



INDIA AI
A MEITY INITIATIVE

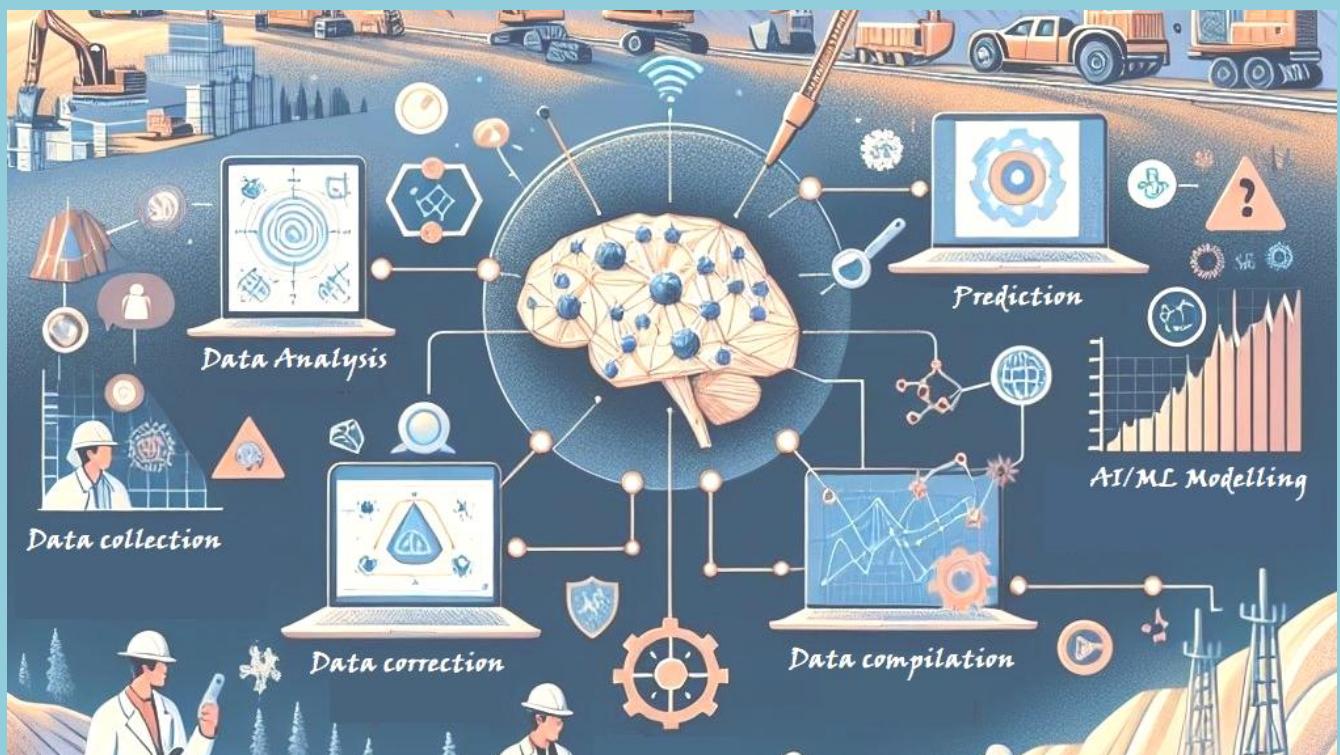
INDIA AI
IN PARTNERSHIP WITH
GEOLOGICAL SURVEY OF INDIA



GT51
YEARS OF GLORIOUS SERVICE

HACKATHON, 2025

**MINERAL TARGETING USING
ARTIFICIAL INTELLIGENCE**



REPORT SUBMITTED

BY

TEAM: GeoML

SOUMYA MITRA
SAPTARSHI MALICK
KSHOUNISH PATRA
SANTU BISWAS

GEOLOGICAL SURVEY OF INDIA

12-JUNE-2025

CONTENTS

1. Abstract.....	1
2. Introduction.....	2
2.1. Problem Statement:	2
2.2. Project Objectives:	2
2.3. Study Area:	2
2.4. Area of Interest:.....	3
3. Resources Utilized.....	4
3.1. Hardware:	4
3.2. Software:.....	4
3.3. Manpower:	5
4. Data Used in Detail.....	6
4.1 Category of Data.....	6
4.2. Data Acquisition and Initial Assessment:.....	6
4.2.1. Geological Data:	6
4.2.2. Geochemical Data:	7
4.2.3. Geophysical Data:	7
4.2.4. Mineral Exploration Data:.....	7
4.2.5. PGRS (Photogeology and Remote Sensing) Data:	7
4.2.6. Technical Reports:	8
5. Conceptual Genetic Models and Targeting Criteria	9
5.1. Overview:	9
5.2. Mineral System:	9
5.3. Geo-scientific Controls:.....	11
5.4. Definition of Mineral Targeting Criteria:	12
5.4.1. Geological concerns:	12
5.4.2 Geophysical data processing and outputs:	14
5.4.3. Geochemical data processing and outputs:	19
5.4.4 Remote Sensing data	25
5.4.5 Aerogeophysical Spectrometric Data.....	25
6. Methodology	26
6.1. Overall Workflow (Flowchart):	26
6.2. Data Processing and Curation:	27

6.2.1. Data Cleaning and Pre-processing	27
6.2.2. Spatial Data Handling and Harmonization	27
6.2.3. Outlier Detection and Handling	28
6.2.4 Raster Data Standardization: Min-Max Normalization	28
6.3. Feature Engineering and Derived Data Layers:	29
6.3.1. List of Derived Features:.....	29
6.4. Training Data Selection and Preparation:.....	30
6.4.1. Definition:.....	30
6.4.2. Handling Imbalance:	31
6.5. Machine Learning Model Development:.....	32
6.5.1. Problem Formulation:	32
6.5.2. Model Selection:	32
6.5.3. Training & Validation	33
6.5.4. Hyperparameter Tuning:	33
7. Results and Discussion.....	35
7.1. Model Performance and Evaluation:	35
7.1.1. Receiver Operating Characteristic (ROC) Curve Comparison	35
7.1.2. Precision-Recall Curve Comparison	36
7.1.3. Comparison of Key Metrics (Bar Plot).....	37
7.1.4. Confusion Matrices	38
7.2. Feature Importance and Layer Contribution Analysis:	40
7.2.1. Feature Importance Plots (for Tree-Based Models)	40
7.3. Predictive Maps:	42
7.3.1. Delineation of New Prospective Zones (Combined Map)	42
8. Conclusions and Recommendations.....	49
8.1. Summary of Key Findings:	49
8.2. Implications for Mineral Exploration:	51
8.3. Future Work and Enhancements:.....	51
8.3.1 Incorporating New and Higher-Resolution Data Types:	52
8.3.2. Testing Advanced ML Algorithms and Architectures:.....	52
9. References	53

ANNEXURE – I Prospectivity map showing proposed blocks

LIST OF FIGURES

<i>Figure 2.1: Geological map showing Mineralisation of the study area</i>	3
<i>Figure 5.1 Factor maps derived from geological layers. Details of these files added in Github.....</i>	13
<i>Figure 5.2a: Bouguer gravity anomaly map of the study area, 2b: RTP Map of the aero-magnetic data of the study area.....</i>	15
<i>Figure 5.3a: Upward Continuation map (5km) of the gravity data and, 3b: Upward Continuation map (5km) of the magnetic data.....</i>	16
<i>Figure 5.4a: Residual gravity map of the study area, 4b:Residual Magnetic map....</i>	16
<i>Figure 5.5a: Tilt derivative map of gravity data in the study area, 5b: 8Analytical signal map of aero-magnetic data in the study area.....</i>	17
<i>Figure 5.6: Radially averaged power spectrum of gravity data for depth estimation</i>	18
<i>Figure 5.7: Euler Solution map of gravity data, superimposed over bouguer anomaly map of the study area.....</i>	19
<i>Figure 5.8: Histogram of the selected geo-chemical elements.....</i>	21
<i>Figure 5.9: Correlation heatmap of the selected geo-chemical elements.....</i>	22
<i>Figure 5.10: PCA loadings heatmap (first 10) of the selected geo-chemical elements</i>	23
<i>Figure 5.11: Scree plot of the PCA.....</i>	23
<i>Figure 5.12: Normalized map of PCA1, PCA3 and PCA4.....</i>	25
<i>Figure 5.13: Normalized alteration maps.....</i>	25
<i>Figure 5.14: Normalized Dose TC and K map.....</i>	25
<i>Figure 6.1: Comprehensive workflow illustrating the stages of data processing, machine learning model development, and output generation for mineral prospectivity mapping.....</i>	26
<i>Figure 7.1: Comparative ROC curves for all evaluated machine learning models, demonstrating their ability to distinguish between mineralized and non-mineralized classes for prospectivity mapping.</i>	36
<i>Figure 7.2: Comparative Precision-Recall curves for all evaluated machine learning models, highlighting their performance in identifying the positive (mineralized) class, particularly relevant for imbalanced datasets.</i>	37
<i>Figure 7.3: Bar plot comparing key performance metrics (Accuracy, Precision, Recall, F1-Score, and AUC-ROC) for all trained machine learning models.</i>	38
<i>Figure 7.4: Individual confusion matrices for each machine learning model, detailing true positive, true negative, false positive, and false negative counts on the test set.</i>	40

<i>Figure 7.5: Bar plots illustrating the relative importance of each geoscientific feature as determined by the tree-based machine learning models (Random Forest, Decision Tree, LightGBM, and XGBoost).</i>	42
<i>Figure 7.6: Combined prospectivity map, highlighting high-confidence target areas for gold (Au) mineralization. These zones represent locations where all top three performing machine learning models predicted a prospectivity score greater than 0.85.</i>	44
<i>Figure 7.7: Combined prospectivity map, highlighting high-confidence target areas for copper (Cu) mineralization. These zones represent locations where all top three performing machine learning models predicted a prospectivity score greater than 0.60.</i>	45
<i>Figure 7.8: Combined prospectivity map, highlighting high-confidence target areas for iron - manganese (Fe-Mn) mineralization. These zones represent locations where all top three performing machine learning models predicted a prospectivity score greater than 0.70.</i>	46
<i>Figure 7.9: Combined prospectivity map, highlighting high-confidence target areas for nickel - PGE (Ni - PGE) mineralization. These zones represent locations where all top three performing machine learning models predicted a prospectivity score greater than 0.80.</i>	47
<i>Figure 7.10: Combined prospectivity map, highlighting high-confidence target areas for rare earth elements (REE) mineralization. These zones represent locations where all top three performing machine learning models predicted a prospectivity score greater than 0.85.</i>	48
<i>Figure 8.1 Prospectivity map showing proposed blocks for gold, copper, iron-manganese, Ni-PGE and REE</i>	50

AI/ML-Driven Mineral Prospectivity Mapping in Karnataka and Andhra Pradesh, India

1. Abstract

This project presents an advanced, data-driven approach to Mineral Prospectivity Mapping (MPM) within a 39,000 sq. km area across Karnataka and Andhra Pradesh, India. Addressing the challenge of identifying new, concealed, and deep-seated mineralized bodies, the initiative utilized cutting-edge AI/ML algorithms for comprehensive data cleaning, integration, modelling, and validation.

A multi-thematic geo-scientific dataset, encompassing detailed geological maps (25K and 50K scale), lineament features, National Geochemical Mapping (NGCM) data, aero geophysical, magnetic and spectrometric surveys, ground gravity data, mineral exploration blocks, and ASTER-derived mineral maps, was systematically processed and integrated. The methodology included rigorous data curation, handling of missing values, and Min-Max normalization of raster layers, followed by Principal Component Analysis (PCA) on geochemical data for dimensionality reduction and feature engineering.

Multiple machine learning algorithms (Logistic Regression, Decision Tree, Random Forest, LightGBM, XGBoost, SVC, KNN, ANN) were trained and comparatively evaluated using a single train-test split, emphasizing metrics critical for imbalanced data, such as AUC-ROC and F1-Score. The insights gained from feature importance analysis further refine understanding of key geological controls on mineralization, validating the project's robust methodology and its significance for mineral exploration in India. The robust machine learning approach successfully delineated significant new prospective zones for various commodities: 26 blocks (1504 Sq. Km) for Gold (Au), 23 blocks (832 Sq. Km) for Copper (Cu), 6 blocks (240 Sq. Km) for Ni-PGE, 8 blocks (1220 Sq. Km) for REE, and 22 blocks (1056 Sq. Km) for Fe-Mn.

The project's key findings are visualized through predictive maps, highlighting high-confidence targets derived from a consensus of top-performing models. This innovative application of AI/ML significantly de-risks early-stage exploration by prioritizing areas for follow-up ground surveys and potential drilling, thereby enhancing the efficiency and success rate of future mineral discovery efforts.

2. Introduction

The geological exploration in India is poised for a significant transformation with the integration of Artificial Intelligence and Machine Learning (AI/ML) technologies. Recognizing the unparalleled ability of AI/ML to process and integrate vast, complex, multi-parametric geoscientific datasets – encompassing geology, geophysics, geochemistry and remote sensing – a major initiative, Hackathon 2025 on "Mineral Targeting using Artificial Intelligence," has been launched by IndiaAI in partnership with the Geological Survey of India (GSI), targeting a 39,000 sq. km area within the Dharwar Craton, spanning Karnataka and Andhra Pradesh.

2.1. Problem Statement:

1. Identifying new potential areas for exploration of critical minerals like REE, Ni-PGE, and copper, as well as other commodities like diamond, iron, manganese, and gold within a pre-defined 39,000 sq. km area in the states of Karnataka and Andhra Pradesh, India.
2. Locating concealed & deep-seated mineralised bodies with depth modelling.
3. Developing AI/ ML algorithms for data cleaning, integration, modelling and validation.
4. Generating mineral predictive maps showing exploration targets visualised through maps, sections, etc.

2.2. Project Objectives:

This report details the strategy for identifying new potential areas for critical minerals such as Rare Earth Elements (REE), Nickel-Platinum Group Elements (Ni-PGE) and copper, as well as diamond, iron, manganese and gold, within the designated area. A key emphasis is given on locating concealed and deep-seated mineralized bodies through advanced depth modelling techniques. During this process, AI/ML algorithms are developed for robust data cleaning, seamless integration, sophisticated modelling and rigorous validation of geo-scientific information. The ultimate outcomes are the production of highly accurate mineral predictive maps, which significantly narrow down exploration targets, thereby reducing exploration risks, costs and time. This enhanced efficiency and accuracy in mineral targeting is crucial for sustainable mining practices and informed economic planning in India.

2.3. Study Area:

A total of 39,000 sq. km area, covering 52 toposheets within the Dharwar Craton, spanning Karnataka and Andhra Pradesh are given as study area. The bounding Latitude-Longitudes of the assigned area are:

	LAT	LONG
1	15° 45' 00" N	76° 00' 00" E
2	15° 45' 00" N	77° 45' 00" E
3	14° 15' 00" N	77° 45' 00" E
4	14° 15' 00" N	77° 15' 00" E
5	13° 45' 00" N	77° 15' 00" E
6	13° 45' 00" N	76° 00'00" E

2.4. Area of Interest:

The Dharwar Craton of southern India is a geologically ancient and economically vital region. It's a treasure trove of diverse mineral deposits, primarily found within its distinctive greenstone belts (Bababudan, Shimoga, Chitradurga, Gadag, Kolar, Huttī, Sargur and Sandur schist belts distributed across its Western and Eastern divisions) and associated granitoid complexes along with few mafic – ultra mafic intrusive. Varied rock types, long tectonic history and extensive network of shear zones have created ideal conditions for various mineral systems. While gold and iron are the main economic drivers, base metals and critical minerals are emerging exploration frontiers. Key exploration targets include shear zones for gold with an IOCG setup, submarine volcanic successions for VMS Cu-Pb-Zn and younger acidic to intermediate intrusive complexes for REE and Cr-Ni deposits.

