (importance > 0.001 or 0.01) showing stable, high accuracies. However, Figure 9 introduces spatial context by evaluating how well the models capture known mineral occurrences within a compact area. Here, the models using the entire feature set and features with importance > 0.001 performed best, capturing a high percentage of occurrences within a smaller area, while models using stricter feature selection or just supergene features captured fewer occurrences over a larger area. This shows that, while all models performed well in accuracy (Fig. 7), their spatial efficiency varied, with supergene features less effective in compactly identifying occurrences (Fig. 9).

DISCUSSION

Feature Overview

We created functional prospectivity maps by integrating publicly available geological and geophysical exploration datasets into a ML framework. These maps highlighted areas with potential for discovering new lateritic Ni–Co resources in the East-Central Lachlan Orogen of NSW, showing strong spatial correlations between high-probability zones and known mineral occurrences while also identifying new exploration targets (Fig. 10). The workflow involved developing models to process

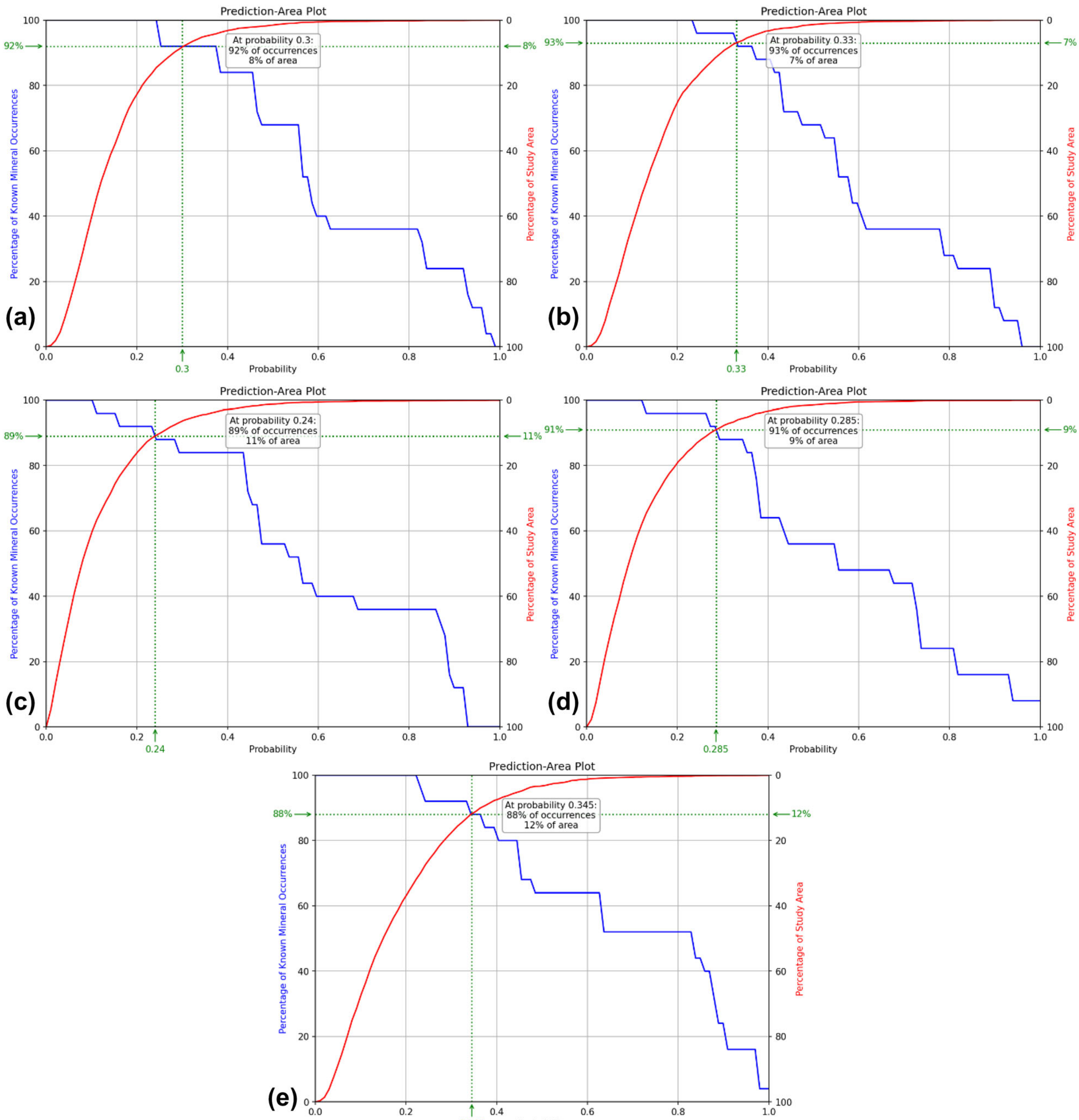


Figure 9. Prediction-area plots for the five models, each based on different feature sets: (**a**) the entire feature set; (**b**) features with importance greater than 0.001; (**c**) features with importance greater than 0.01; (**d**) hypogene features; and (**e**) supergene features. These plots illustrate the relationship between predicted mineral potential area and actual occurrence of mineral deposits.

multidimensional exploration data and generate prospective targets for lateritic Ni–Co. The feature datasets used to train the models aligned with those documented for the lateritic Ni–Co mineral system, producing expected results for prospectivity over mapped source rocks in the Lachlan Orogen.

The prospectivity maps demonstrate the effectiveness of the methods in delineating mineralized zones within the complex geological terrain of the

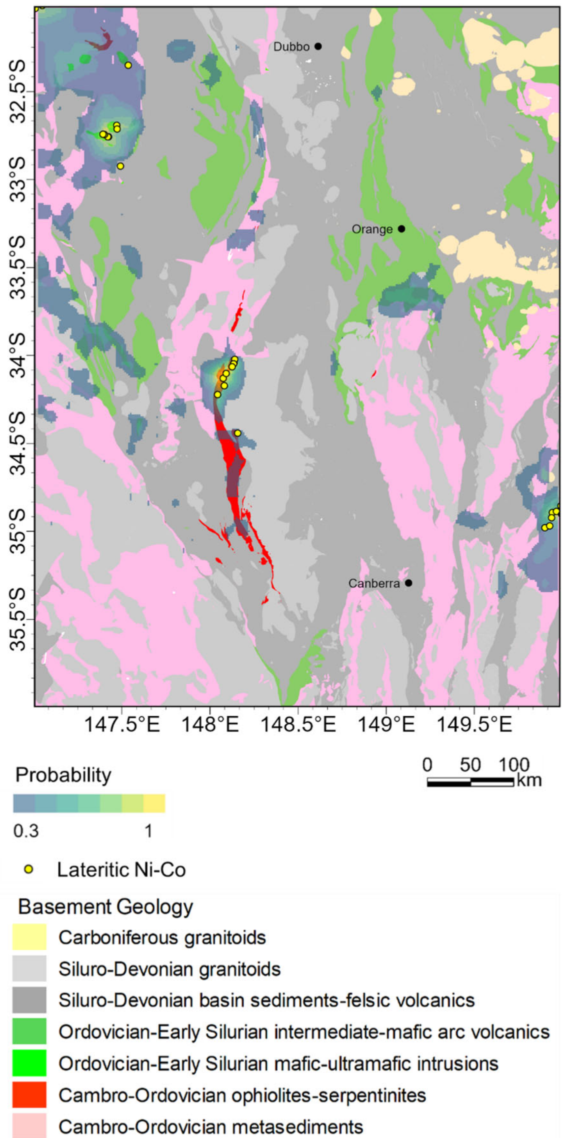


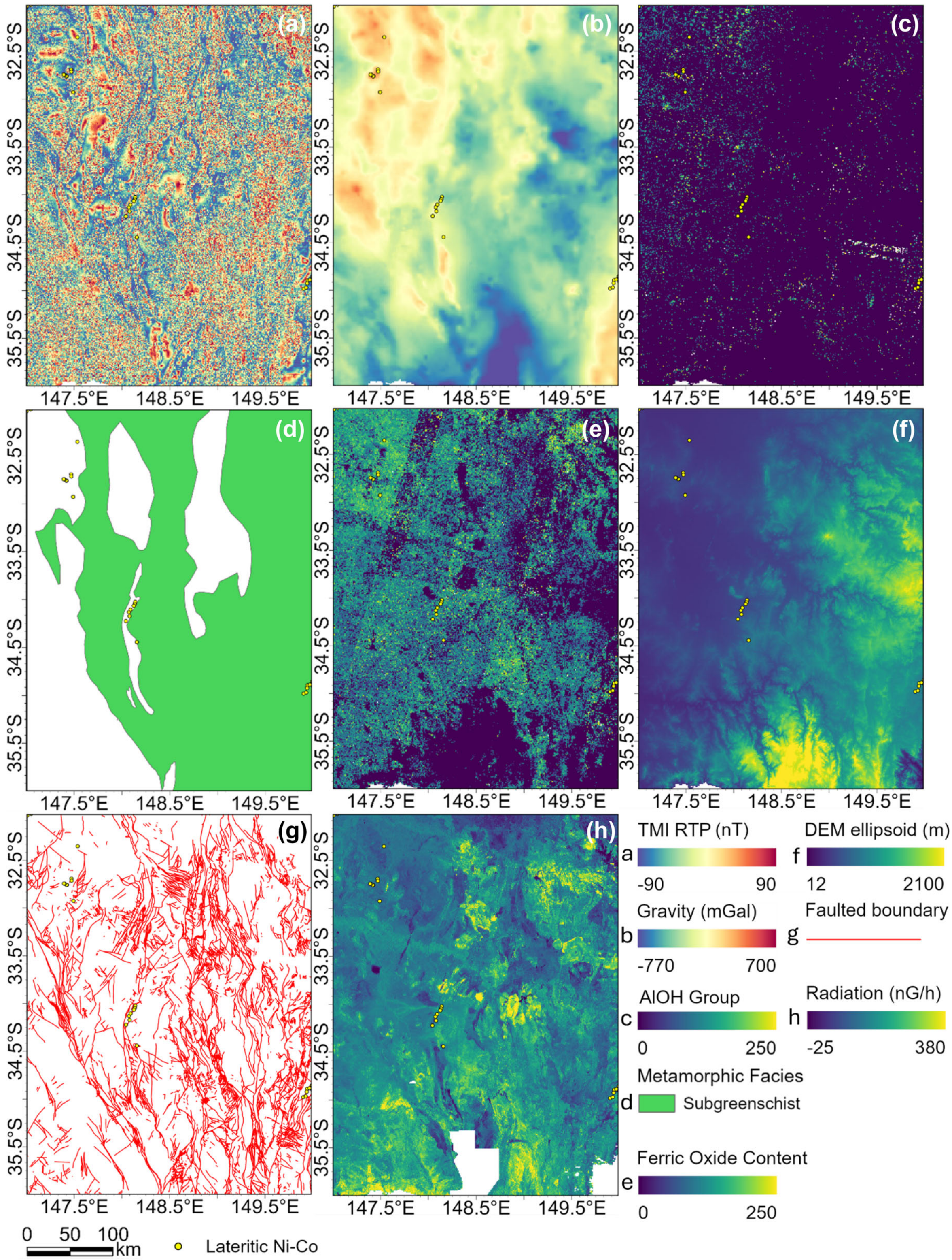
Figure 10. Feature map of the Lachlan Orogen showing rock units, overlayed by high-probability zones.

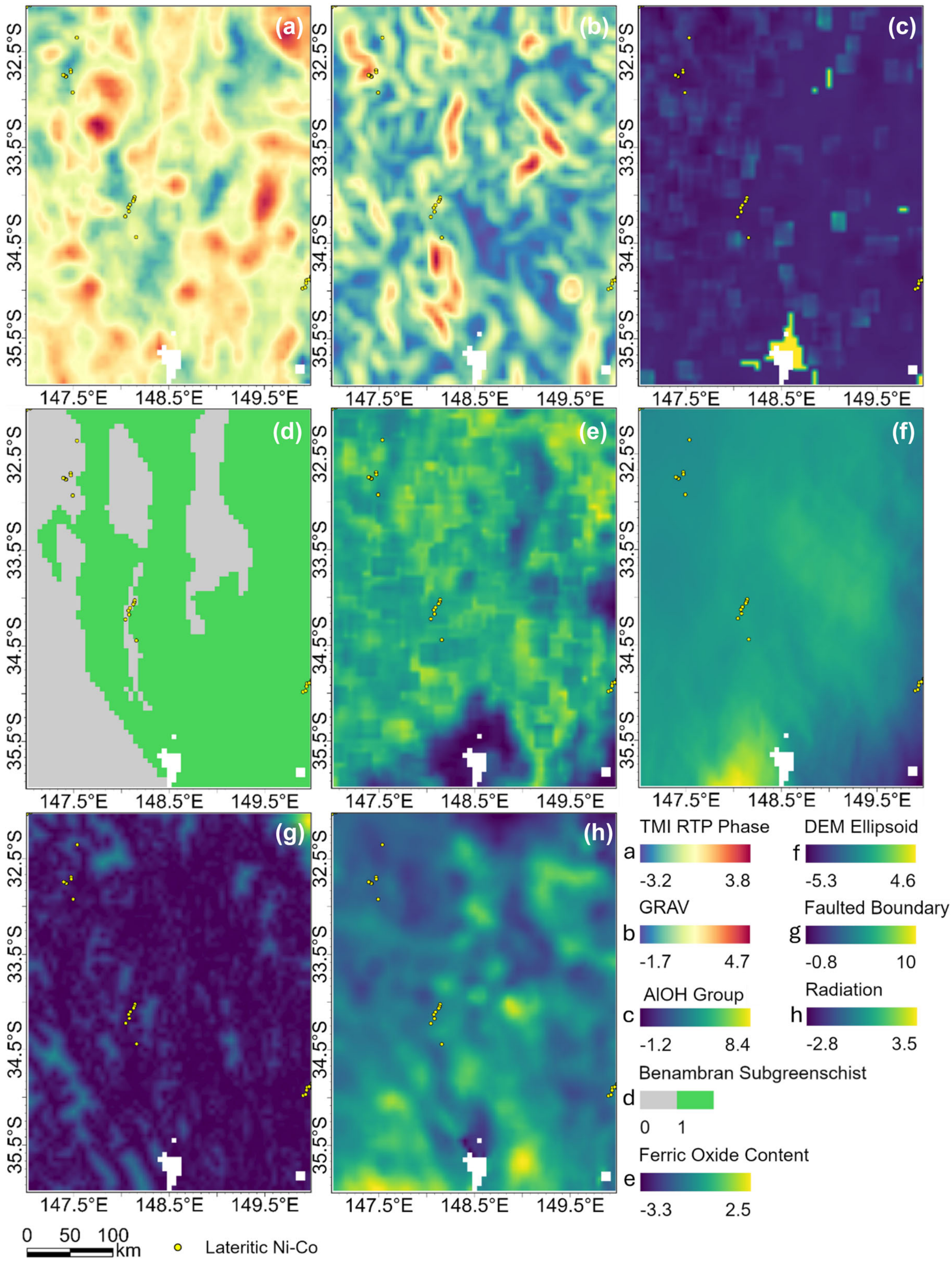
Figure 11. Some of the most important exploration data layers used in this study: (**a**) total magnetic intensity reduce-to-pole phase; (**b**) gravity; (**c**) AlOH group content; (**d**) subgreenschist in Benambran metamorphic facies; (**e**) ferric oxide content; (**f**) ellipsoid DEM; (**g**) Lachlan Orogen faulted boundaries; and (**h**) terrestrial radiation dose rate. ▶