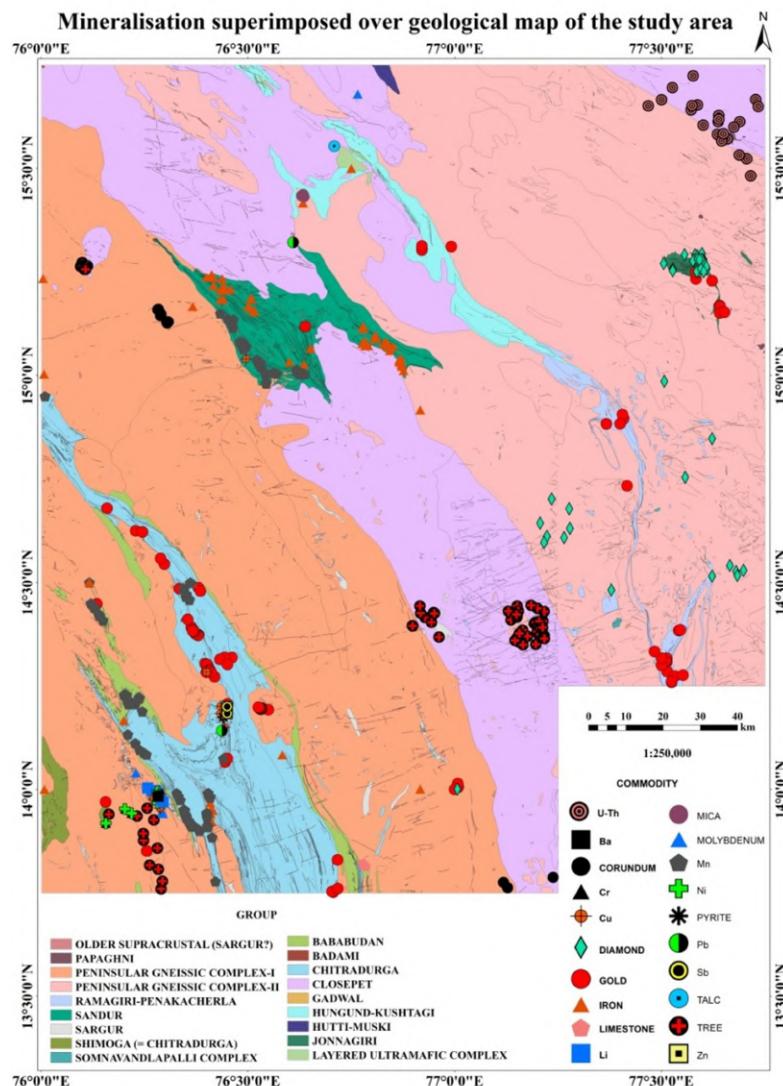


Figure 2.1: Geological map showing Mineralisation of the study area

3. Resources Utilized

3.1. Hardware:

The computational tasks for this project were primarily executed on a dedicated workstation with the following specifications:

Processor: 12th Gen Intel(R) Core(TM) i7-12700 @ 2.10 GHz

Installed RAM: 16.0 GB (15.7 GB usable)

Graphics Processing Unit (GPU): Intel UHD Graphics 770, (8154 MB)

Storage: 930 GB Local Disk

Operating System: Windows 11 Pro (64-bit operating system, x64-based processor)

3.2. Software:

The project leveraged a combination of specialized geospatial software and a robust Python environment for data processing, machine learning, and visualization.

- **GIS Software:** All Geographic Information System (GIS) related tasks, including data visualization, spatial analysis, and map composition, were performed using licensed version of ArcGIS Pro v.3. Geosoft Software used for geophysical data processing.
- **Programming Environment:** The majority of data processing, machine learning model development, and advanced analytical tasks were conducted using Python. The key Python libraries and frameworks utilized include:

pandas (for data manipulation and analysis)

numpy (for numerical operations)

seaborn (for statistical data visualization)

matplotlib.pyplot (for plotting and visualization)

rasterio (for reading and writing raster data, and warping)

geopandas (for working with geospatial data in tabular format)

shapely (for geometric objects)

Scikit-learn (sklearn) (for machine learning models, pre-processing, dimensionality reduction, and evaluation metrics)

lightgbm (for LightGBM model)

xgboost (for XGBoost model)

imblearn (for handling class imbalance, if used)

os (for interacting with the operating system, e.g., path manipulations, directory creation)

3.3. Manpower:

The details of the team of 4 members are as follows:

1. **Soumya Mitra**, Senior Geologist, GSI, NER, M-III, Geodata, Shillong, (Team Leader), (EMP ID: 121493)
2. **Saptarshi Mallick**, Senior Geologist, SU: Meghalaya, GSI, NER, Shillong. (Team Member) (EMP ID: 121257)
3. **Kshounish Patra**, Senior Geologist, SU: Meghalaya, GSI, NER, Shillong. (Team Member) (EMP ID: 121380)
4. **Santu Biswas**, Senior Geophysicist, GSI, ER, Kolkata. (Team Member) (EMP ID: 122222)

4. Data Used in Detail

4.1 Category of Data

A comprehensive and multi-thematic dataset has been compiled from various geo-scientific disciplines to capture the diverse characteristics of the study area. The primary datasets used for feature extraction and subsequent machine learning model development.

Geological: Multi-Layer Geological Map of Karnataka and Andhra Pradesh (25K Scale)

Multi-Layer Geological Map of Karnataka and Andhra Pradesh (50K Scale)

Lineament Features of Karnataka and Andhra Pradesh

Geochemical: National Geochemical Mapping Data of Karnataka and Andhra Pradesh

Geophysical: Aero-geophysical Magnetic Data of Karnataka and Andhra Pradesh

Aero-geophysical Spectrometric Data of Karnataka and Andhra Pradesh

Ground Gravity Data of Karnataka and Andhra Pradesh

Mineral Exploration: Mineral Exploration Blocks in Karnataka and Andhra Pradesh

PGRS Data: Mineral Maps from Advanced Space-borne Thermal Emission and Reflection Radiometer (ASTER) Data – Karnataka and Andhra Pradesh

Technical reports: Reports on Geological mapping, geochemical mapping, geophysical mapping and mineral exploration of Karnataka and Andhra Pradesh.

4.2. Data Acquisition and Initial Assessment:

A comprehensive and multi-thematic dataset was compiled from various geo-scientific disciplines to capture the diverse characteristics of the study area, which spans approximately 39,000 sq. km across the states of Karnataka and Andhra Pradesh, India. These primary datasets served as the foundation for feature extraction and subsequent machine learning model development.

4.2.1. Geological Data:

The geological framework of the study area was established using multi-layer geological maps, sourced from the Geological Survey of India (GSI). This includes detailed maps at both 1:25,000 scale and 1:50,000 scale, providing granular information on lithological units, structural features such as faults and folds and the contacts between different rock types, as well as intrusive bodies with the country rock. These vector datasets are crucial for understanding the bedrock geology and are subsequently re-projected to a common Coordinate Reference System (CRS) and rasterized to the target resolution to enable integration with other grid-based data. Complementing the geological maps, lineament features for Karnataka and Andhra Pradesh were utilized. These features, derived from satellite imagery and geological mapping, provide essential information on the orientation, density, and length of linear geological structures like faults, fractures and dykes, which often serve as controls for mineralizing fluids and their emplacement. These have also been re-projected and converted into density or proximity rasters.

4.2.2. Geochemical Data:

The National Geochemical Mapping (NGCM) Data for Karnataka and Andhra Pradesh, obtained from the Geological Survey of India (GSI), formed a vital component of the dataset. This point-based data includes concentrations of a wide array of major oxides, trace elements, and critical precious metals such as Gold (Au), Platinum (Pt), Palladium (Pd), Rare Earth Elements (REE), Copper (Cu), Nickel (Ni) and Arsenic (As). This dataset has undergone extensive pre-processing, including the treatment of values recorded as '0' as missing data (NaN), filtering for data completeness (elements with less than 50% null values were retained), and subsequent imputation and normalization. Furthermore, Principal Component Analysis (PCA) was applied to this high-dimensional dataset for dimensionality reduction and to extract composite geochemical signatures.

4.2.3. Geophysical Data:

Three distinct types of geophysical datasets have been provided for subsurface information's which are crucial for identifying concealed mineralization. Aero-geophysical Magnetic Data, acquired by the GSI for Karnataka and Andhra Pradesh, provided information on the Total Magnetic Intensity (TMI) and its various derivatives (e.g., Analytic Signal), which are instrumental in mapping lithological boundaries, structural features, and magnetic bodies often associated with mineralization. Aero-geophysical Spectrometric Data, also from the GSI, offered insights into the distribution of naturally radioactive elements like Potassium (K), Thorium (Th), and Uranium (U), along with their ratios, which can indicate alteration zones. Lastly, Ground Gravity Data, provided as both point and gridded data by the GSI, was used to understand subsurface density contrasts, represented by Bouguer Anomaly and Residual Gravity and their derivatives. All geophysical raster datasets were resampled to a common grid resolution, normalized and subjected to various derivative filters to enhance structural and lithological features.

4.2.4. Mineral Exploration Data:

Information regarding Mineral Exploration Blocks in Karnataka and Andhra Pradesh was compiled from sources such as the GSI, State Mining Departments and the public domain. This vector-based data includes the boundaries of existing exploration blocks and license areas. While not directly used as a predictive feature, this dataset served as critical contextual information, aiding in the definition of 'known' mineralized (positive) areas for model training and 'unexplored' (negative) areas, as well as for overlaying with the final predictive maps to assess the delineation of new prospective zones relative to existing exploration activities.

4.2.5. PGRS (Photogeology and Remote Sensing) Data:

Derived Mineral Maps from Advanced Space-borne Thermal Emission and Reflection Radiometer (ASTER) Data for Karnataka and Andhra Pradesh contributed significantly to surface alteration mapping. These raster-based maps provide key alteration indices, such as those related to hydroxyl minerals, argillic alteration, iron oxides and mafic minerals, as well as features for lithological discrimination. This remote sensing data was pre-processed for atmospheric correction and noise reduction and the calculated indices were subsequently normalized for consistency with other input layers.

4.2.6. Technical Reports:

A collection of technical reports from the Geological Survey of India, academic institutions, and exploration companies provided invaluable contextual information. These documents, covering geological mapping, geochemical mapping, geophysical surveys and previous mineral exploration findings in Karnataka and Andhra Pradesh, have extensively utilized for developing the conceptual genetic models of the targeted mineral systems, understanding regional metallogeny and validating the geological plausibility of the machine learning model's predictions.

5. Conceptual Genetic Models and Targeting Criteria

5.1. Overview:

The Dharwar Craton of southern India is a significant Archean continental crustal block, formed between 3.6 and 2.5 billion years ago (Naqvi & Rogers, 1987). It is broadly divided into two major tectonic blocks: the Western Dharwar Craton (WDC) and the Eastern Dharwar Craton (EDC), separated by the prominent Chitradurga Shear Zone (Chadwick et al., 2000). The WDC is characterized by older lithologies, including ancient tonalite-trondhjemite-granodiorite (TTG) gneisses, often referred to as the Peninsular Gneiss (dating back over 3.0 Ga), and older greenstone sequences of the Sargur Group (3.1-3.3 Ga). These are unconformably overlain by the younger Dharwar Supergroup (2.9-2.6 Ga), which includes the Bababudan and Chitradurga groups, composed of volcano-sedimentary rocks. Metamorphic grades generally increase from greenschist facies in the north to granulite facies in the south (Ramakrishnan & Vaidyanadhan, 2010). The EDC, in contrast, is dominated by relatively younger Neoarchean granitoids (2.7-2.5 Ga), including the voluminous Closepet Batholith, which intrude narrower, volcanic-dominated greenstone belts (e.g., Kolar Group). The EDC shows more evidence of active magmatism and tectonic reworking (Ramakrishnan & Vaidyanadhan, 2010).

The Dharwar Craton provides crucial insights into early Earth processes, with evidence for both vertical (mantle plume related) and horizontal (subduction-accretion) tectonic regimes in its evolutionary history. It is also economically significant, hosting important deposits of gold, iron and base metals (Naqvi & Rogers, 1987).

Stratigraphic Table of the Dharwar Craton

Group/Unit	Approximate Age (Ga)	Dominant Lithology / Characteristics	Primary Location
Younger Granitoids	2.7 - 2.5 (Bhaskar Rao et al., 2000)	Closepet Batholith, other granitic intrusions, post-tectonic	EDC (predominantly)
Dharwar Supergroup	2.9 - 2.6 (Chadwick et al., 2000)	Volcanic and sedimentary rocks (Bababudan & Chitradurga Groups), overlies Peninsular Gneiss	WDC & EDC
Sargur Group	3.1 - 3.3 (Bhaskar Rao et al., 2003)	Older greenstone sequences, mafic-ultramafic rocks, metasediments, enclaves within Peninsular Gneiss	WDC
Peninsular Gneiss Complex	>3.0 (Peucat et al., 1993)	TTG (Tonalite-Trondhjemite-Granodiorite) gneisses, basement rocks	WDC (widespread)

5.2. Mineral System:

The mineral system concept is vital for mineral prospectivity because it allows predicting where new deposits might be found. The mineral system is not just about the ore itself, but the entire journey of the ore from their source rocks, through transport mechanisms (like hot fluids), along pathways (such as faults), to eventually accumulate in a trap where they deposit. A driver, often tectonic or magmatic activity,

provides the energy, and finally, preservation ensures the deposit isn't destroyed. Therefore, by understanding the entire system, explorers can identify areas that have all the right conditions for mineralization, even if surface clues are subtle. It shifts exploration from simply looking for known deposit types to understanding the underlying geological processes, leading to more efficient, less risky, and ultimately more successful discoveries in both well-explored and frontier regions.

Major Mineral Systems of entire Dharwar Craton:

1. Orogenic Gold System

- Tectonic Setting: Archean greenstone-granite terranes, formed during crustal shortening and shearing events.
- Host Rocks: Amphibolites, meta-volcanics, BIFs, quartz veins and intrusive granites.
- Controls: Regional shear zones, faults, lithological contacts.
- Deposits: Kolar Gold Fields, Huttı Gold Mines, Gadag, Ramagiri, Ajjanahalli prospects

2. Volcanogenic Massive Sulphide (VMS) System

- Tectonic Setting: Submarine volcanic sequences in extensional back-arc or rift environments.
- Host Rocks: Mafic to felsic volcanics, tuffs, and BIFs.
- Controls: Synvolcanic faults, exhalative sedimentary facies.
- Deposits: Ingaldhal (Cu), Chitradurga Cu belts, Sandur belt (Cu, Zn)

3. Banded Iron Formation (BIF)-Hosted Iron System

- Tectonic Setting: Marine basins associated with greenstone belts.
- Host Rocks: Banded iron formations (oxide facies).
- Controls: Stratigraphic layering, basin architecture.
- Deposits: Sandur, Donimalai, Bababudan, Kudremukh

4. SEDEX (Sedimentary Exhalative) Base Metal System

- Tectonic Setting: Sedimentary basins with basin-wide hydrothermal circulation.
- Host Rocks: Graphitic schists, greywackes, cherty units.
- Controls: Syn-sedimentary faults, basin margin structures.
- Deposits: Zangamrajupalle (Pb-Zn-Cu), Sandur Sub-basin – Cu-Zn horizons

5. Carbonatite-REE-Nb-Ta System

- Tectonic Setting: Intraplate magmatic activity, mantle-derived alkaline intrusions.
- Host Rocks: Carbonatites, nepheline syenites, alkali granites.
- Controls: Ring structures, rift zones, intrusive contacts.
- Deposits: Pajarla-Korlapahad– REE, Nb, Pulivendla and adjacent alkaline complexes

6. Lateritic REE System

- Tectonic Setting: Weathering profiles over REE-rich rocks (carbonatites, granites).
- Host Rocks: Laterites developed on syenites and carbonatites.
- Controls: Paleosurface, climate, and depth of weathering.
- Deposits: Pajarla & Srikakulam belt (AP) – Monazite-bearing laterites

7. Magmatic Nickel-Cr-PGE System

- Tectonic Setting: Archean ultramafic-mafic intrusions in stable cratonic settings.
- Host Rocks: Dunite, peridotite, pyroxenite, serpentinite.

- Controls: Layered intrusions, cumulate layering, shear zones.
- Deposits: Nuggihalli Schist Belt – Chromite, Ni

8. Pegmatite-Li-Be-Ta (Rare Metal) System

- Tectonic Setting: Late-stage granitic differentiation during craton stabilization.
- Host Rocks: Pegmatites, granites.
- Controls: Fractures, granitoid contacts.
- Potential Areas: Eastern Dharwar Craton margins, Basin margins with pegmatite swarms

5.3. Geo-scientific Controls:

Protracted geological evolution, involving multiple episodes of magmatism, metamorphism, deformation and fluid activity of the Dharwar Craton, has led to the concentration of valuable commodities such as copper, iron, manganese and gold. Understanding the geoscientific controls on their formation and the associated geological, geophysical, geochemical, and remote sensing signatures is crucial for effective exploration and resource targeting within this terrane.

Mineral Resource	Typical Geological Controls	Typical Geophysical Signatures	Typical Geochemical Signatures	Typical Remote Sensing Signatures
Rare Earth Elements (REE)	Alkaline-carbonatite complexes, peralkaline granites, pegmatites. Also secondary enrichment	High magnetic (magnetite/pyrrhotite in carbonatites), elevated radiometric (Th, U, K), positive gravity (carbonatites).	High LREE/HREE, enrichment in Sr, Ba, P, F, Nb, Ta, Zr, Hf.	Tonal anomalies & distinctive spectral responses. Structural lineaments indicative of fluid pathways.
Nickel-Platinum Group Elements (Ni-PGE)	Magmatic sulfide deposits in komatiites, layered mafic-ultramafic intrusions. Sulfide saturation. Shear zones for remobilization.	Strong EM conductors (sulfides), variable magnetic (pyrrhotite/magnetite), positive gravity (dense sulfides).	High Ni, Cu, Co, PGE, elevated Se, Te, As. Chalcophile element enrichment.	Limited direct signatures. Structural controls (faults, shear zones) and mafic-ultramafic lithologies from imagery.
Copper (Cu)	Volcanogenic Massive Sulfides (VMS) in greenstone belts, porphyry copper (less common), sediment-hosted, epigenetic vein-type. Associated with intermediate-felsic volcanics/intrusives.	EM conductors (massive sulfides), IP chargeability (disseminated sulfides), variable magnetic, K enrichment in alteration (radiometric).	Elevated Cu, Mo, Au, Ag, Pb, Zn.	Characteristic alteration assemblages (phyllitic, potassic, propylitic, argillic) Structural controls.
Iron (Fe)	Banded Iron Formations (BIFs - Algoma/Superior type). Lateritic iron ores (supergene enrichment).	Strong, characteristic high magnetic anomalies (magnetite-rich BIFs). Positive gravity (dense iron oxides).	High Fe content (hematite, magnetite, goethite). Low P, S, Si.	Distinctive reddish-brown coloration (gossans). Linear features reflecting BIF strike. Topographic highs on resistant BIFs.
Manganese (Mn)	Sedimentary deposits associated with BIFs	Variable magnetic (if associated iron oxides).	High Mn content (pyrolusite,	Dark, blocky outcrops. Limited

Mineral Resource	Typical Geological Controls	Typical Geophysical Signatures	Typical Geochemical Signatures	Typical Remote Sensing Signatures
	or volcano-sedimentary sequences. Supergene enrichment.	Potentially low resistivity. Positive gravity (dense Mn oxides).	psilomelane, braunite, rhodochrosite). May be associated with Fe, Si, P.	direct spectral signatures but occur in mappable stratigraphic horizons.
Gold (Au)	Orogenic gold deposits in shear zones within greenstone belts (quartz veins, sulfides, carbonate alteration). Magmatic-hydrothermal systems. Placer gold.	Magnetic lows (magnetite destruction) or highs (pyrrhotite). Linear magnetic breaks (shear zones). IP chargeability (disseminated sulfides). EM (conductive sulfides).	Au, Ag, As, S, Sb, Bi, Te, W. Alteration (silicification, carbonatization, sulfidation, sericitization). Geochemical halos.	Linear topographic features, prominent lineaments (shear zones). Tonal/vegetation anomalies (alteration, quartz veins).

5.4. Definition of Mineral Targeting Criteria:

5.4.1. Geological concerns:

While preparing mineral prospectivity map, these key geological elements and their spatial relationships were taken care of:

- **Greenstone Belts:** Extent and internal stratigraphy of different greenstone belts (Sargur, Bababudan, Chitradurga, Kolar, Huttī, Sandur, etc.) as they are fundamental for gold, BIF-iron, manganese and base metals.
- **Shear Zones and Major Structures:** The regional and local shear zones, faults and fold patterns are crucial for orogenic gold and influence the emplacement of other deposit types.
- **Granitoid Intrusions:** The type, age and extent of various granitoids (TTG gneisses, Closepet Granite and its equivalents, younger K-rich granites, alkali granites) and their contacts and associated hydrothermal systems are important for gold, U, REE and potentially Sn-W.
- **Lithological Contacts:** Contacts between different rock types (e.g., BIF and volcanics, ultramafics and surrounding rocks, carbonates and clastics) are often favorable loci for mineralization.
- **Metamorphic Grade:** The metamorphic history (greenschist to amphibolite facies) influences fluid generation and mobilization for orogenic gold.
- **Unconformities:** Particularly important for uranium mineralization where Proterozoic basins overlie the Archaean basement.

From these geological understanding the following geological layers has been constructed: Normalized Lithological layers and Normalized Group layers for selected commodity, Normalized lineament intersection buffer layer, Normalized dyke layer, Normalized lineament ring buffer layer, Normalized contact layer between intrusive granitoids and schist belts, Normalized lineament density layer (All the files are shown in Figure 5.1) and Normalized Known mineralisation layers for a) Au, b) Cu, c) NiPGE, d) REE and e) Fe-Mn.

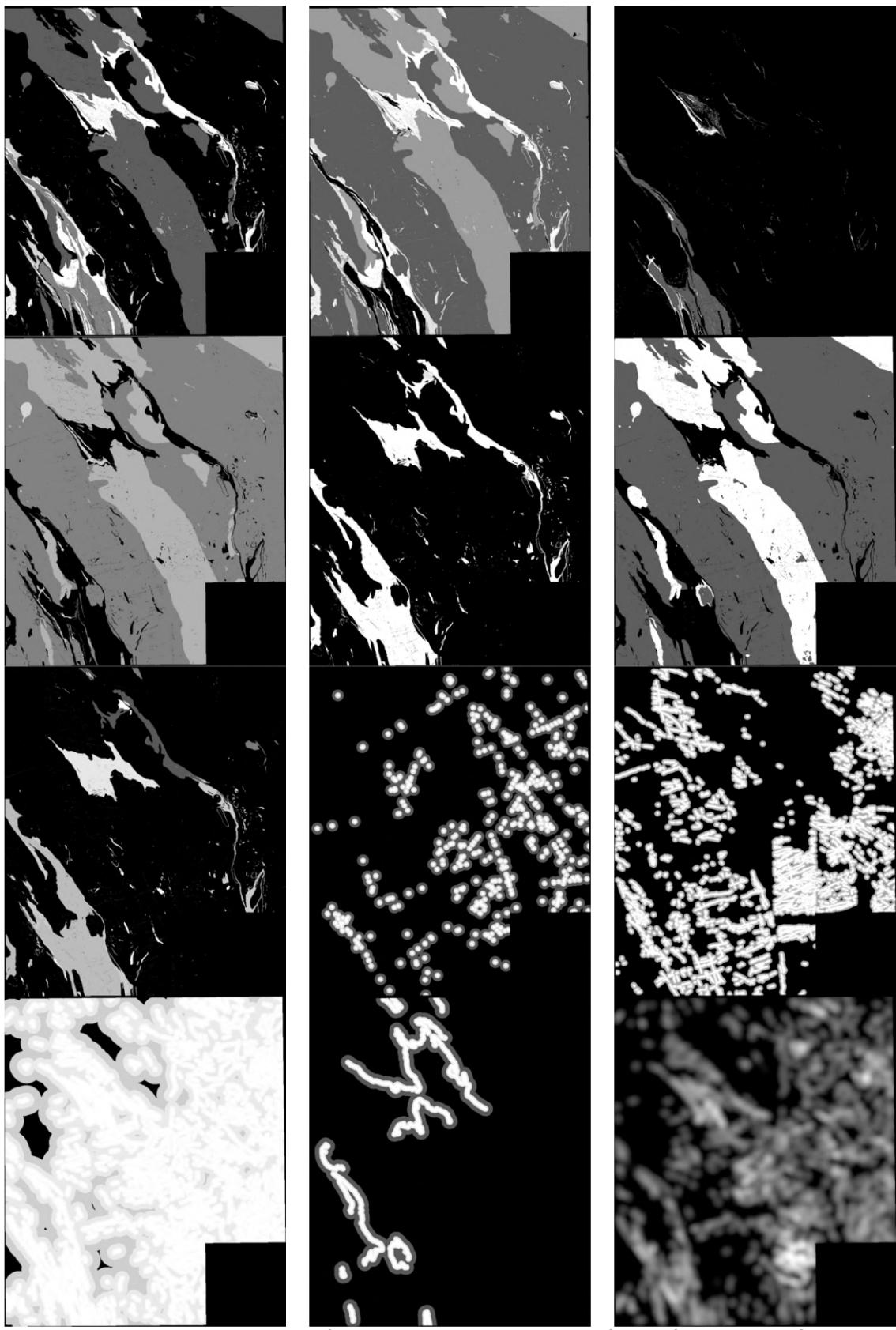


Figure 5.1 Factor maps derived from geological layers. Details of these files added in Github.

5.4.2 Geophysical data processing and outputs:

Geophysical support for mineral exploration is a multi-stage, iterative process which begins with defining survey objectives (e.g., choosing ground or airborne magnetic surveys), followed by data acquisition, which involves planning and field measurements. Data processing corrects, enhances and displays the data to highlight relevant geological information. The processed data can be displayed in various formats. Finally, data interpretation analyses the geophysical data to create a geological model of the area. The type of survey and data acquisition impact the resolution of geological information, and interpretation is closely tied to how the data is processed and displayed. Interpretation isn't a standalone task but an integral part of the iterative analysis.

Gravity and magnetic data reduction

Gravity Data: Field gravity measurements were acquired using a CG-5 gravimeter. To prepare the raw data for interpretation, several corrections were applied:

- Tidal correction: Applied in the field using the gravimeter's in-built program.
- Drift, latitude, elevation (Free Air), Bouguer slab, and terrain corrections: Applied through Excel sheets and Geosoft.
- Density assumption: 2.67 g/cc (standard crustal density).
- Terrain correction: Computed using Shuttle Radar Topography Mission (SRTM) data (90 m resolution), with values ranging from 0.56 to 5.70 mGal.

Magnetic Data: Magnetic readings were acquired and tuned at 45000 nT after calibration.

The following corrections were applied:

- Diurnal variations and IGRF corrections (International Geomagnetic Reference Field) were applied manually using Excel.

Data processing and preparation of maps

The Oasis Montaj (version 9.3) package of Geosoft Inc. has been used for the processing and generation of gravity and magnetic anomaly maps. The Bouguer gravity anomaly and IGRF corrected magnetic anomaly values have been projected through UTM projection system with WGS 84 datum in zones 43N and, a digital base map has been prepared. Bouguer gravity and magnetic anomaly values have been gridded at 500 m intervals and, contours have been drawn on 1mGal and 50 nT intervals, respectively to prepare anomaly maps. The gridded gravity and magnetic data have been filtered by digital filtering techniques using Fast Fourier Transform (FFT) (Bhattacharyya, 1967) to enhance the response of a geological structure, suppressing the effect of noise.

Results and Discussions

Gravity and magnetism involve the interaction of objects at a distance through the respective fields surrounding the objects. A gravity field is caused by an object's mass, a magnetic field by its magnetism. Bouguer gravity anomaly map brings out the density contrast amongst geological features and formations. It enables to distinguish individual units from an assemblage based on their characteristic gravity signatures. Basic and ultrabasic rocks are denser than felsic and sedimentary rocks and hence, yield higher gravity signals.

Bouguer gravity anomaly contour map:

The Bouguer Anomaly (BA) map of the study area has been prepared and presented in Figure 5.2a. A total gravity variation of 66 mGal ranging from -111 mGal in the southern-eastern and eastern part to a maximum of -45 mGal in the central part of the area. The gravity anomaly map reflects a prominent trend in contours in northwest and southeast direction over peninsular gneissic complex. The regional trends in the area are in good correlation with structural trend of geological formation. The Bouguer gravity picture on the block has been made quite significant with the presence of three gravity lows and highs that aligned roughly in NW-SE direction.

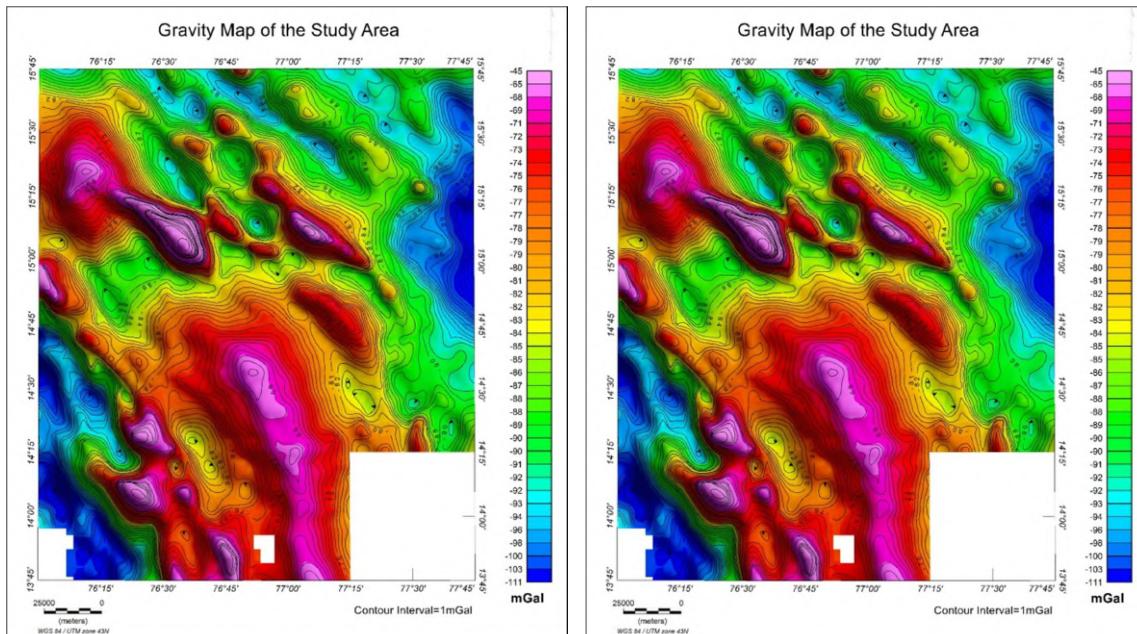


Figure 5.2a: Bouguer gravity anomaly map of the study area, 2b: RTP Map of the aero-magnetic data of the study area.

The magnetic method is based on susceptibility contrast due to subsurface distribution of magnetic minerals, primarily magnetite. The Magnetic Anomaly (MA) map of the Study area has been prepared and reduced to pole and presented in Figure 5.2b. A total magnetic variation of 26977 nT, ranging from -4595 nT to 22382 nT, has been observed in the area. Reduction to the Pole (RTP) is a standard part of magnetic data processing method, especially for large-scale mapping. RTP operation can transform a magnetic anomaly caused by an arbitrary source into the anomaly that the same source would produce if it is located at the pole and magnetized by induction only. Interpretation of magnetic data can further be helped by RTP in order to remove the influence of magnetic latitude on the anomalies, which is significant for anomalies caused by crust.

For enhancement of the anomalies pertaining to regional features in suppression of the responses of local features, an upward continuation filter is used. It calculates the potential field at an elevation higher than at which the field is actually measured. The gravity field from an upward continuation level of 'Z' retains anomalies from sources below a depth of $Z_0 = (1/2) * Z$, whereas sources above are more attenuated (Jacobsen, 1987; Lyngsie et al, 2006). The regional Bouguer anomaly map and magnetic anomaly map of the study area is presented in the form of images in Figure 5.3a and 3b. The regional map has been accepted

for upward continuation map at 5000m where all the local anomalies have been removed. The map has clearly brought out the general trend of regional features in the area.

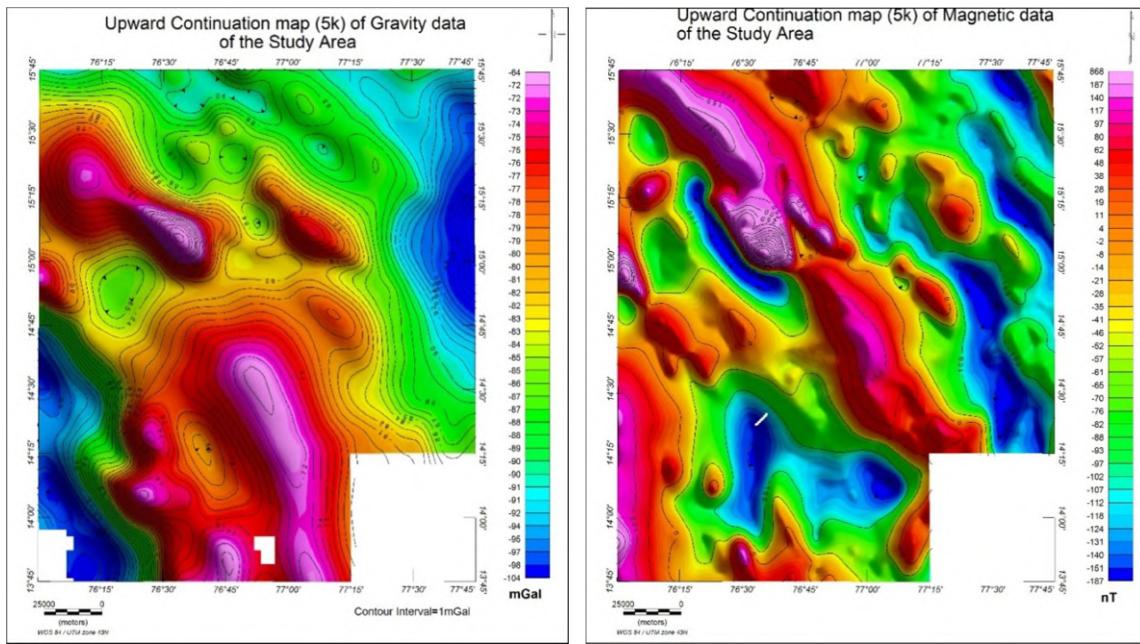


Figure 5.3a: Upward Continuation map (5km) of the gravity data and, **3b:** Upward Continuation map (5km) of the magnetic data

For a better understanding of surface geology with the data, the regional component of the signature is removed from the gravity and magnetic signal using Geosoft Software by applying grid math expression builder. This map (Figure 5.4a and 4b) depicts the zones of residual highs and lows, possibly correlatable with various geological formations.