Lachlan Orogen, which has undergone plate convergence, accretion, crustal extension, magmatism, and complex Cenozoic landform evolution. To validate the results, Model 2 which was our best model based on the P–A plots (Fig. 9b), was superimposed on the mapped distribution of mafic–ultramafic intrusions and serpentinite belts, based on prior knowledge and the mineral system model that identifies these lithologies as source rocks for lateritic Ni–Co in the Lachlan Orogen (Downes, 2021; Farahbakhsh et al., 2023). A strong correlation exists between these mafic–ultramafic intrusions, serpentinite basement rocks, and the prospective areas highlighted in Model 2. The Girilambone Group sandstone also shows a spatial relationship with known mineral occurrences, likely due to its interaction with serpentinite or intrusion by mafic–ultramafic rocks.

Lithology data features did not rank highly in the models, likely due to challenges in incorporating binary data into the SMOTE–GAN algorithm and the limited scale and distribution of ultramafic rock and serpentinite polygons. High-ranking geophysical feature layers, such as DEM, remote sensing, and radiometric data, primarily map surficial and modern environmental elements. These layers can serve as proxies for laterite preservation and weathering/mineralization zones but are unable to trace prehistoric preservation processes due to variability since the Mesozoic and Cenozoic eras. We identified areas with high mineralization potential by overlapping complementary supergene-related signals with hypogene signals. Future studies would benefit from more sophisticated analysis of radiometric data integrated with regional geochemistry and multispectral remote sensing imagery to better understand landform evolution and the potential preservation of target deposits (see Barbosa et al., 2013; Iza et al., 2018; Motta & Junior, 2016; Wilford, 2013).

According to Table 8, magnetic and gravity data layers were prominent due to the higher magnetic susceptibilities and bulk densities of mafic–ultramafic rocks and serpentinites. The magnetic intensity feature (TMI RTP Phase) in Figure 11a revealed detailed stratigraphy and structural frameworks of the Lachlan Orogen basement rocks. Serpentinite belts are shown as narrow formations trending north–northwest to south–southeast, characterized by elevated magnetics, while high-amplitude magnetic anomalies mark mafic–ultramafic intrusions. Structural discontinuities along these formations may indicate major faults and fracture systems, aligning with known mineral occurrences. The mean TMI RTP phase map in Figure 12a provided a


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
32.5°S
33.5°S
34.5°S
35.5°S
147.5°E
148.5°E
149.5°E
TMI RTP (nT)
a
-90 90
Gravity (mGal)
b
-770 700
AlOH Group
c
0 250
Metamorphic Facies
d
Subgreenschist
Ferric Oxide Content
e
0 250
DEM ellipsoid (m)
f
12 2100
Faulted boundary
g
Radiation (nG/h)
h
-25 380
0 50 100
km
Lateritic Ni-Co


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
32.5°S
33.5°S
34.5°S
35.5°S
147.5°E
148.5°E
149.5°E
TMI RTP Phase
a
-3.2 3.8
GRAV
b
-1.7 4.7
AlOH Group
c
-1.2 8.4
Benambran Subgreenschist
d
0 1
Ferric Oxide Content
e
-3.3 2.5
DEM Ellipsoid
f
-5.3 4.6
Faulted Boundary
g
-0.8 10
Radiation
h
-2.8 3.5
0 50 100
km
Lateritic Ni-Co

◀**Figure 12.** Some of the most important features extracted from data layers: (**a**) mean of total magnetic intensity reduce-to-pole phase; (**b**) standard deviation of gravity; (**c**) correlation of AlOH group content; (**d**) subgreenschist in Benambran metamorphic facies; (**e**) dissimilarity of ferric oxide content; (**f**) horizontal derivative of ellipsoid DEM along the *x*-axis; (**g**) Lachlan Orogen faulted boundaries; and (**h**) mean of radiation dose rate.

broader view, averaging TMI values and showing that lower values correlate with known mineral occurrences, which are often located near magnetic highs linked to source rock or intrusions. The gravity feature (Gravity CSCBA) in Figure 11b highlighted significant density contrasts, with known mineral occurrences aligning along the edges of these contrasts in basement rocks, corresponding to serpentinite belts and mafic–ultramafic intrusions.

In the western study area, elevated gravity responses may indicate higher-grade metasedimentary rocks of the Girrilambone Group. The gravity map showed a high-density signature in the top-left, matching areas of medium mineralization probability. A high-prospect zone is aligned with a high-density signature near an isolated mineral occurrence. The gravity standard deviation map in Figure 12b captured the NNW–SSE trend of ultramafics, correlating strongly with Ni–Co occurrences. Linear high-density areas may indicate underlying structural pathways. The AlOH group content in Figure 11c mapped clay and mica distribution, with high clay signatures correlating to medium-high-probability zones. The AlOH correlation map in Figure 12c provided more detail in the western region, where high concentrations of AlOH minerals are located near Ni–Co clusters. The subgreenschist of Benambran metamorphic facies in Figures 11d and 12d illustrates the distribution of basement rocks metamorphosed during the Early Silurian Benambran Orogeny, which aligns with known Ni–Co occurrences.

The ferric oxide mineral group in Figures 11e and 12e mapped the distribution of hematite, goethite, and jarosite, which correlate with Ni–Co occurrences, although the patterns were broad and less defined. The DEM in Figure 11f reflects elevation variations, with lower relief areas correlating with preserved lateritic weathering profiles. The horizontal derivative of the DEM mapped highlights elevation gradients, with Ni–Co occurrences typically found in low-relief areas. The Bounding faults

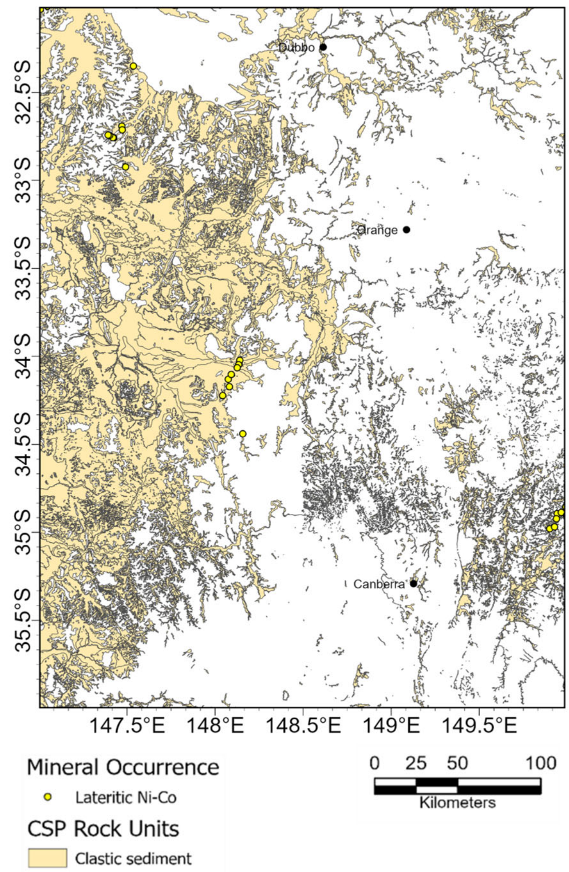


Figure 13. Cenozoic sedimentary province, displaying clastic sediment rock unit polygons. Mineral occurrences are marked with yellow circles.

in Figures 11g and 12g depict major faulted contacts, aligning with serpentinite belts and mafic–ultramafic intrusions, indicating deeper, primary pathway structures for parent magmas. The radiation dose rate in Figures 11h and 12h showed the distribution of radiogenic elements, with low radiation signatures correlating to known Ni–Co occurrences, although some high radiation dose patterns mirrored the trends of serpentinite belts.