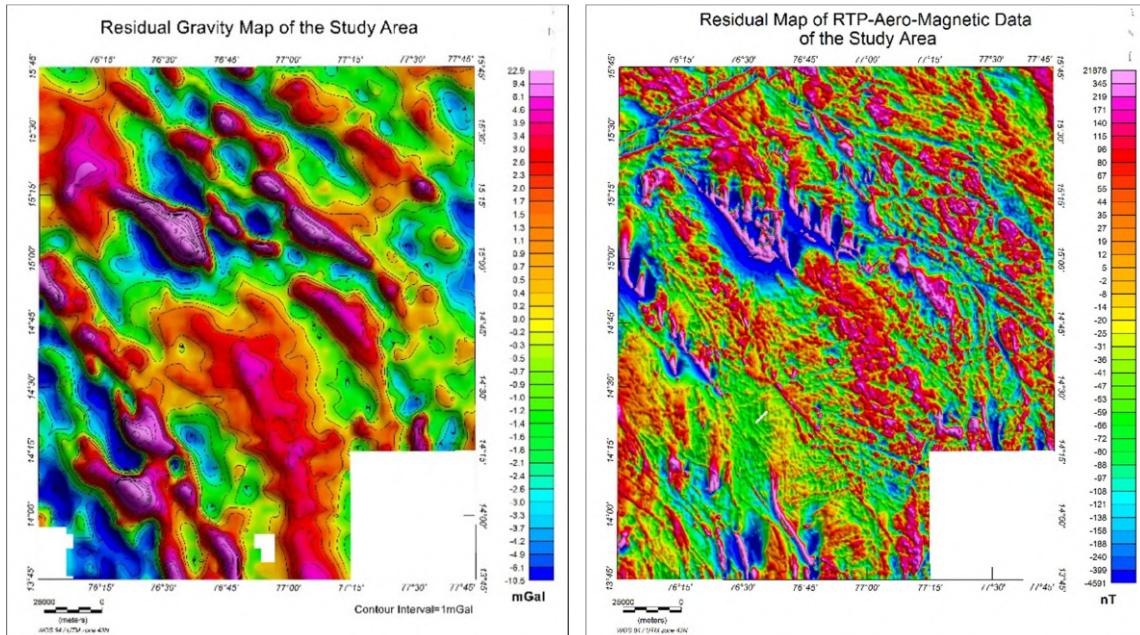


Figure 5.4a: Residual gravity map of the study area, **4b:** Residual Magnetic map.

Tilt derivative (TDR) is a potential field data processing technique that is adequate to detect edges of structural bodies by edge enhancement. TDR is obtained from the ratio of the vertical gradient of the field to the absolute horizontal gradient.

A tilt derivative map (Figure 5.5a) of gravity data in the study area has been prepared to visualize the lateral extension of near-surface and deeper features. It may be easily visualized that the signal from the near-surface features in the Tilt derivative map is further enhanced in comparison to BA contour map. The Tilt Derivative map of gravity anomaly has brought out a number of local/hidden features including secondary geological structures like contacts, faults/folds, ridges, and troughs which were not visible in original BA contour map prominently. The proper disposition of major linear structures, lateral continuity of Dhanjori metabasalt, Singhbhum granite etc. are well reflected in the map.

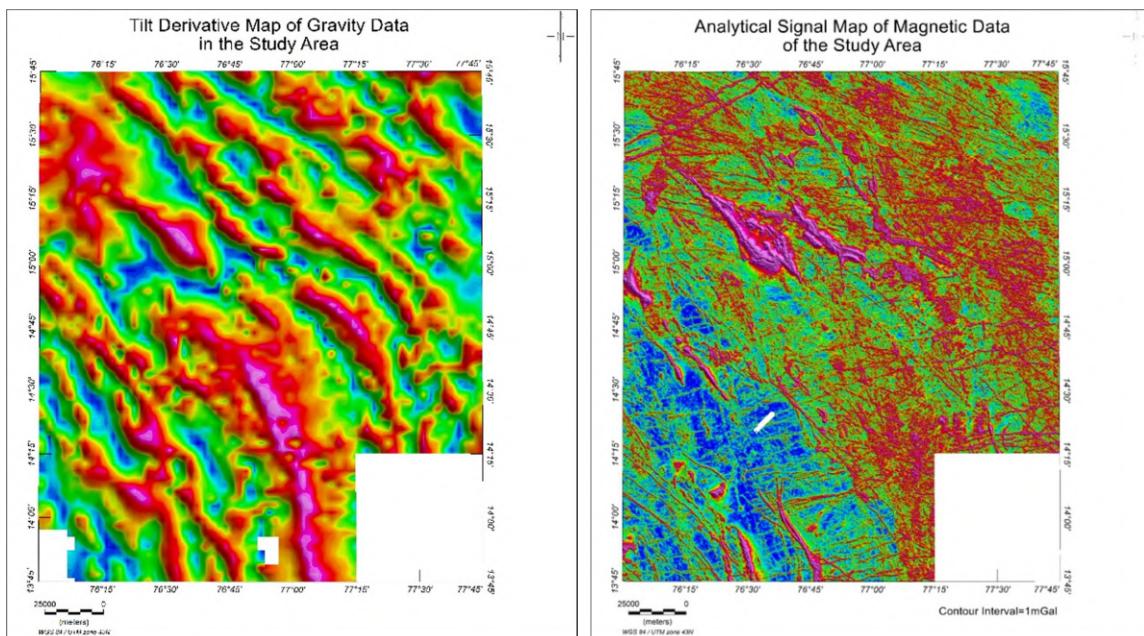


Figure 5.5a: Tilt derivative map of gravity data in the study area, 5b: Analytical signal map of aero-magnetic data in the study area.

Analytic Signal (AS) maps (Figure 5.5b) highlight short-wavelength anomalies and, hence, are useful for locating shallow anomalous bodies and delineating their edges of anomalous bodies as well as formation contacts (Nabighian, 1972). The AS maxima occur directly over the causative body.

Spectral analysis of gravity and magnetic data is a conventional technique for the determination of depths of geological features, particularly the basement (Spector and Grant, 1970; Maus and Dimri, 1996). Spectrum segment with smaller wave number corresponds to deeper features whereas higher wave numbers represent shallow features. Slopes of spectral segments yield the estimates of average depths to the source bodies. The depth of ensemble of sources responsible for each segment of a power spectrum is computed from the following relation:

$$\text{Depth} = - \frac{\text{Slope}}{4\pi}$$

In the present case, the radially averaged power spectra of gravity anomaly (Figure 5.6) have been computed for the study area, which shows two spectrum segments for which

regression lines are fitted. The density interfaces are detected at depths of about 15 km and 4km.

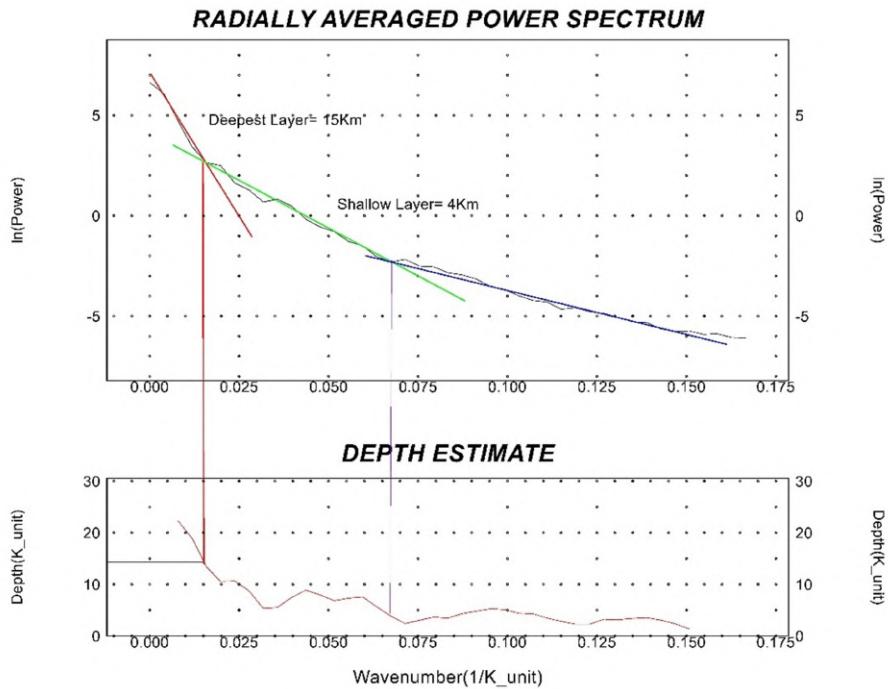


Figure 5.6: Radially averaged power spectrum of gravity data for depth estimation

The 3D Euler deconvolution method of Reid et al. (1990) is widely used to determine depth persistence and subsurface locations of causative sources for discrete gravity and magnetic anomalies based on the gradient of the potential field. Euler solutions are particularly used for delineating the horizontal and vertical contacts, depth, and nature of the causative source. Each calculation has been run for grid cell size 200m, Structural Index (source body geometry) 0, 0.5 and 1, and window size 3, 5 and 10 to derive depth solutions for the causative body. Acceptable solutions for Structural Index 0 (for contacts) and window size 10 have been considered for the interpretation of results. The depth solutions from gravity anomalies have been superimposed on the gravity anomaly map (Figure 5.7). Maximum solutions have been obtained for signal source depths in between 1000m – 4000m. Few of the solutions came from less than 1000m which indicates shallow depth persistence in the area.

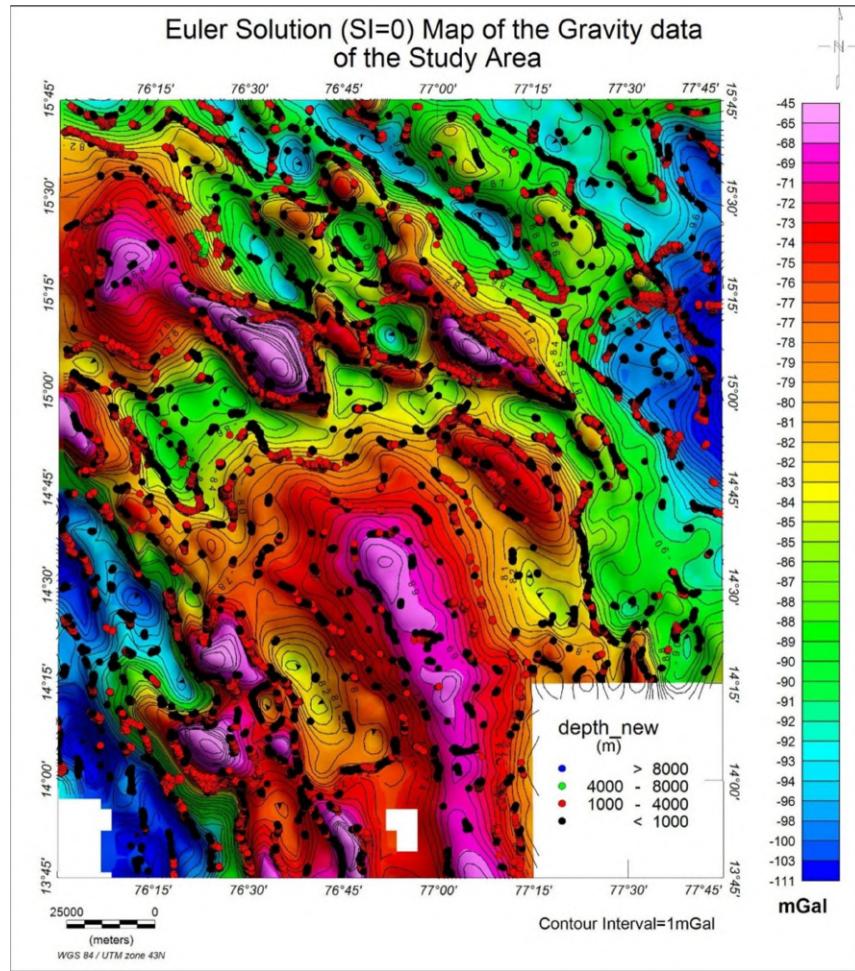


Figure 5.7: Euler Solution map of gravity data, superimposed over bouguer anomaly map of the study area.

5.4.3. Geochemical data processing and outputs:

Principal component analysis (PCA) is a dimension reduction and de-correlation technique that transforms a correlated multivariate distribution into orthogonal linear combinations of the original variables (Hotelling, 1933; Pearson, 1901). In geological context, this dimensionality reduction method is used in conjunction with other techniques to propose high resolution correlation schemes and provide an interpretation of geological conditions (Michael and Craigie, 2021; Scheibe and Michael, 2023).

Objective:

The main objective of the analysis of geochemical data through PCA is mostly related to Scree plot and analysis which determines element-mineral links and relationships between elements and geological conditions (e.g. depositional environment, diagenesis, and weathering). In this report, geochemical data of stream sediment samples are taken into consideration for analysis.

Data Loading and Initial Inspection:

The geochemical dataset, sourced from the NGCM.csv file, serves as a primary input for identifying geochemical signatures related to mineralization. This dataset comprises

10004 entries and 53 columns, including geographical coordinates (Longitude, Latitude) and concentrations of 51 major, trace, and precious elements (e.g., SiO₂ %, Al₂O₃ %, Fe₂O₃ %, Au_ppb, Pt_ppb, Pd_ppb, In_ppm, F_ppm, Te_ppm, Ti_ppm, etc.).

Geochemical Data Pre-processing and Quality Control

A rigorous pre-processing pipeline was implemented to prepare the raw geochemical data for robust analysis:

Handling Below Detection Limit Values: A common challenge in geochemical datasets is the presence of values recorded as 0 which often represent concentrations below the analytical detection limit. To accurately reflect this, all 0 values within the geochemical element columns were systematically replaced with NaN (Not a Number) to be treated as missing data.

Missing Value Analysis: After treating 0s as NaN, a detailed analysis of null percentages for each element was conducted. This revealed significant variations in data completeness. Elements such as In_ppm, F_ppm, Te_ppm, Ti_ppm, Pt_ppb, and Pd_ppb exhibited very high percentages of missing data (often exceeding 80-90%).

Feature Selection based on Completeness: To ensure the reliability of subsequent analyses and model training, elements with a high proportion of missing values (specifically, those with greater than 50% null values) were excluded from the dataset. This filtering step retained a subset of 45 geochemical elements (eda_df) that possessed sufficient data completeness for meaningful statistical analysis and dimensionality reduction. The distribution of the selected elements has been checked again in histogram (Figure 5.8).

Imputation and Scaling: Missing values in the filtered geochemical dataset (eda_df) have been imputed (e.g., using mean, median) and then scaled (e.g., using StandardScaler) before applying Principal Component Analysis. This is standard practice to ensure that PCA is not biased by missing values or by the differing scales and units of the various geochemical elements.

Exploratory Data Analysis (EDA): Correlation Analysis

As part of the exploratory data analysis, a comprehensive correlation matrix was computed for the 45 retained geochemical elements.

Purpose: The correlation matrix quantifies the linear relationships between all pairs of geochemical elements. This is crucial for understanding element associations, identifying potential pathfinder elements, and detecting multicollinearity within the dataset. Strong positive or negative correlations can indicate common geological processes or mineralizing events.

Visualization: An annotated heatmap of this correlation matrix was generated and saved (Figure 5.9). This visualization provides a clear and intuitive representation of the geochemical interdependencies, allowing for rapid identification of element groups that behave similarly.

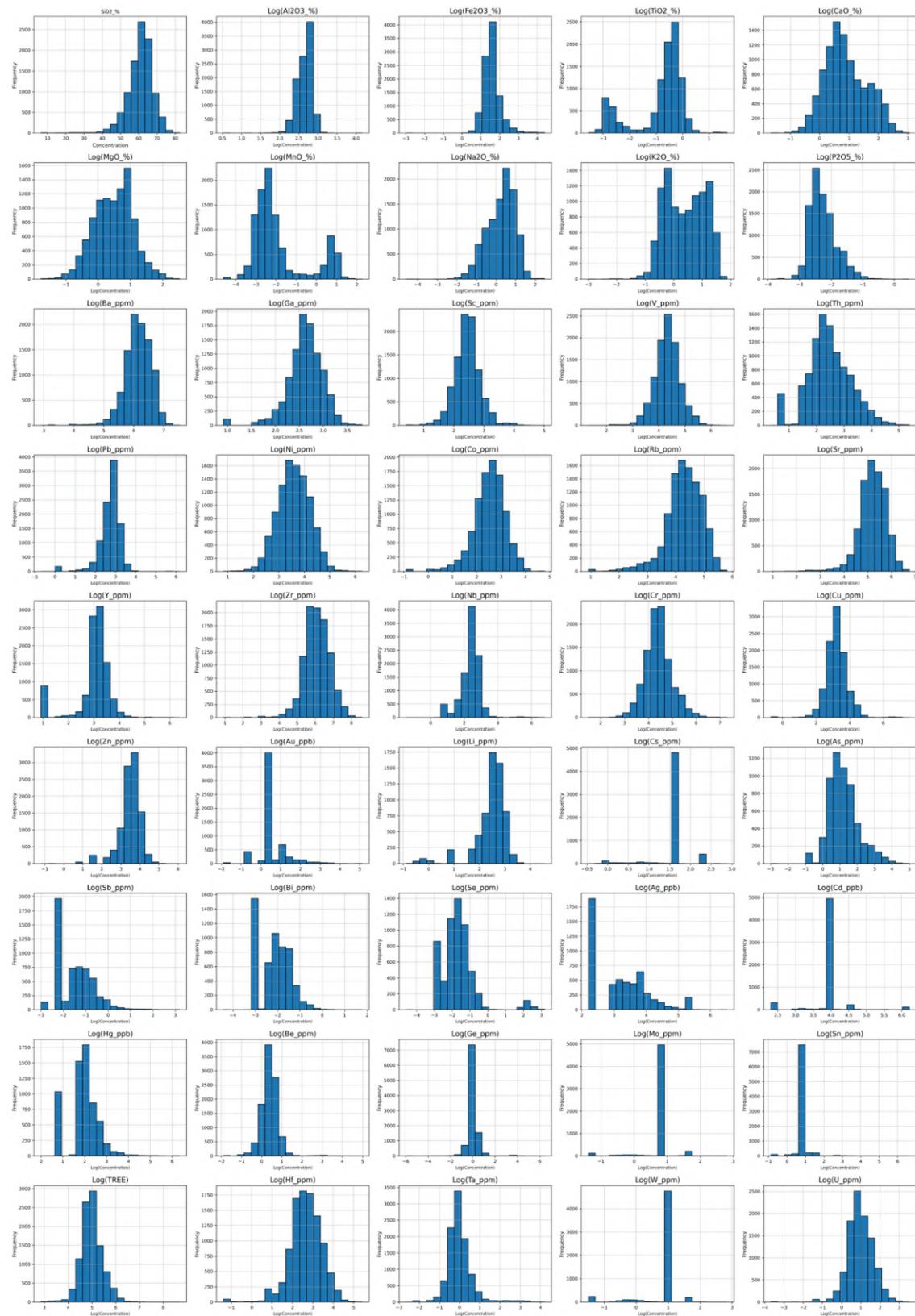


Figure 5.8: Histogram of the selected geo-chemical elements

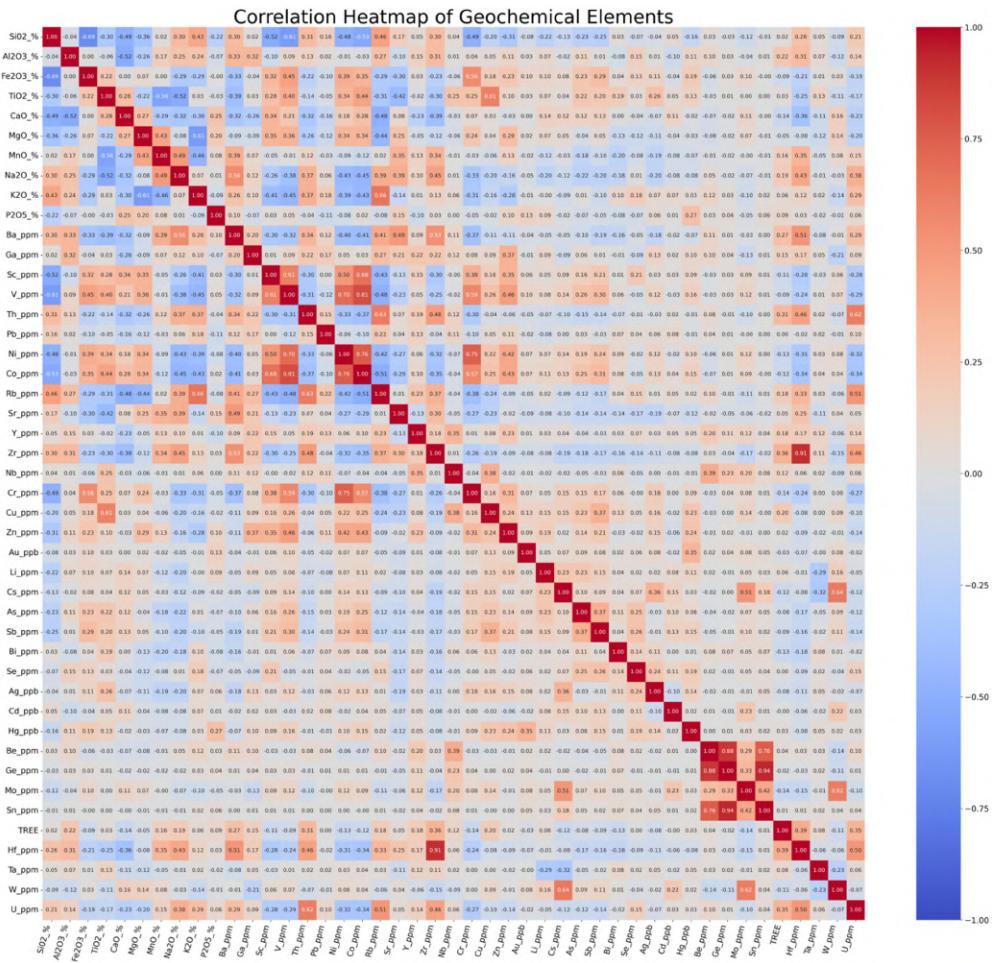


Figure 5.9: Correlation heatmap of the selected geo-chemical elements

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) was applied to the pre-processed and scaled geochemical data as a powerful dimensionality reduction technique.

Rationale: Geochemical datasets often contain a large number of correlated variables, which can lead to multi-collinearity issues in machine learning models and obscure underlying patterns. PCA transforms the original, possibly correlated, variables into a new set of orthogonal (uncorrelated) variables called Principal Components (PCs) (Figure 5.8). Each PC captures a certain amount of the total variance in the dataset, with the first few components capturing the most significant variations.

Outcome: PCA effectively reduced the dimensionality of the geochemical data while retaining the most important information (Figure 5.10 and 5.11). The resulting Principal Components represent composite geochemical signatures. Specifically, selected Principal Components (e.g., norm_pca4.tif in ML_inputdata_Au.ipynb file) were then used as a derived feature in the overall mineral prospectivity modelling. These components encapsulate the collective behaviour of multiple elements, potentially highlighting subtle geochemical anomalies associated with mineralization that might not be apparent from individual element concentrations.

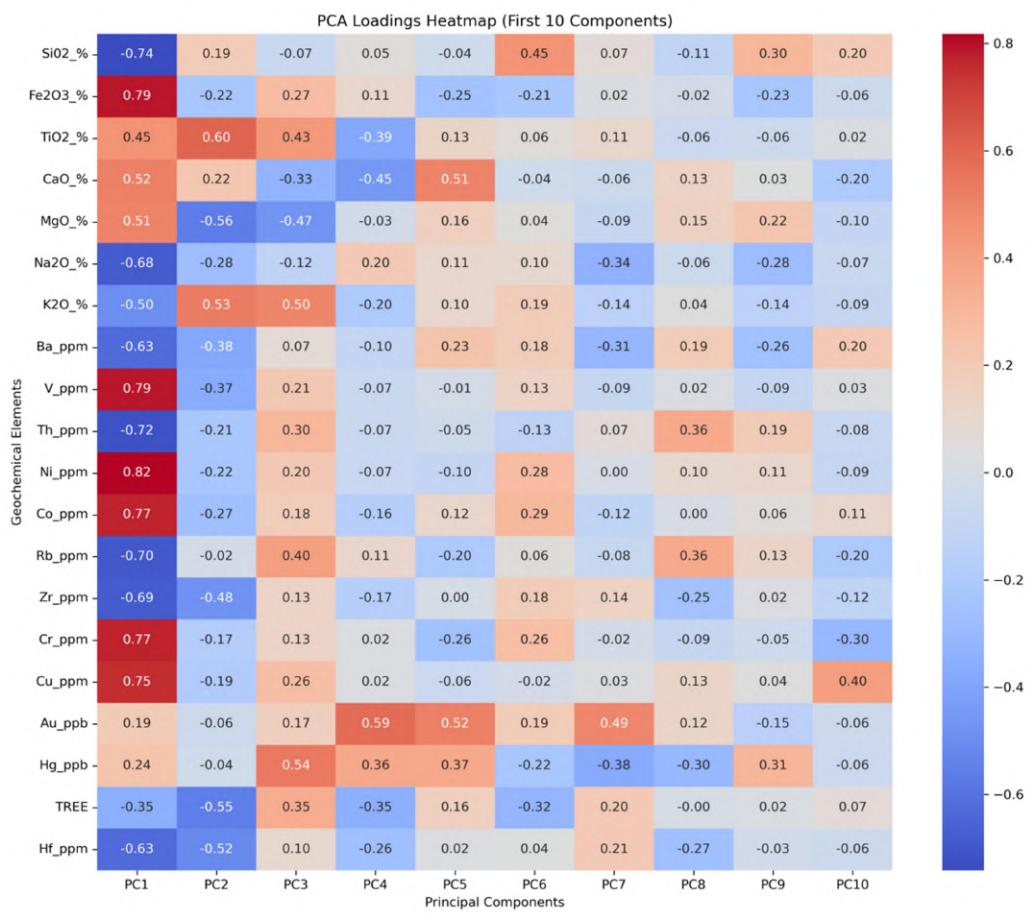


Figure 5.10: PCA loadings heatmap (first 10) of the selected geo-chemical elements

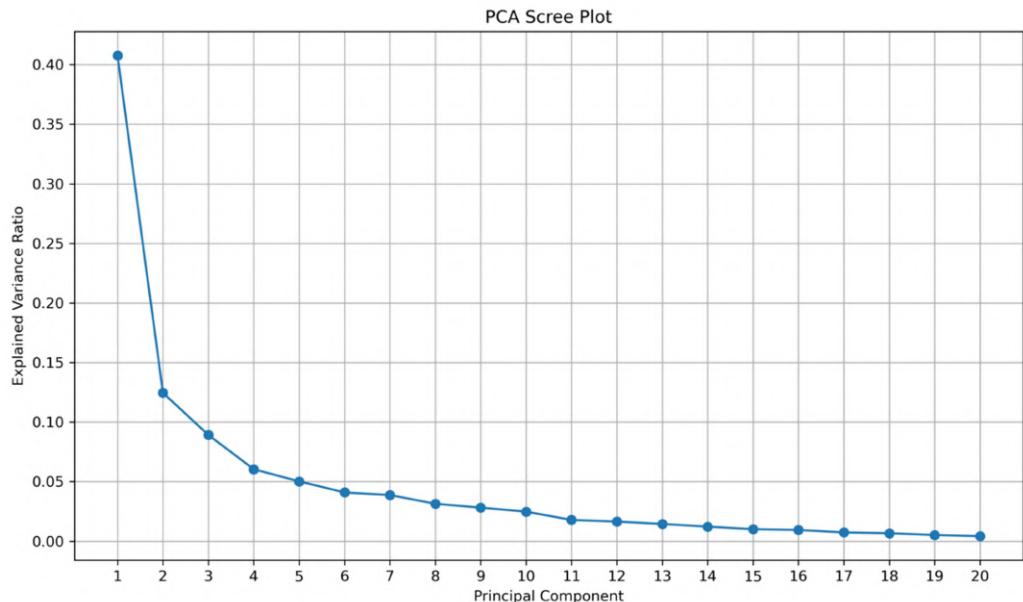


Figure 5.11: Scree plot of the PCA

Interpretation:

Factor-1: Eigenvector-1 reveals a strong geochemical association between the transition metals chromium (Cr), cobalt (Co), nickel (Ni), copper (Cu), and vanadium (V) and the oxides Fe₂O₃, TiO₂, CaO, and MgO. This characteristic elemental and oxide assemblage is specifically indicative of orthomagmatic systems. Within these systems, the critical process for sulfide mineralization is the sulfur saturation of the magma. This triggers the segregation of an immiscible sulfide liquid, which acts as a highly efficient collector, specifically scavenging chalcophile metals due to their strong affinity for sulfur. This dense, sulfide-rich liquid subsequently undergoes gravitational settling and accumulation. Consequently, the resulting sulfide mineralization in these orthomagmatic deposits is unequivocally syngenetic, forming directly as a product of the primary magmatic differentiation process.

Factor-3: A strong presence of TiO₂, K₂O, Rb, Hg, REE (Rare Earth Elements), and Th in an igneous rock indicates it has undergone significant magmatic differentiation, leading to an enrichment of incompatible elements—those that concentrate in the residual melt during crystallization. Rocks most likely to exhibit this geochemical signature are highly evolved, felsic igneous rocks:

- Granites and Rhyolites (especially per-alkaline varieties): These are commonly enriched in K₂O, Rb, REE, and Th due to extensive fractional crystallization, with per-alkaline types being particularly rich in incompatible elements and often hosting exotic minerals.
- Pegmatites: As the very last, coarse-grained products of granitic magma crystallization, they are well-known for concentrating incompatible elements like REE, Th, and sometimes Hg.
- Carbonatites: Though less common and with variable silica/potassium, these unique igneous rocks are also notable for significant REE and Th enrichment.

Factor-4: Na₂O (sodium oxide), Au (gold), and Hg (mercury) don't typically have direct, strong chemical associations in most natural geological settings but their co-occurrence and indirect relationships arise from shared geological processes, especially those involving hydrothermal fluids and mineralizing events, and in some cases, anthropogenic activities.

The most common geological condition is Hydrothermal Systems and Mineralization where Na as a highly mobile element can move through hydrothermal fluid and replace the existing minerals (like potassium feldspar or plagioclase of a different composition) by albite (sodium feldspar) during Sodic alteration and may associate with gold and mercury mineralization. This association also indicates the occurrences of Iron Oxide Copper-Gold (IOCG) deposits which is characterized by voluminous Na-Ca-K metasomatism (alteration) and often contain elevated Cu, Au, REE, and other elements.

In summary, Sodium-rich fluids can be responsible for altering the rocks that subsequently host gold mineralization, and mercury can be a volatile component of those same fluid systems, migrating further and acting as a guide to the gold.

The normalised factor maps are given in Figure 5.12.

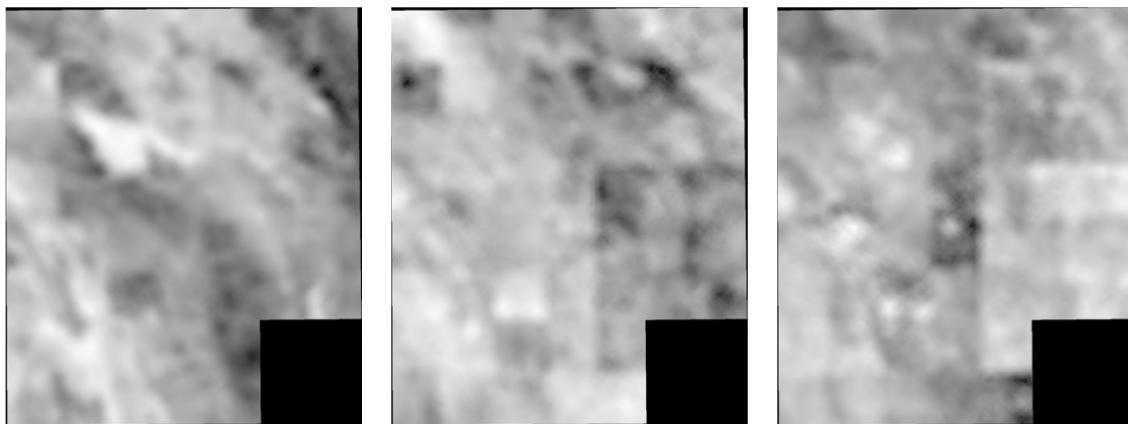


Figure 5.12: Normalized map of PCA1, PCA3 and PCA4.

5.4.4 Remote Sensing data

Processed ASTER data has been used to generate a) normalised chloritic, b) silica, c) ferrous oxide alteration zone maps (Figure 5.13).

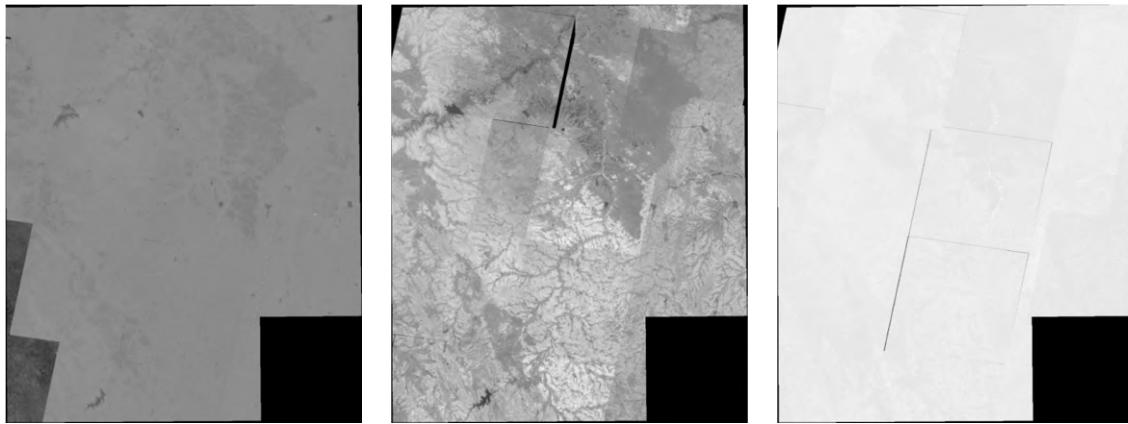


Figure 5.13: Normalized alteration maps

5.4.5 Aerogeophysical Spectrometric Data

The Aerogeophysical Spectrometric Data has been used for preparation of normalised a) Dose-Total Count map, b) K map.