Figure 13 shows the Cenozoic sedimentary province, mapping younger cover sediments that run parallel to known serpentinite belts and mafic–ultramafic intrusions. Known Ni–Co occurrences are strongly associated with the CSP edges, with the algorithm prioritizing surficial/regolith features over basement signatures. Lastly, the box and whisker plots in Figure 14 illustrate a similar distribution of key features between the real and synthetic positive samples, with overlapping interquartile ranges and median values. However, median values may not fully capture the true data distribution and were primarily used for ease of calculation and outlier management.

Hypogene (Primary) and Supergene (Secondary) Models Comparison

The two hybrid models created in this study were based on high-importance features and align with the mineral system model to differentiate between primary and secondary mineralization in laterites. The outputs largely mirrored the feature importance ranking from the initial full dataset model (Model 1), with secondary mineralization features dominating over primary ones. This indicates that the algorithm prioritizes features with greater surface reach over those related to deeper sources. As a result, the two models highlight different prospective areas based on their varying spatial correlations with mineral occurrences.

The hypogene model pinpointed high-probability regions in the southwestern part of the study area with reduced noise, likely due to the influence of geological data (e.g., ultramafic distribution, geological boundaries, gravity anomalies), which helped constrain the prospective zones. In contrast,

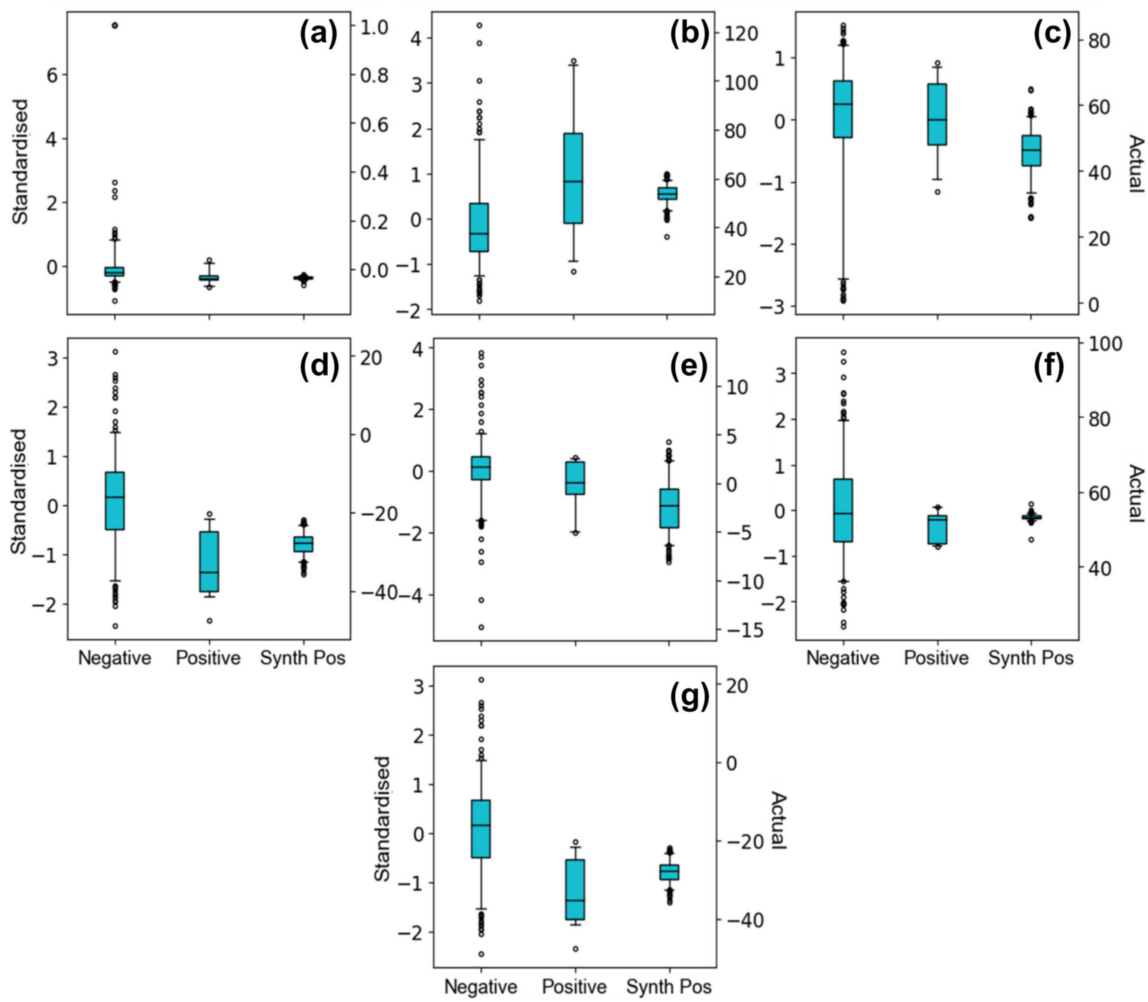


Figure 14. Box and whisker plots for the important features: (**a**) mean of total magnetic intensity reduce-to-pole phase; (**b**) correlation of AlOH group content; (**c**) standard deviation of gravity; (**d**) dissimilarity of ferric oxide content; (**e**) Lachlan Orogen faulted boundaries; (**f**) horizontal derivative of ellipsoid DEM along the *x*-axis; (**g**) mean of radiation dose rate, displaying the common distribution of interquartile ranges and median values for the negative, positive, and synthetic positive samples.

the supergene model placed more emphasis on surface and regolith data (e.g., the first vertical derivative of TMI, remote sensing, radiation), indicating zones of weathering and lateritization. In the full data model (Model 1), supergene-related features were given higher priority, causing prospective zones to lean toward surface-level correlations while also factoring in proximity to hypogene features. The highly prospective areas identified by Models 1 and 2 aligned with the supergene model but exhibited less noise, likely due to the geological constraints from the hypogene model.

Exploration Targets

The models used to generate the lateritic Ni–Co prospectivity maps and exploration targets yielded promising results, offering potential guidance for exploration strategies aimed at discovering additional critical mineral resources in the Lachlan Orogen. These models identified multiple targets, including well-explored areas with significant mineral resources, such as Thuddungra and Syerston, as well as underexplored regions like Byng and Bungonia. Low-cost exploration techniques, such as prospecting and orientation sampling, could be employed in these greenfield areas to quickly determine the presence of indicator source rocks, deep chemical weathering, or shallow cover units (such as Cenozoic sediments, basaltic lavas, or transported regolith) that may conceal prospective source rocks and/or laterite profiles.