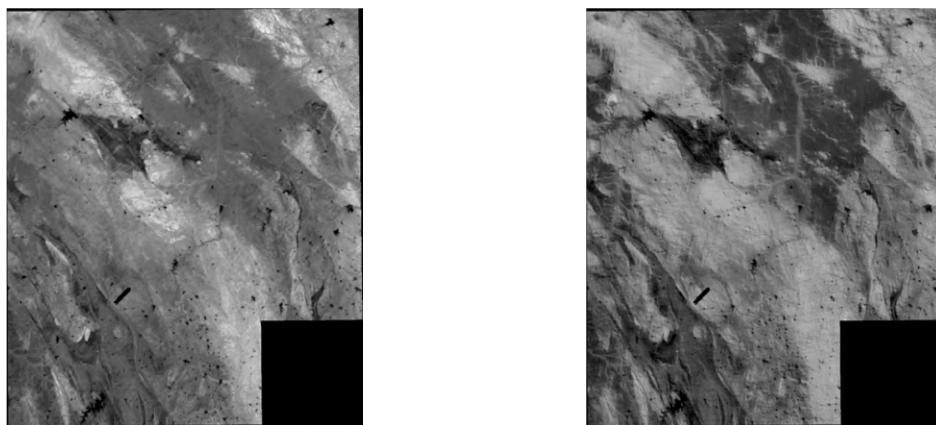


Figure 5.14: Normalized Dose TC and K map

6. Methodology

6.1. Overall Workflow (Flowchart):

The entire process, from the initial acquisition of raw geo-scientific data to the generation of the final mineral prospectivity maps, is systematically outlined in the flowchart presented below (Figure 6.1). This diagram illustrates the sequential steps involved in data processing, integration, machine learning model development, and visualization of results, highlighting the interdependencies of various stages.

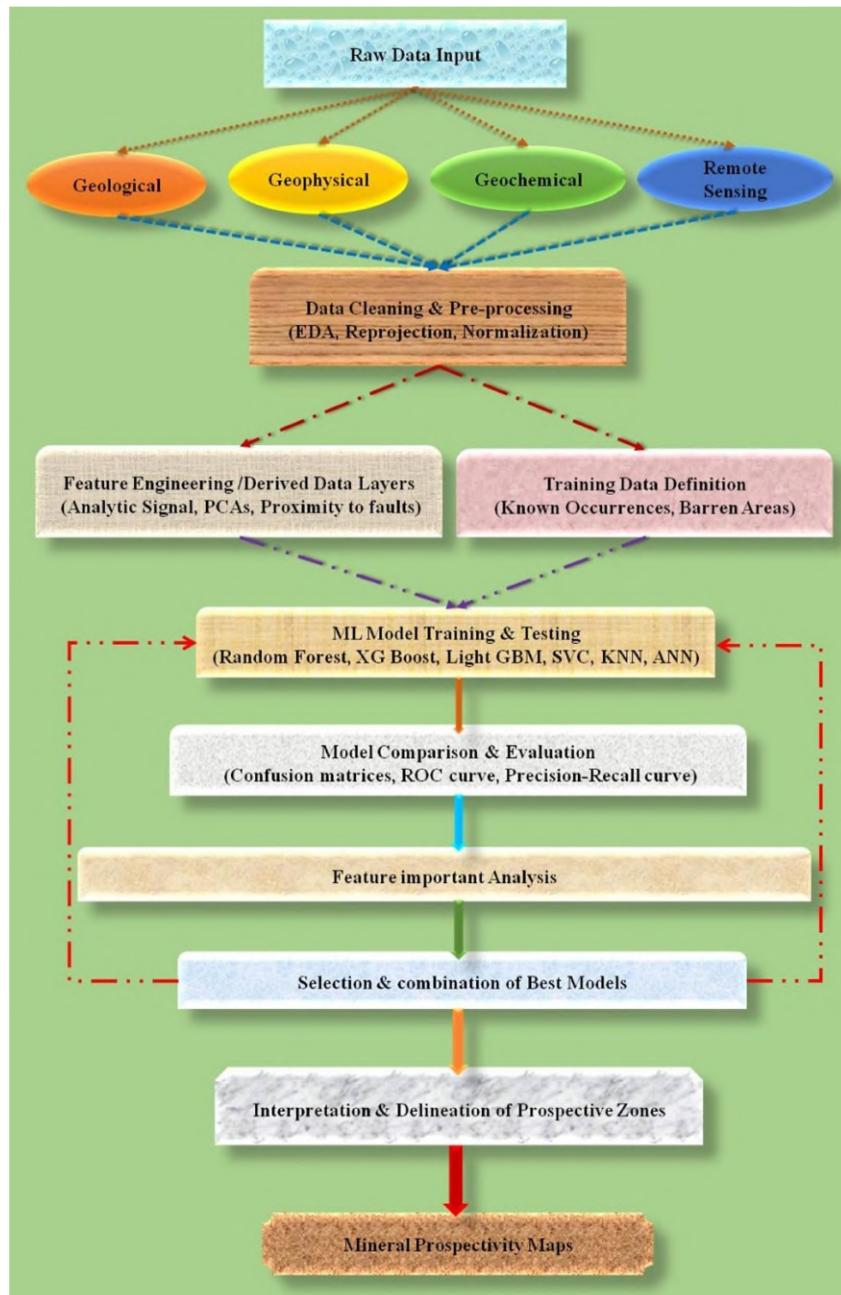


Figure 6.1: Comprehensive workflow illustrating the stages of data processing, machine learning model development, and output generation for mineral prospectivity mapping.

6.2. Data Processing and Curation:

Effective data processing and curation are foundational to the success and reliability of any machine learning model, especially in complex geoscience applications. This stage involved a meticulous series of steps to transform raw, disparate geo-scientific datasets into a clean, consistent, and analysis-ready format suitable for feature engineering and machine learning. The primary objectives were to ensure data quality, spatial and thematic harmonization, and appropriate scaling for model input.

6.2.1. Data Cleaning and Pre-processing

This initial phase focused on identifying and rectifying errors, inconsistencies, and redundancies within and across various datasets. Key activities included:

- **Initial Data Assessment (QA/QC):** A thorough review of each dataset was conducted to identify missing values, corrupted records, data entry errors, and inconsistencies in naming conventions (e.g., lithology codes, element units). This involved statistical summaries and visual inspection of data distributions.
- **Spatial Data Integrity:** Vector datasets (e.g., geological features, fault lines, litho contacts) were checked for topological errors such as dangles, overlaps, or slivers. Raster datasets (e.g., geophysical grids, satellite imagery) were inspected for geometric distortions or data gaps. Erroneous geometries were corrected or flagged.
- **Handling Missing Values:** Strategies for addressing missing data were implemented based on the nature and extent of the missingness. For instance, for continuous data with limited missing values, imputation techniques such as mean, median, or K-Nearest Neighbors (KNN) imputation were considered. For categorical data, mode imputation or creation of a 'Missing' category was applied. In cases of extensive missingness, certain features or records might have been excluded if their informational value was deemed insufficient.

6.2.2. Spatial Data Handling and Harmonization

Integrating multi-source, multi-format geoscientific data necessitated robust spatial handling to ensure all layers were geospatially aligned and compatible.

- **Common Coordinate Reference System (CRS) Projection:** All datasets, regardless of their original projection, were re-projected to a unified Universal Transverse Mercator (UTM) Zone [e.g., 43N] with the WGS84 datum. This ensured precise spatial overlay and consistent distance measurements across all layers.
- **Clipping to Area of Interest (AOI):** Datasets extending beyond the defined 39,000 sq. km study area were clipped to the precise boundaries of the AOI to remove extraneous data and focus computational efforts.
- **Rasterization and Resampling:** Vector datasets (e.g., geological polygons, fault lines) were converted to raster formats with a common cell size (e.g. 200m) to enable integration with grid-based geophysical and remote sensing data. Similarly, raster datasets with varying resolutions were resampled to this common grid cell size using appropriate interpolation methods (e.g., bilinear for continuous data, nearest neighbour for categorical data).
- **Thematic Harmonization:** For categorical data (e.g., lithology units from different geological maps), consistency checks were performed to harmonize categories where aliases or slight variations existed (e.g., 'Granite' vs. 'Granite_gneiss' unified to 'Granite').

6.2.3. Outlier Detection and Handling

Outliers, while sometimes representing true anomalies, can significantly bias machine learning models. A balanced approach was adopted to identify and manage them.

- **Statistical Methods:** For numerical data, methods such as the Interquartile Range (IQR) rule (values falling outside $Q1 - 1.5 \times IQR$ and $Q3 + 1.5 \times IQR$) were applied. Z-scores were also used to identify extreme values relative to the mean. This is adopted mainly for geochemical dataset (NGCM).
- **Visual Inspection:** Box plots, scatter plots, and histograms were extensively used to visually identify unusual data points or distributions that might indicate outliers or errors.
- **Domain Expertise Review:** Identified outliers, particularly in geochemical or geophysical data, were cross-referenced with geological knowledge and known measurement errors to ascertain if they represented genuine anomalies (e.g., a high gold value at a known prospect) or data errors.
- **Handling Strategy:** Depending on their nature and impact, outliers were transformed (e.g., logarithmic transformation for highly skewed geochemical data), or, in rare cases of clear data errors, removed.

6.2.4 Raster Data Standardization: Min-Max Normalization

Prior to integration into the unified dataset for machine learning, all continuous raster datasets were subjected to Min-Max normalization. This process scales the values of each raster band to a common range, typically between 0 and 1.

- **Purpose:** The primary objective of normalization is to standardize the range of independent variables. In multi-thematic geospatial analysis, different data layers (e.g., magnetic intensity, elevation, geochemical concentrations) often have vastly different units and scales. Without normalization, algorithms sensitive to magnitude (such as those using distance calculations or gradient descent) might implicitly assign higher importance to features with larger numerical ranges, regardless of their actual predictive power.
- **Methodology:**

Data Loading and Resampling: The raster data was loaded using `rasterio`. If the original resolution differed from the target resolution (e.g., 200m x 200m), the raster was resampled to the consistent target resolution using bilinear interpolation. Bilinear interpolation is suitable for continuous data as it calculates the new pixel value based on a weighted average of the four nearest pixel values.

NoData Value Handling: Original NoData values (e.g., $-3.4028234663852886e+38$) were identified and replaced with `np.nan` to ensure they are correctly excluded from normalization calculations and maintained as missing data throughout the workflow.

Min-Max Scaling: The non-NaN pixel values of the raster were then scaled using the Min-Max normalization formula: $X_{\text{normalized}} = (X - X_{\min}) / (X_{\max} - X_{\min})$ where X is the original pixel value, X_{\min} is the minimum non-NaN value in the raster, and X_{\max} is the maximum non-NaN value in the raster. If all non-NaN values in a raster were identical, the normalized values were set to 0.0 (while preserving NaN values).

Output Generation: The normalized raster data was saved as a new georeferenced GeoTIFF file (e.g., `norm_Silica.tif`), ensuring its spatial integrity. Additionally, for visual inspection and tabular representation, a PNG image (e.g., `norm_Silica.png`) and a CSV file (`norm_Silica.csv`) containing the normalized pixel values and their coordinates were generated.

- **Significance:** This systematic normalization ensures that all input features are on a comparable scale, which is essential for the optimal performance of the machine learning algorithms. It prevents features with naturally larger values from dominating the model's learning process and allows the model to learn true underlying relationships based on the relative variation within each feature.

6.3. Feature Engineering and Derived Data Layers:

Feature engineering is a critical step in preparing multi-thematic geoscientific data for machine learning. It involves transforming raw data into meaningful and predictive features that can capture the subtle signatures of mineralization. This process enhances the information content of the input data, enabling the machine learning models to effectively learn the complex relationships indicative of mineral deposits. All derived raster layers were subsequently normalized using Min-Max scaling (as detailed in Section 6.2.4) to ensure comparable ranges for model input.

6.3.1. List of Derived Features:

The following features were engineered and derived from the primary geological, geochemical, geophysical, and remote sensing datasets.

From Geological Data: (section 5.4.1 and Figure 5.1)

- Normalized Lithology feature, specific to target commodity. Label Encoding is deployed considering knowledge driven scoring of lithology.
- Normalized Group member feature specific for target commodity based on knowledge.
- Normalized Lineament Density, defines density map of lineaments.
- Normalized Lineament Multi Ring buffer, for influence of lineaments.
- Normalized Lineament intersection Density, defines density map of intersection points.
- Normalized Dyke vein buffer.
- Normalized Quartz Vein buffer.

From Geochemical Data (NGCM):

- Normalized Principal Component 1 from geochemical data.
- Normalized Principal Component 3 from geochemical data.
- Normalized Principal Component 4 from geochemical data.
- Normalized elemental distribution map of U from geochemical data.
- Normalized elemental distribution map of Th from geochemical data.

From Geophysical Data:

- Normalized Bouguer Anomaly.
- Normalized Bouguer Anomaly Residual.
- Normalized Vertical Derivative of Bouguer Anomaly Residual.
- Normalized Tilt Derivative of Bouguer Anomaly.

- Normalized Upward Continuation (500m) of Bouguer Anomaly.
- Normalized Total Magnetic Anomaly.
- Normalized Magnetic Analytic Signal.
- Normalized Magnetic Residual.
- Normalized Magnetic RTP Residual.
- Normalized Magnetic RTP.
- Normalized Upward Continuation (500m) of Total Magnetic Anomaly.

From PGRS Data (ASTER):

- Normalized Chlorite alteration index.
- Normalized Iron Oxide alteration index.
- Normalized Silica alteration index.

From Aerogeophysical Spectrometric Data:

- Normalized Dose index.
- Normalized K index.

6.4. Training Data Selection and Preparation:

The effectiveness of a supervised machine learning model hinges critically on the quality and representativeness of its training data. This section details the systematic approach employed to define and prepare the training dataset, comprising both positive (mineralized) and negative (non-mineralized) samples, and the strategies adopted to address potential class imbalance.

6.4.1. Definition:

The creation of the training dataset involved the careful selection and labelling of data points representing either known mineralized areas (positive samples) or demonstrably barren areas (negative samples) within the study region.

Positive Samples (Mineralized Areas):

- **Sources:** Positive samples were primarily derived from comprehensive databases of known mineral occurrences, prospects, and operating mines for the targeted commodities (REE, Ni-PGE, Copper, Diamond, Iron, Manganese, Gold) within Karnataka and Andhra Pradesh. These databases typically included details such as deposit type, principal commodities, location coordinates, and, where available, geological context.
- **Spatial Representation:** Given that mineral occurrences are often point-based or represent small, defined areas, these point occurrences were converted into a spatial representation suitable for the grid-based modeling approach. This typically involved:
- **Buffer Zones:** Creating small buffer zones (e.g., 500m radius, depending on the scale of mineralization) around known occurrences to represent the immediate prospective footprint. These buffered areas were labeled as positive.
- **Expert Delineation:** In certain well-studied areas, expert geological interpretation was used to delineate known mineralized footprints more accurately, which were then rasterized to define positive grid cells.

- **Filtering/Refinement:** To ensure the quality of positive samples, records with ambiguous location data or very low confidence in mineralization were excluded. Emphasis was placed on confirmed occurrences to minimize noise in the positive class.

Negative Samples (Non-Mineralized / Barren Areas):

- **Selection Criteria:** Identifying unequivocally barren areas is often more challenging than identifying positive ones. Negative samples were strategically selected from regions believed to be non-prospective for the targeted mineralization based on multiple lines of evidence:
- **Absence of Known Mineralization:** Areas with no recorded mineral occurrences despite extensive exploration activity or regional geological mapping.
- **Unfavourable Geology:** Regions characterized by geological units or structural settings considered highly unfavourable for the genesis of the targeted mineral systems.
- **Geophysical/Geochemical Background:** Areas exhibiting clear background geophysical or geochemical signatures that are inconsistent with known mineralization types (e.g., extremely low magnetic susceptibility where high susceptibility is a key indicator, absence of pathfinder elements).
- **Spatial Distribution:** Care was taken to ensure negative samples were spatially distributed across the study area and did not cluster exclusively around known deposits, thereby allowing the model to learn broader regional patterns.
- **Random Sampling with Constraints:** A common approach involved random sampling of points or grid cells from the non-prospective areas, while ensuring these samples did not overlap with positive samples or their buffer zones.
- **Verification:** Selected negative samples were cross-referenced with available geological maps and geophysical data to minimize the inadvertent inclusion of potentially mineralized or geologically ambiguous areas.

6.4.2. Handling Imbalance:

In mineral exploration, the number of known mineral occurrences (positive samples) is typically orders of magnitude smaller than the vast non-mineralized areas (negative samples). This inherent class imbalance can significantly bias machine learning models, leading them to favour the majority class (non-mineralized) and perform poorly in identifying the rare positive class. To mitigate this challenge, following techniques were applied:

- **Evaluation Metric Focus:** Initially, standard evaluation metrics were adjusted to prioritize the performance on the minority class. Metrics such as Recall (Sensitivity), Precision, F1-score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC) were emphasized over simple accuracy, as they provide a more robust assessment in imbalanced scenarios.
- **Undersampling of the Majority Class:** A portion of the majority class (non-mineralized areas) samples were randomly or strategically removed. While this can balance the dataset, it risks discarding potentially valuable information from the majority class. This was considered when the majority class was overwhelmingly large and computational efficiency was a concern.
- **Algorithm-Level Techniques - Class Weighting:** For algorithms that support it (e.g., Logistic Regression, Support Vector Machines, Random Forests, Gradient Boosting), higher weights were assigned to the minority class during model training. This tells the

model to penalize misclassifications of positive samples more heavily, forcing it to pay closer attention to these rare occurrences.

The specific combination of techniques employed was determined empirically, based on cross-validation performance and the desired balance between identifying new prospects (recall) and minimizing false positives (precision). This careful management of training data ensured that the machine learning model was effectively trained to identify the subtle signatures of mineralization within the vast non-mineralized background.

6.5. Machine Learning Model Development:

The core of this project involves leveraging machine learning algorithms to systematically identify and delineate potential mineralized areas within the defined study region. This stage encompasses the definition of the problem, the selection and justification of the chosen model, the strategy for training and validating the model, and the optimization of its performance through hyperparameter tuning.

6.5.1. Problem Formulation:

The mineral prospectivity mapping problem was formulated as a binary classification task. The primary objective was to classify each spatial unit (e.g., grid cell) within the study area as either 'mineralized' (positive class, typically represented as 1) or 'non-mineralized' (negative class, typically represented as 0). This approach allows for the direct identification of areas exhibiting geological, geophysical, and geochemical signatures indicative of potential mineralization. The selected models learn to discriminate between these two classes based on the integrated geoscientific features derived from the multi-thematic input data.

6.5.2. Model Selection:

A comparative approach was adopted for model selection, evaluating a diverse suite of machine learning algorithms to identify the most effective performer for mineral prospectivity mapping. The selection was guided by the need for models robust to complex, heterogeneous geoscientific data, capable of handling potential non-linear relationships, and providing reliable predictive performance.

The following models were initialized and evaluated:

- **Logistic Regression:** A fundamental linear model providing interpretability, often serving as a strong baseline.
- **Decision Tree Classifier:** A non-linear model capable of capturing complex decision boundaries, useful for understanding feature interactions.
- **Random Forest Classifier:** An ensemble learning method based on decision trees. It was chosen for its robustness to noisy data, ability to handle high-dimensional features, reduced overfitting compared to single trees, and the capability to provide feature importance scores, which are crucial for understanding the geological drivers of prospectivity. The specified `max_depth=4` was set to control complexity and prevent overfitting.
- **LightGBM (Light Gradient Boosting Machine):** A gradient boosting framework that uses tree-based learning algorithms. Known for its high efficiency and speed, particularly on large datasets, while maintaining high accuracy.

- **XGBoost (Extreme Gradient Boosting):** Another powerful and popular gradient boosting framework, recognized for its speed, performance, and regularization capabilities that help prevent overfitting. The scale_pos_weight parameter was explicitly set to address class imbalance, emphasizing the positive class.
- **SVC (Support Vector Classifier):** A powerful discriminative classifier that works well in high-dimensional spaces, even with limited data. The radial basis function (rbf) kernel was used to capture non-linear relationships, and probability=True was set to allow for prospectivity score estimation.
- **K-Nearest Neighbors (KNN):** A simple, non-parametric instance-based learning algorithm that classifies based on the majority class among its nearest neighbours.
- **ANN (MLP - Multi-layer Perceptron):** A foundational type of artificial neural network. It was included to explore the potential of deep learning to capture highly complex, non-linear patterns and interactions within the geoscientific dataset. The architecture (hidden_layer_sizes=(100, 50)), activation function (relu), and optimization (adam) were specified, along with early_stopping for regularization.

All classification models, where applicable, were initialized with class_weight='balanced' to automatically adjust weights inversely proportional to class frequencies, directly addressing the inherent class imbalance commonly observed in mineral prospectivity datasets. For XGBoost, scale_pos_weight was explicitly calculated and applied for this purpose.

6.5.3. Training & Validation

The training and validation of the selected machine learning models were conducted using a single train-test split approach (70 : 30). This involved partitioning the pre-processed and harmonized dataset into distinct training and testing subsets.

The training set (X_train_scaled, y_train) was used to fit each model, allowing the algorithms to learn the underlying patterns and relationships between the geoscientific features and the presence or absence of mineralization.

The testing set (X_test_scaled, y_test), which consisted of data unseen by the models during training, was then used to evaluate their predictive performance. This provides an independent assessment of how well each model generalizes to new, unobserved spatial locations.

6.5.4. Hyperparameter Tuning:

The initial selection of hyperparameters for each model was based on common best practices, established defaults, and an understanding of their typical performance in similar geoscientific classification tasks. For instance, n_estimators=100 for Random Forest and max_depth=4 were chosen to provide a balance between model complexity and interpretability. Similarly, specific layer sizes and optimization settings were chosen for the ANN. In this phase, the models were trained and evaluated with these pre-defined hyperparameter settings. The chosen parameters for each model represent a baseline configuration.

The raster files used for modelling of each commodity are tabulated below:

Layer	Au	Cu	Ni-PGE	Fe-Mn	REE
normalized_Lithology_REE.tif					
normalized_Lithology_NiCrPGE.tif					
normalized_Lithology_FeMn.tif					
normalized_Lithology_Cu.tif					
normalized_Lithology_Au.tif					
normalized_Group_Au_Cu_Fe-Mn.tif					
normalized_Group_Ni-PGE.tif					
normalized_Group_REE.tif					
normalized_Line_density.tif					
normalized_Line_ring.tif					
normalized_intersect.tif					
normalized_Boundary_Intru_Granite-Schist Belt.tif					
normalized_pca1.tif					
normalized_pca3.tif					
normalized_pca4.tif					
normalized_vein.tif					
normalized_dyke.tif					
normalized_BA.tif					
normalized_BA_res.tif					
normalized_BA_res_vdr.tif					
normalized_BA_tdr.tif					
normalized_BA_up5.tif					
normalized_MA.tif					
normalized_MA_ana.tif					
normalized_MA_res.tif					
normalized_MA_up5.tif					
normalized_MA_RTP.tif					
normalized_MA_RTP_res.tif					
normalized_chl.tif					
normalized_Silica.tif					
normalized_feo.tif					
normalized_mf.tif					
norm_Dose_TC.tif					
norm_K.tif					
norm_U.tif					
norm_Th.tif					
Au_target.tif					
NiCrPGE_target.tif					
FeMn_target.tif					
Cu_target.tif					
REE_target.tif					

7. Results and Discussion

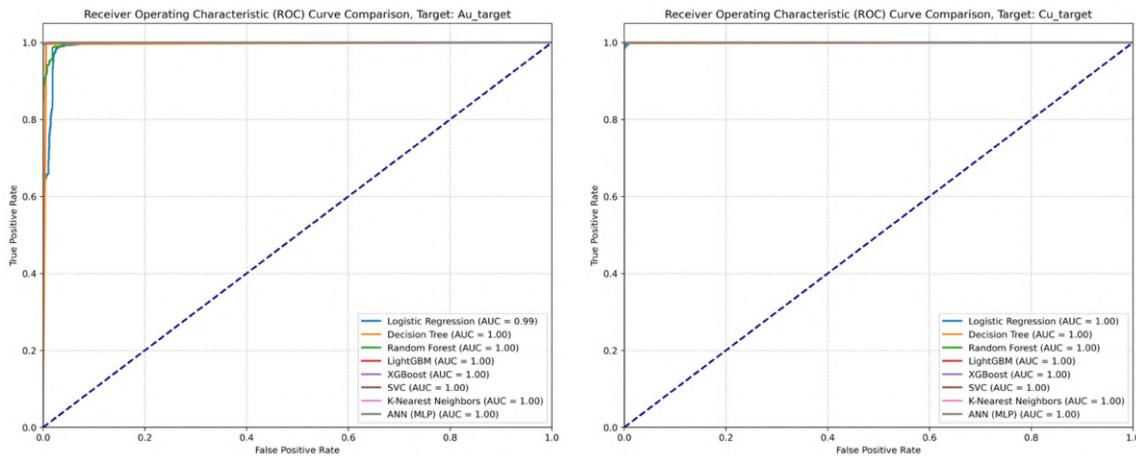
7.1. Model Performance and Evaluation:

To comprehensively assess the performance of the various machine learning models and facilitate a comparative analysis, a suite of visualization techniques was employed. These plots provide critical insights into model accuracy, discriminatory power, and behaviour, particularly concerning the identification of the minority positive class (mineralization).

7.1.1. Receiver Operating Characteristic (ROC) Curve Comparison

The Receiver Operating Characteristic (ROC) curve is a fundamental tool for evaluating the performance of classification models, particularly when dealing with imbalanced datasets. It illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied.

- **Purpose:** Each ROC curve plots the True Positive Rate (TPR, or Recall) against the False Positive Rate (FPR) at various threshold settings.
- **Interpretation:** A model with perfect discrimination would have a point at the top-left corner (0 FPR, 1 TPR). The closer the curve is to the top-left corner, the better the model's performance. The area under the ROC curve (AUC-ROC) quantifies this performance, with an AUC of 1.0 representing a perfect classifier and 0.5 representing a random classifier.
- **Significance in MPM:** In mineral prospectivity mapping, a high AUC-ROC indicates that the model is effectively distinguishing between mineralized and non-mineralized areas across different probability thresholds. This helps in selecting models that are robust in identifying potential targets while minimizing false alarms.



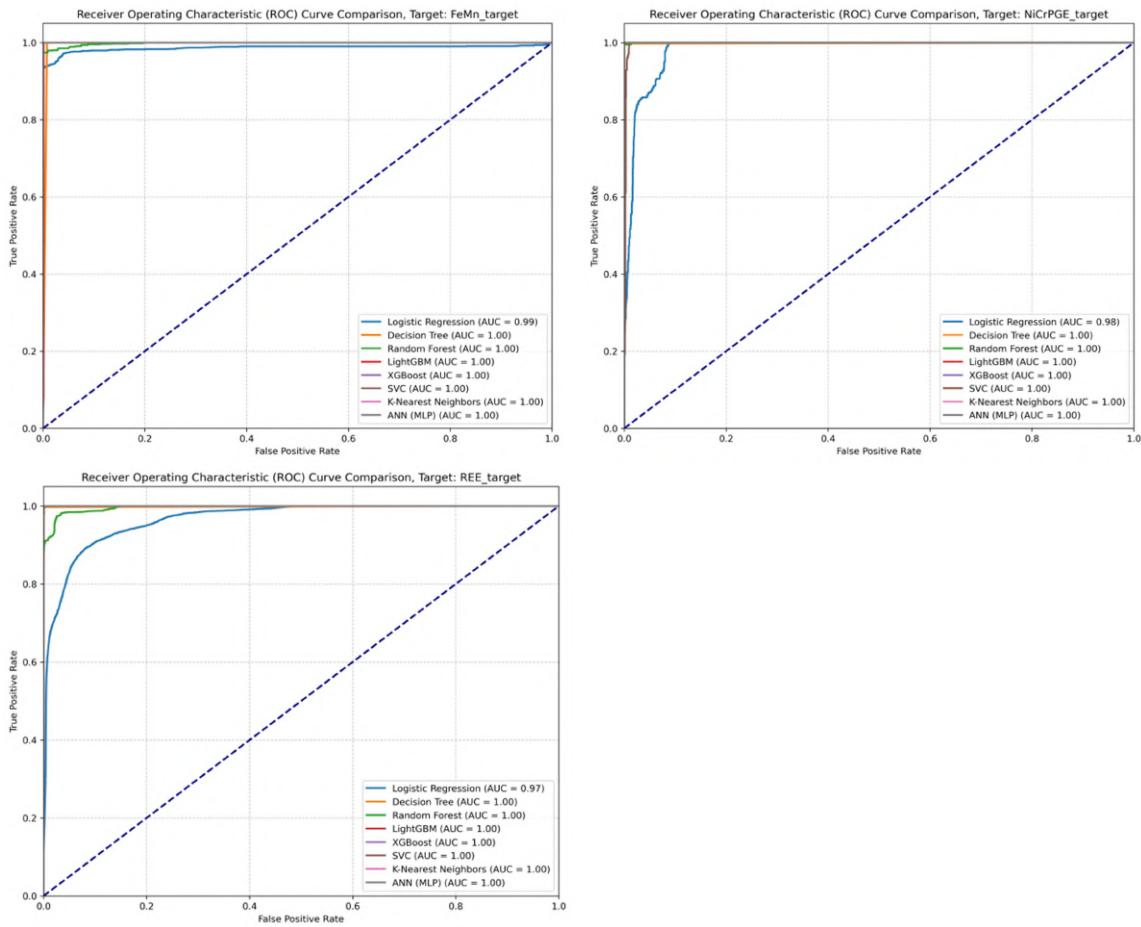


Figure 7.1: Comparative ROC curves for all evaluated machine learning models, demonstrating their ability to distinguish between mineralized and non-mineralized classes for prospectivity mapping.

7.1.2. Precision-Recall Curve Comparison

While the ROC curve is useful, the Precision-Recall (PR) curve is often more informative for highly imbalanced datasets, such as those typically encountered in mineral exploration where the positive class (mineralization) is rare.

- **Purpose:** The PR curve plots Precision (the proportion of positive identifications that were actually correct) against Recall (the proportion of actual positives that were identified correctly) at various classification thresholds.
- **Interpretation:** A high area under the Precision-Recall curve (Average Precision - AP) indicates that the model is able to find a high proportion of positive examples while making few false positive errors. The curve shows the trade-off between identifying all possible prospects and ensuring that identified prospects are genuinely mineralized.
- **Significance in MPM:** For mineral exploration, a strong PR curve is crucial as it directly relates to the efficiency of follow-up exploration efforts. A model with high average precision means that field validation of its predicted targets is more likely to yield true mineralization, saving significant resources.

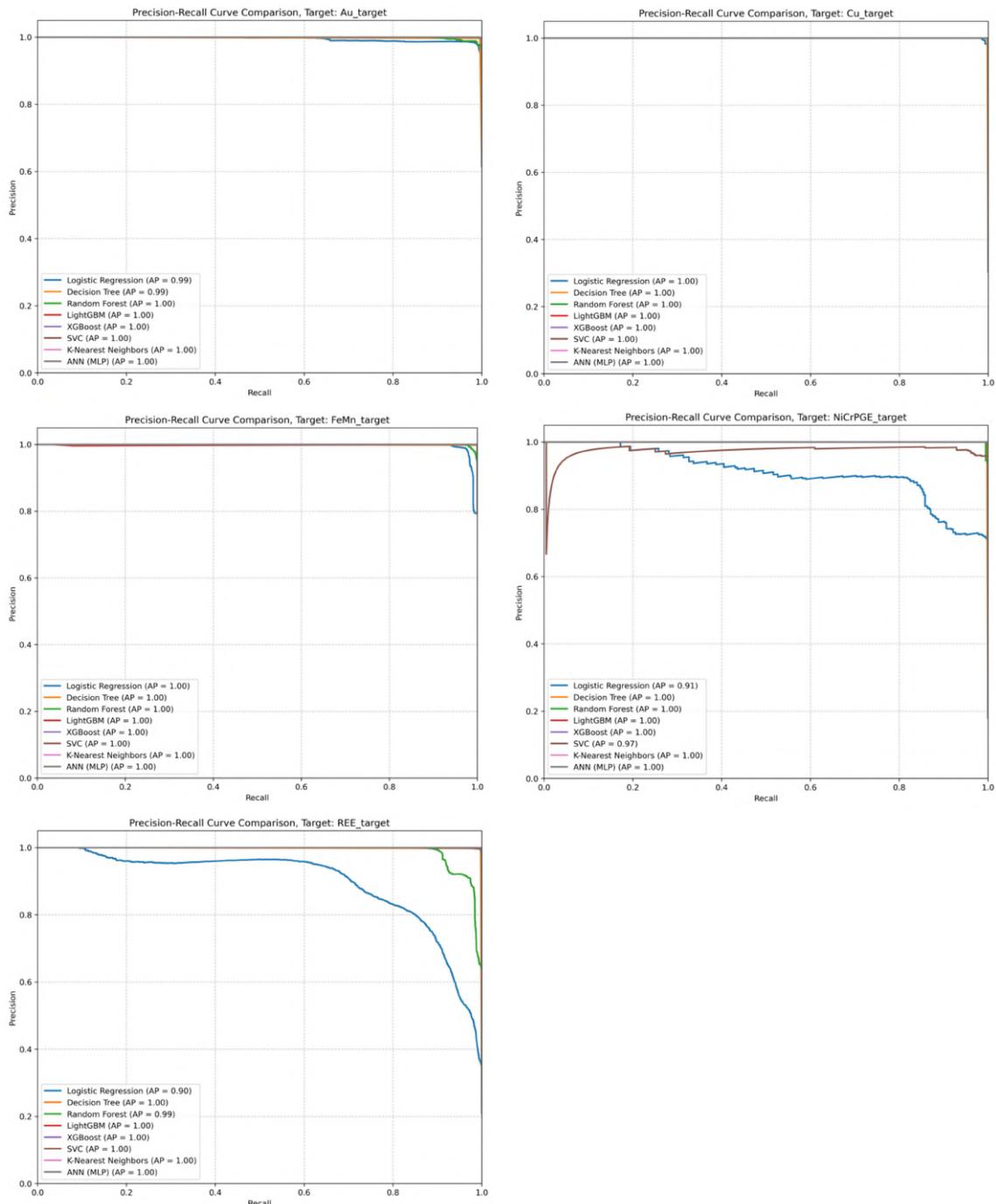


Figure 7.2: Comparative Precision-Recall curves for all evaluated machine learning models, highlighting their performance in identifying the positive (mineralized) class, particularly relevant for imbalanced datasets.

7.1.3. Comparison of Key Metrics (Bar Plot)

To provide a concise overview of each model's overall performance across several critical metrics, a bar plot was generated comparing Accuracy, Precision, Recall, F1-Score, and AUC-ROC.

- **Purpose:** This visualization allows for a quick side-by-side comparison of how each model performs on different aspects of classification.