The prospectivity maps revealed moderate to high likelihood of additional lateritic Ni–Co mineralization in areas surrounding and extending from known mineral occurrence clusters (Figs. 8, 15). Key exploration targets include:

- Two subparallel linear/curvilinear belts, each spanning 200–300 km, trending from NNW/SSE to NNE/SSW on the western side of the study area (Coolac–Thuddungra serpentinite belts), with one belt containing a large cluster of known lateritic Ni–Co occurrences.
- A large NNW/SSE-oriented lobate high-heat feature (30 km by 70 km) located in the northwest corner of the study area (Syerston–Homeville–Nyngan mafic–ultramafic intrusion field), which contains numerous known lateritic Ni–Co occurrences and extends northwestward into the Nyngan district.

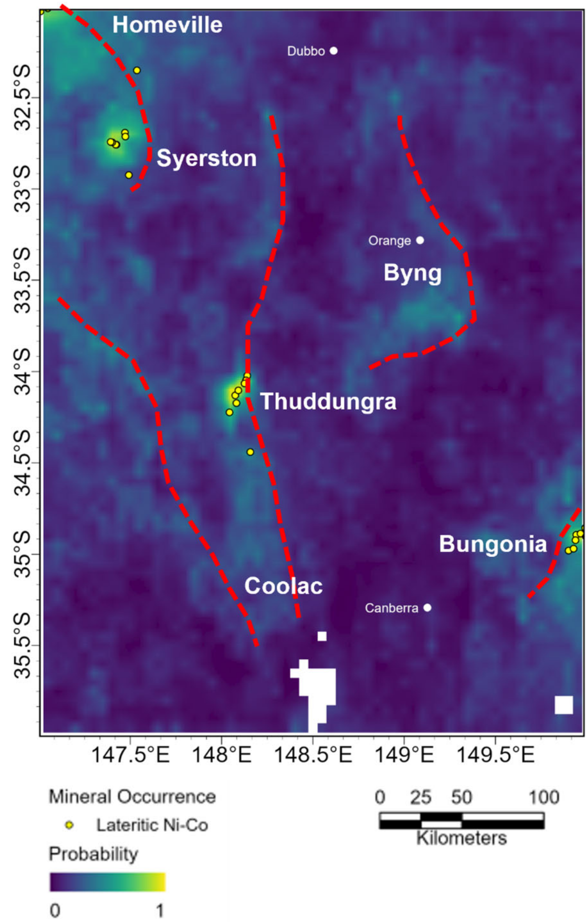


Figure 15. Prospectivity map of lateritic Ni–Co mineralization based on all features (Model 1), highlighting the identified exploration targets (more details in Table 1). The heat bar indicates areas with moderate to high potential for mineralization, ranging from blue to yellow. Gaps in the prospectivity map (at the bottom) result from missing values in the radiometric data layers and do not necessarily indicate a lack of mineralization.

- An arcuate linear-lobate feature, approximately 150 km long, located on the eastern side of the study area with few known lateritic Ni–Co occurrences. This feature is likely associated with the shoshonitic ultramafic lavas of the Byng Volcanics and may represent an underexplored zone for lateritic mineralization, potentially preserved beneath Cenozoic lava flows.
- A large mottled lobate feature on the southeastern edge of the study area containing a cluster of exotic supergene Co occurrences in Cenozoic manganiferous grits and sandstone (Bungonia). The Co source is unknown but is likely in basement rocks beneath the Cenozoic

cover, making this an intriguing target for exploration.

Effectiveness of Machine Learning Workflow

The prospectivity maps identified potential undiscovered lateritic Ni–Co deposits in underexplored regions. The ML workflow produced models with acceptable accuracy (0.75) and promising P–A plots (Fig. 9) despite integrating a large database with only a small number of mineral occurrences. High-importance features successfully predicted 6 out of 8 mineral occurrences. The workflow is flexible, allowing for adding or removing datasets based on new insights, and holds the potential for improved accuracy through hyperparameter tuning and further data refinement. Positive and unlabeled learning was applied to train a binary model, addressing the lack of negative training samples by labeling unknown samples. SMOTE–GAN was used to generate synthetic positive samples, mitigating class imbalance and increasing the minority class size through nearest-neighbor interpolation and generation–discrimination processes. Random Forest served as the primary classifier to evaluate the model and predict prospective areas.

Due to the unavailability of real negative samples, performance metrics such as the F1-score were calculated based on the combination of real and synthetic data, while accuracy can be solely calculated based on real positive samples. As this is a preliminary prospectivity map, future studies should include metrics with real negative samples to improve the results. As shown in Table 7, model accuracies combining real and synthetic data ranged 0.944–0.994, with Models 4 and 5 slightly lower. High F1-scores (0.958–0.992) indicate consistent accuracy in predicting the positive class, with few false positives or negatives. However, class imbalance remains a concern, which can only be fully addressed with real negative samples.

Nonetheless, the P–A plots (Fig. 9) of all models show promising results, revealing that a significant number of known mineral occurrences are located within a small portion of the total study area, corresponding to high likelihood values. This highlights that a large number of features is not always required for an accurate prospectivity model; rather, selecting the right features can produce strong results. Boxplots for high-importance features in Figure 14 show that the interquartile ranges of positive and synthetic positive samples have similar distributions and medians, which improves model accuracy and helps distinguish positive from negative feature values. The final feature importance list represents the average of multiple random samplings. However, the mode calculation of binary categorical data may have reduced their importance in the SMOTE–GAN algorithm, affecting accuracy.

In the investigation of hypogene and supergene models, the ML algorithm successfully distinguished between these mineralization stages, emphasizing supergene features in the combined model. However, incorporating mineral system knowledge into two hybrid models did not affect accuracy, likely because the same high-ranking features dominated feature importance. The limitations arose because the initial data-driven models relied on calculating spatial correlations to identify relationships between the input data and positive samples, whereas understanding the mineral system requires a more comprehensive hybrid approach. One challenge was differentiating between primary and secondary mineral system processes; for example, a deep-reaching gravity signal should correlate more with the source material than with high-frequency surficial deposits. This led to questions about the inclusion of gravity maps in the final prospectivity map for lateritic Ni–Co, as source material is necessary for secondary mineralization but does not guarantee its occurrence. In contrast, the presence of lateritic Ni–Co deposits indicates both underlying source material and secondary mineralization. Thus, a combination of primary and secondary mineral system features was necessary.

Data selection challenges included binary categorical data, which are incompatible with the SMOTE–GAN algorithm. The mode of binary numbers was used to produce numerical inputs, but this may have affected accuracy rankings by misrepresenting true binary values. While an alternative technique, the logistic transformation function, could have been used to generate binary variables, it was deemed unnecessary since geophysical data provided similar signatures to categorical data and were more useful in this study. Common sense and expert knowledge were applied when interpreting results, selecting data, and adjusting the workflow to mitigate the risk of poor data leading to poor outputs. For instance, investigations were made into why expected features, such as ultramafic rocks, did not rank highly. Overall, the ML workflow has successfully identified prospective areas for further investigation, supporting additional geophysical and geochemical data collection for future refinement.

Additional Data Layers

Surface geochemistry (including stream sediments, soils, and rocks) and drill hole geochemistry (e.g., rotary air blast, reverse circulation, and diamond drilling) can provide valuable data layers if there is adequate sample coverage and density within the study area. These methods offer direct detection and measurement of the grade and distribution of Ni, Co, and related metals in the lateritic weathering profile overlying the source rocks. Regolith mapping helps identify and characterize different types of weathered surface materials that cover the bedrock and any associated mineral deposits. Regional regolith mapping is particularly useful in prospectivity analysis, as different regolith components have distinct physical and chemical properties that reflect varying degrees and styles of weathering, as well as the concentration and distribution of Ni and Co within the lateritic profile. Regolith mapping can pinpoint the location, thickness, and quality of the preserved lateritic Ni–Co zone, distinguishing it from other regolith types or bedrock. Combining geochemical and regolith data layers would be a powerful tool for training ML algorithms to directly detect exposed or near-surface lateritic Ni–Co mineralization in the study area (González-Álvarez et al., 2016; González-Álvarez et al., 2022; Farahbakhsh et al., 2023).

CONCLUSIONS

This paper introduces a novel exploration method that employs ML algorithms on multidimensional data layers to generate prospectivity maps, addressing a knowledge gap in identifying additional critical mineral resources associated with lateritic Ni–Co deposits in the Lachlan Orogen of NSW. The ML-based workflow seamlessly integrates publicly available geological and geophysical data layers into a data-driven model and then selectively incorporates elements of the lateritic Ni–Co mineral system into hybrid models to analyze the diagnostic feature weightings. The study employed data augmentation techniques to address common ML challenges, such as transforming geological and geophysical data into training proxy layers using mathematical equations and GLCM textures. It also addressed issues like negative training samples through positive and unlabeled learning, tackled class imbalances with SMOTE–GAN, and produced a high-accuracy model using the Random Forest classifier.

Incorporating knowledge of the lateritic Ni–Co mineral system involved creating two hybrid models to gain deeper insights into the algorithm's outputs. The hypogene (primary) and supergene (secondary) features of the mineral systems model generated prospectivity maps with similar probability patterns. However, the hypogene map appeared less noisy and better constrained, while the supergene map displayed more dispersed signatures related to surface processes. These maps closely resemble the combined data maps, revealing that the ML algorithm favors features representing the supergene mineral system. The prospectivity maps for lateritic Ni–Co deposits show a strong spatial correlation between high probabilities and known mineral occurrences, predicting extensions of these occurrences and identifying potential new greenfield exploration areas. The models highlight multiple targets, including well-explored areas with significant resources, such as Thuddungra and Syerston, and potentially new, underexplored areas like Byng and Bungonia.

While these prospective areas offer promising exploration opportunities, further refinements to the workflow and outputs are recommended. These include adjusting hyperparameters to improve model accuracy and incorporating additional data layers, such as surface (stream sediment, soil, and rock chip) and drill hole geochemistry for direct Ni–Co target detection. The integration of regional (stream sediment) and province-scale regolith maps would also help differentiate and characterize the surficial geology of the study area. Ground-truthing the identified targets could be done cost-effectively through orientation prospecting and sampling.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Geological Survey of New South Wales and Geoscience Australia for making a variety of exploration datasets available to the public. EF was supported by the Australian Research Council grant

LP210100173. We also thank the anonymous reviewers for their invaluable feedback, which significantly enhanced the quality and rigor of our paper.

FUNDING

Open Access funding enabled and organized by CAUL and its Member Institutions.

DATA AVAILABILITY

The workflow repository with the related Python notebooks and datasets can be found here: https://doi.org/10.5281/zenodo.10681931.

DECLARATIONS

Conflict of 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.

OPEN ACCESS

REFERENCES

Alwosheel, A., Cranenburgh, S. V., & Chorus, C. G. (2018). Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of Choice Modelling, 28*, 167–182.

Anand, R. R. (2005). Weathering history, landscape evolution and implications for exploration. In R. R. Anand & P. De Broekert (Eds.), *Regolith landscape evolution across Australia: A compilation of regolith landscape case studies with regolith landscape evolution models* (pp. 2–40). CRC LEME Monograph.

Geoscience Australia. (2021). 2.10 Surficial mineral systems. https://www.ga.gov.au/data-pubs/data-and-publications-search/publications/critical-commodities-for-a-high-tech-world/surficial-mineral-systems.

Geoscience Australia. (2023). Critical minerals at geoscience Australia. https://www.ga.gov.au/scientific-topics/minerals/critical-minerals.

Barbosa, I. O., Pires, A. C. B., Lacerda, M. P. C., & Carmelo, A. C. (2013). Geology, airborne geophysics, geomorphology and soils in the individualization of the Niquelândia mafic-ultramafic complex, Goiás state, Brazil. *Revista Brasileira de Geofísica, 31*(3), 463–481.

Barnes, R., & Downes, P. (undated). Nickel exploration opportunities in New South Wales. https://www.dpi.nsw.gov.au/__data/assets/pdf_file/0011/316838/Nickel.pdf.

Bedell, R. L. (2004). Remote sensing in mineral exploration. *SEG Newsletter, 58*, 1 and 8–14. https://doi.org/10.5382/SEGnews.2004-58.fea.

Belousova, E. A., Graham, I. T., Glen, R. A., Griffin, W. L., & Martin, L. (2022). Zircons from the Wambidgee Serpentinite Belt, southern Lachlan Orogen: Evidence for oceanic crust at the Cambrian-Ordovician boundary. *Australian Journal of Earth Sciences, 69*(3), 406–418.

Brand, N. W., Butt, C. R. M., & Elias, M. (1998). Nickel laterites: classification and features. *AGSO Journal of Australian Geology and Geophysics, 17*(4), 81–88.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Brownlee, J. (2020). How to calculate precision, recall, and F-measure for imbalanced classification. Machine learning mastery, 1. https://machinelearningmastery.com/precision-recall-and-f-measure-for-imbalanced-classification/.

Butt, C. R. M. (2016). The development of regolith exploration geochemistry in the tropics and sub-tropics. *Ore Geology Reviews, 73*, 380–393.

Butt, C. R. M., & Chuzel, D. (2013). Nickel laterite ore deposits: weathered serpentinites. *Elements, 9*, 123–128.

Carranza, E. J. M., & Laborte, A. G. (2015). Random forest predictive modelling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Computers and Geosciences, 74*, 60–70.

Champion, D. C. (2016). Geodynamic synthesis of the Phanerozoic of eastern Australia (2nd edn.). Record 2016/07. Geoscience Australia, Canberra. https://doi.org/10.11636/Record.2016.007.

Chassé, M., Griffin, W. L., O'Reilly, S. Y., & Calas, G. (2017). Scandium speciation in a world-class lateritic deposit. *Geochemical Perspective Letters, 3*, 105–114.

Collins, W. J., Huang, H-Q., Bowden, P., & Kemp, A. I. S. (2019). Repeated S–I–A-type granite trilogy in the Lachlan Orogen and geochemical contrasts with A-type granites in Nigeria: implications for petrogenesis and tectonic discrimination. In: Janousek, V., Bonin, B., Collins, W.J., Farina, F., & Bowden, P. (Eds.), Post-Archean granitic rocks: Petrogenetic processes and tectonic environments (Vol. 491(1), pp. 53–76). Geological Society, London, Special Publications.

Crawford, A. J., Meffre, S., Squire, R. J., Barron, L. M., & Falloon, T. J. (2007). Middle and Late Ordovician magmatic evolution of the Macquarie Arc, Lachlan Orogen, New South Wales. *Australian Journal of Earth Sciences, 54*, 181–214.

Cudahy, T., Caccetta, M., Lau, I., Rodger, A., Laukamp, C., Ong, C., Chia, L., Collings, S., Rankine, T., Fraser, R., Woodcock, R., Vote, J., Warren, P., Thomas, M., Tyler, I., Mauger, A., Close, D., Jones, M., & Abrams, M. (2020). Satellite ASTER

Geoscience Map of Australia. CSIRO. Retrieved from https://data.csiro.au/collection/csiro:6182.

Dentith, M., & Mudge, S. Y. (2014). *Geophysics for the mineral exploration geoscientist*. Cambridge University Press. https://doi.org/10.1007/s00126-014-0557-9.

Downes, P. M. (2021). A mineral system model for scandium-enriched regolith-hosted deposits in NSW. *Geological Survey of New South Wales*, Report GS2021/0343. https://doi.org/10.13140/RG.2.2.13012.73603.

Eggleton, R. A. (ed.) (2001). The regolith glossary CRC for landscape evolution and mineral exploration. Floreat Park, Western Australia. http://hdl.handle.net/102.100.100/366737?index=1.

Elias, M. (2002). Nickel laterite deposits—geological overview, resources and exploitation. In D.R. Cooke and J. Pongratz (Eds)., Giant Ore Deposits: Characteristics, genesis and exploration, *CODES Special Publication 4*, Centre for Ore Deposit Research, University of Tasmania (pp. 205–220).

Elias, M. (2006). Lateritic nickel mineralization of the Yilgarn craton. In Barnes, S. J. (Ed.), Nickel deposits of the Yilgarn craton: Geology, geochemistry and geophysics applied to exploration (Vol. 13, pp. 195–210). Society of Economic Geology Special Publication.

Elliot, S. J., & Martin, A. R. (1991). The Owendale Alaskan-type intrusive complex and associated PGE mineralization, Fifield belt, New South Wales. In *6th International Platinum Symposium, Perth, July 1991*. Guidebook for the Pre-Symposium field excursion (pp. 12–24).

Farahbakhsh, E., Maughan, J., & Müller, R. D. (2023). Prospectivity modelling of critical mineral deposits using a generative adversarial network with oversampling and positive-unlabelled bagging. *Ore geology Reviews, 162*, 105665.

Fergusson, C. L. (2017). Mid to late Paleozoic shortening pulses in the Lachlan Orogen, southeastern Australia: a review. *Australian Journal of Earth Sciences, 64*, 1–39.

Figueroa, R. L., Zeng-Treitler, Q., Kandula, S., & Ngo, L. H. (2012). Predicting sample size required for classification performance. *BMC Medical Informatics and Decision Making, 12*, 1–10. https://doi.org/10.1186/1472-6947-12-8.

Fitzpatrick, K. R. (1968). The Mineral Industry of New South Wales No. 45 Nickel. Department of Mines, Geological Survey New South Wales, Sydney. https://search.geoscience.nsw.gov.au/api/digs/ocr/8bc157de-e79e-4cf1-8a9e-9e39238a74f0.

Ford, A., Peters, K. J., Partington, G. A., Blevin, P. L., Downes, P. M., Fitzherbert, J. A., & Greenfield, J. E. (2019). Translating expressions of intrusion-related mineral systems into mappable spatial proxies for mineral potential mapping: Case studies from the Southern New England Orogen, Australia. *Ore Geology Reviews, 111*, 1–18.

Foster, D. A., & Gray, D. R. (2000). Evolution and structure of the Lachlan fold belt (orogen) of eastern Australia. *Annual Review of Earth and Planetary Sciences, 28*, 47–80.

Freyssinet, Ph., Butt, C. R. M., Morris, R. C., & Piantone, P. (2005). Ore-forming processes related to lateritic weathering. In *Economic geology 100th anniversary volume* (pp. 681–722).

Gan, S., Alshahrani, M., & Liu, S. (2022). Positive-unlabeled learning for network link prediction. *Mathematics, 10*(18), 3345.

Gleeson, S. A., Butt, C. R. M., & Elias, M. (2003). Nickel laterites, a review. *Society of Economic Geologists Newsletter, 54*, 11–18.

Glen, R. A. (2005). The Tasmanides of eastern Australia. In A. P. M. Vaughan, P. T. Leat, & R. J. Pankhurst (Eds.), *Terrane processes at the margins of Gondwana* (Vol. 246, pp. 23–96). London: Special Publication of the Geological Society.

Glen, R. A., Crawford, A., & J., Cooke, D.R. (2007a). Tectonic setting of porphyry Cu–Au mineralisation in the Ordovician-Early Silurian Macquarie Arc, Eastern Lachlan Orogen, New South Wales. *Australian Journal of Earth Sciences, 54*(2–3), 465–479.

Glen, R. A., Meffre, S., & Scott, R. J. (2007b). Benambran orogeny in the Eastern Lachlan Orogen, Australia. *Australian Journal of Earth Sciences, 54*, 385–415.

Golightly, J.P. (2010). Progress in understanding the evolution of nickel laterites. In: Goldfarb R.J., Marsh. E.E., Monecke, T. (Eds.), *The Challenge of Finding New Mineral Resources: Global Metallogeny, Innovative Exploration, and New Discoveries* (Vol. 15, pp. 451–475). Society of Economic Geologists Special Publication.

González-Álvarez, I., Boni, M., & Anand, R. R. (2016). Mineral exploration in regolith-dominated terrains: Global considerations and challenges. *Ore Geology Reviews, 73*, 375–379.

González-Álvarez, I., Krapf, C., Albrecht, T., Kelka, U., Martínez, C., Pawley, M., Ibrahimi, T., Irvine, J., Petts, A., Gum, J., & Klump, J. (2022). Connecting the cover and the fabric of the basement in the central Gawler Craton, South Australia. *MESA Journal, 96*, 4–21.

Granek, J. (2016). Application of machine learning algorithms to mineral prospectivity mapping. Ph.D. Thesis, University of British Columbia. https://open.library.ubc.ca/collections/ubctheses/24/items/1.0340340.

Gray, D. R., & Foster, D. A. (2004). Tectonic evolution of the Lachlan Orogen, southeast Australia: historical review, data synthesis and modern Perspectives. *Australian Journal of Earth Sciences, 51*, 773–817.

Hall-Beyer, M. (2017). GLCM texture: A tutorial v. 3.0 March 2017. *Creative Commons Attribution-NonCommercial 4.0 International License*. https://doi.org/10.13140/RG.2.2.12424.21767.

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Texture features for image classification. *IEEE Transactions on Systems, Man and Cybernetics, SMC-3*(6), 610–621.

Herzberg, C., Vidito, C., & Starkey, N. A. (2016). Nickel–cobalt contents of olivine record origins of mantle peridotite and related rocks. *American Mineralogist, 101*, 1952–1966.

Hewson, R., Robson, D., Mauger, A., Cudahy, T., Thomase, M., & Jones, S. (2015). Using the Geoscience Australia-CSIRO ASTER maps and airborne geophysics to explore Australian geoscience. *Journal of Spatial Science, 60*(2), 1–25.

Hough, M. A., Bierlein, F. P., & Wilde, A. R. (2007). A review of the metallogeny and tectonics of the Lachlan Orogen. *Mineralium Deposita, 42*, 435–448.

Huston, D. L., Blewett, R. S., & Champion, D. C. (2012). Australia through time: a summary of its tectonic and metallogenic evolution. *Episodes, 35*(1), 23–43.

Huston, D. L., Champion, D. C., Mernagh, T. P., Downes, P. M., Jones, P., Carr, G., Forster, D., & David, V. (2016). Metallogenesis and geodynamics of the Lachlan Orogen: New (and old) insights from spatial and temporal variations in lead isotopes. *Ore Geology Reviews, 76*, 257–267.

Isles, D. J. & Rankin, L. R. (2013). Geological Interpretation of Aeromagnetic Data. The Australian Society of Exploration Geophysicists, Victora (ASEG eBook). https://www.aseg.org.au/publications/geological-interpretation-aeromagnetic-data.

Iza, E. R. H. F., Horbe, A. M. C., Castro, C. C., & Herrera, I. L. I. E. (2018). Integration of geochemical and geophysical data to characterize and map lateritic regolith: An example in the Brazilian Amazon. *Geochemistry, Geophysics, Geosystems, 19*, 3254–3271.

Jaskie, K., & Spanias, A. (2019). Positive and unlabeled learning algorithms and applications: A survey. In *Proceedings IEEE 10th international conference on information, intelligence, systems and applications (IISA)* (pp. 1–8). https://doi.org/10.1109/IISA.2019.8900698.

Keays, R. R., Holwell, D. A., & Prichard, H. M. (2022). Platinum mineralisation in the Owendale Uralian-Alaskan-type complex, New South Wales, Australia: The effects of serpentinization on Cu-PGE-Ni sulphides. *Ore Geology Reviews, 130*, 103928.

Koehrsen, W. (2018b). A conceptual explanation of Bayesian hyperparameter optimization for machine learning. Retrieved from https://towardsdatascience.com/a-conceptual-explanation-of-bayesian-model-based-hyperparameter-optimization-for-machine-learning-b8172278050f.

Koehrsen, W. (2018a). Hyperparameter Tuning the random forest in python. Retrieved from https://towardsdatascience.com/hyperparameter-tuning-the-random-forest-in-python-using-scikit-learn-28d2aa77dd74.

Kreuzer, O. P., Miller, A. V. M., Peters, K. J., Payne, C., Wildman, C., Partington, G. A., Puccioni, E., McMahon, M. E., & Etheridge, M. A. (2015). Comparing prospectivity modelling results and past exploration data: A case study of porphyry Cu–Au mineral systems in the Macquarie Arc, Lachlan Fold Belt, New South Wales. *Ore Geology Reviews, 71*, 516–544.

Marsh, E., Anderson, E., & Gray, F. (2013). Nickel-cobalt laterites—A deposit model, Chapter H of Mineral deposit models for resource assessment. U.S. Geological Survey Scientific Investigations Report 2010-5070-H. https://doi.org/10.3133/sir20105070H.

McCuaig, T. C., Beresford, S., & Hronsky, J. (2010). Translating the mineral systems approach into an effective exploration targeting system. *Ore Geology Reviews, 38*, 128–138.

Minty, B. R. S. (1997). The fundamentals of airborne gamma-ray spectrometry. *AGSO Journal of Australian Geology and Geophysics, 17*(2), 39–50.

Minty, B., Franklin, R., Milligan, P., Richardson, M., & Wilford, J. (2009). The radiometric map of Australia. *Exploration Geophysics, 40*, 325–333.

Mordelet, F., & Vert, J. P. (2014). A bagging svm to learn from positive and unlabeled examples. *Pattern Recognition Letters, 37*, 201–209.

Motta, J. B., & Junior, I. R. F. (2016). A mineral potential mapping approach for supergene nickel deposits in southwestern São Francisco Craton, Brazil. *Brazilian Journal of Geology, 46*(2), 261–273.

Mudd, G. M., & Jowitt, S. M. (2022). The new century for nickel resources, reserves, and mining: Reassessing the sustainability of the Devil's metal. *Economic Geology, 117*(8), 249–279.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*(10), 2825–2830.

Porwal, A. K., & Kreuzer, O. P. (2010). Introduction to the special issue: Mineral prospectivity analysis and quantitative resource estimation. *Ore Geology Reviews, 38*, 121–127.

Rad, P., Roopaei, M., Beebe, N., Shadaram, M., Au, Y. A. (2018). AI thinking for cloud education platform with personalized learning. In *Proceedings of the 51st Hawaii international conference on system sciences* (pp. 3–12). https://doi.org/10.24251/HICSS.2018.003.

Retallack, G. J. (2010). Lateritization and bauxitization events. *Economic Geology, 105*, 655–667.

Rosenbaum, G. (2018). The tasmanides: Phanerozoic tectonic evolution of Eastern Australia. *Annual Review of Earth and Planetary Sciences, 46*, 291–325.

Scikit-Image Team. (2023). Skimage.feature. https://scikit-image.org/docs/stable/api/skimage.feature.html#skimage.feature.graycoprops.

Seone, J. C. S., Castro, N. A., Osako, L. S., & Baars, F. J. (2009). Multispectral imagery applied to nickel laterite exploration: The Conceição do Araguaia discovery. *Reviews in Economic Geology, 16*, 109–122.

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. *Remote Sensing, 12*(8), 1261.

Shirmard, H., Farahbakhsh, E., Heidari, E., Beiranvand Pour, A., Pradhan, B., Müller, D., & Chandra, R. (2022a). A comparative study of convolutional neural networks and conventional machine learning models for lithological mapping using remote sensing data. *Remote Sensing, 14*(4), 819.

Shirmard, H., Farahbakhsh, E., Müller, R. D., & Chandra, R. (2022b). A review of machine learning in processing remote sensing data for mineral exploration. *Remote Sensing of Environment, 268*, 112750.

Sun, T., Li, H., Wu, K., Chen, F., Zhu, Z., & Hu, Z. (2020). Data-driven predictive modelling of mineral prospectivity using machine learning and deep learning methods: A case study from Southern Jiangxi Province, China. *Minerals, 10*(2), 1–28.

Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., & Yu, T. (2014). Scikit-image: image processing in Python. *PeerJ, 2*, e453.

Webb, K. (2014). The age and tectonometallogenic setting of the Fifield ''Alaskan-type'' intrusive suite, Lachlan Orogen, NSW (BSc. Honours thesis). University of Wollongong, Wollongong, NSW. http://ro.uow.edu.au/thsci/79.

Wilford, J. R. (2013). New regolith mapping approaches for old Australian landscapes. Ph.D. Thesis, University of Adelaide, South Australia. https://hdl.handle.net/2440/85930.

Wilford, J. (2012). A weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. *Geoderma, 183–184*, 124–142.

Wilford, J. R., Bierwirth, P. N., & Craig, M. A. (1997). Application of airborne gamma-ray spectrometry in soil/regolith mapping and applied geomorphology. *AGSO Journal of Australian Geology and Geophysics, 17*(2), 201–216.

Yousefi, M., & Carranza, E. J. M. (2015). Prediction–area (P–A) plot and C-A fractal analysis to classify and evaluate evidential maps for mineral prospectivity modeling. *Computers and Geosciences, 79*, 69–81.

Yousefi, M., Carranza, E. J. M., Kreuzer, O. P., Nykanen, V., Hronsky, J. M. A., & Mihalasky, M. J. (2021). Data analysis methods for prospectivity modelling as applied to mineral exploration targeting: State-of-the-art and outlook. *Journal of Geochemical Exploration, 229*, 106839.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014. ECCV 2014. Lecture notes in computer science.* (Vol. 8689). Springer. https://doi.org/10.1007/978-3-319-10590-1_53.

Zhou, K., Sun, T., Liu, Y., Feng, M., Tang, J., Mao, L., Pu, W., & Huang, J. (2023). Prospectivity mapping of tungsten mineralization in southern Jiangxi Province using few-shot learning. *Minerals, 13*(669), 1–21.

Zuo, R., & Wang, Z. (2020). Effects of random negative training samples on mineral prospectivity mapping. *Natural Resources Research, 29*, 3443–3455.