- **Accuracy:** Overall correctness of predictions.
- **Precision:** Ability to avoid false positives (important for minimizing wasted exploration efforts).
- **Recall (Sensitivity):** Ability to find all actual positives (important for not missing potential deposits).
- **F1-Score:** Harmonic mean of precision and recall, providing a balanced measure.
- **AUC-ROC:** Overall discriminatory power.
- **Significance in MPM:** This plot facilitates the selection of the most suitable model by weighing the importance of different metrics. For instance, a high recall might be prioritized in early-stage exploration to ensure no potential targets are missed, even at the cost of some false positives.

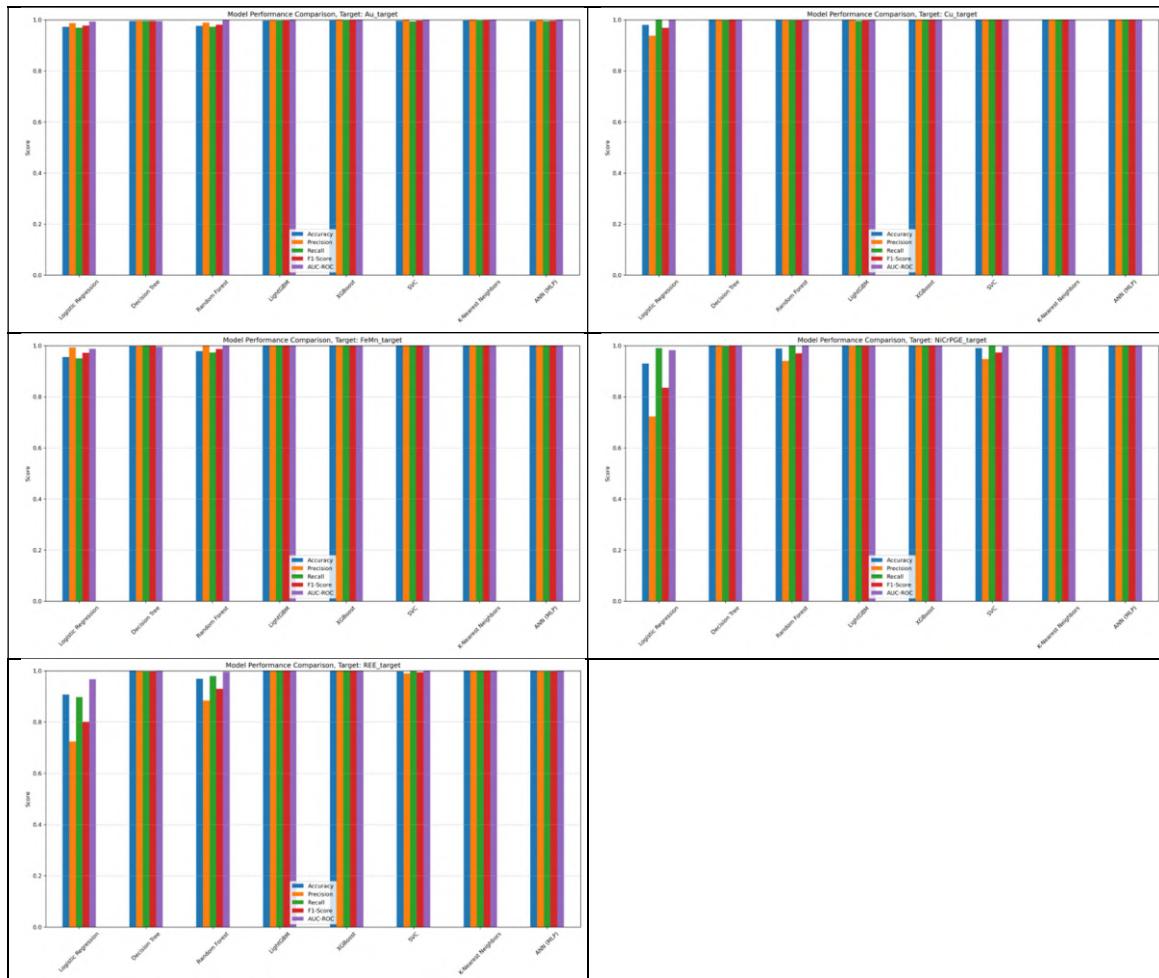


Figure 7.3: Bar plot comparing key performance metrics (Accuracy, Precision, Recall, F1-Score, and AUC-ROC) for all trained machine learning models.

7.1.4. Confusion Matrices

Confusion matrices provide a detailed breakdown of the model's classification performance on the test set for each class.

- **Purpose:** For each model, a confusion matrix illustrates the number of True Positives (correctly identified mineralized areas), True Negatives (correctly identified barren areas),

False Positives (barren areas incorrectly predicted as mineralized), and False Negatives (mineralized areas incorrectly predicted as barren).

- **Interpretation:** These matrices directly show the types of errors each model is making. High False Negatives indicate missed opportunities, while high False Positives suggest inefficient targeting.
- **Significance in MPM:** Analysing confusion matrices is crucial for understanding the practical implications of a model's errors. For example, a model with a very low number of False Negatives (high recall for the positive class) might be preferred in reconnaissance stages to ensure comprehensive coverage, even if it leads to some False Positives.



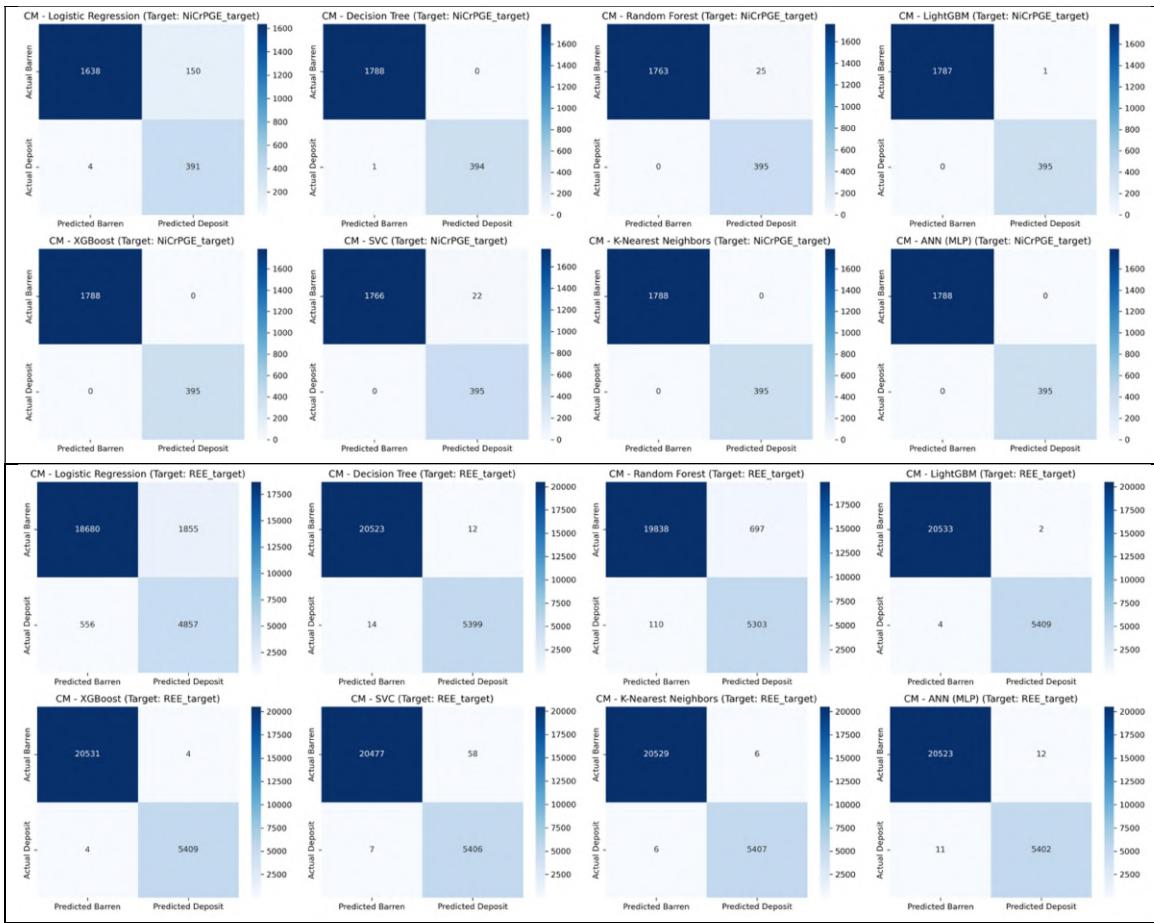


Figure 7.4: Individual confusion matrices for each machine learning model, detailing true positive, true negative, false positive, and false negative counts on the test set.

7.2. Feature Importance and Layer Contribution Analysis:

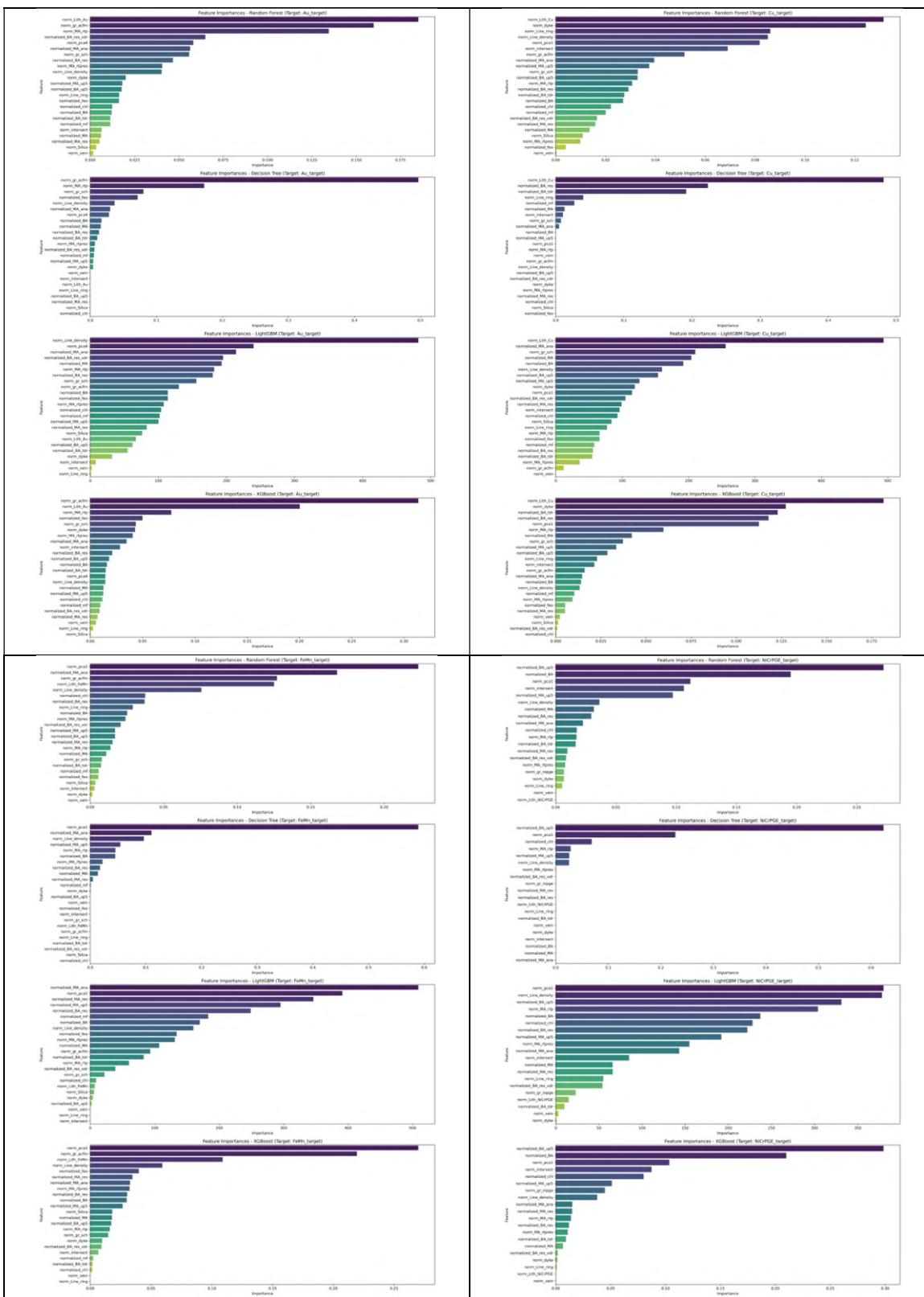
Understanding which input features (geological, geophysical, geochemical layers) are most influential in the model's predictions is vital for geological interpretation and validating the conceptual genetic model.

7.2.1. Feature Importance Plots (for Tree-Based Models)

For ensemble tree-based models such as Random Forest, Decision Tree, LightGBM, and XGBoost, feature importance scores are extracted directly from the trained models.

- Purpose:** These plots visually represent the relative contribution or importance of each input feature (derived data layer) to the model's predictive decisions.
- Interpretation:** Higher importance scores indicate that a feature had a greater impact on the model's ability to accurately classify mineralized versus non-mineralized areas. The specific metric used (e.g., Gini importance for Random Forest) quantifies how much each feature reduces impurity or error across all trees in the ensemble.
- Significance in MPM:** This analysis provides valuable geological insights by identifying which geo-scientific proxies are most strongly correlated with mineralization within the study area. This can help confirm or refine the conceptual genetic model, prioritize data acquisition in future exploration, and highlight the most promising indicators for gold prospectivity. For instance, if norm_Lith_Au.tif and normalized_MA_ana.tif consistently

show high importance, it suggests lithology and magnetic anomaly features are key drivers of gold mineralization in the region.



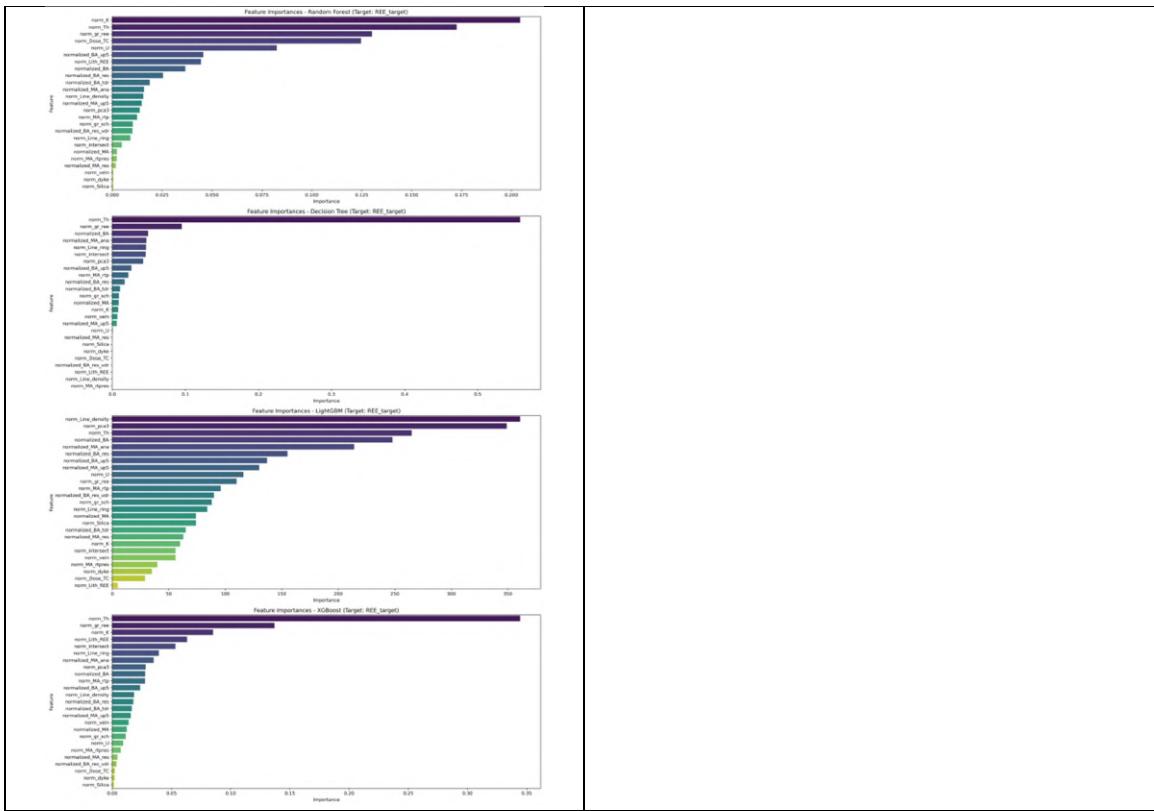


Figure 7.5: Bar plots illustrating the relative importance of each geoscientific feature as determined by the tree-based machine learning models (Random Forest, Decision Tree, LightGBM, and XGBoost).

7.3. Predictive Maps:

The ultimate outcome of the machine learning modelling process is the generation of predictive maps that delineate areas with high mineral prospectivity. These maps serve as critical tools for guiding future exploration efforts by visually representing the spatial distribution of potential mineralization. Beyond individual model predictions, a consensus-based approach was employed to identify robust, high-confidence targets.

7.3.1. Delineation of New Prospective Zones (Combined Map)

To identify the most compelling and robust exploration targets, a consensus-driven approach was implemented using the top-performing models identified in the model evaluation phase (Section 7.1). This strategy aims to reduce false positives by requiring multiple independent models to agree on an area's prospectivity.

- **Top Model Selection:** The top three machine learning models were selected based on manually evaluation of their overall performance, feature importance criteria, prioritizing a high F1-Score, which provides a balanced measure of precision and recall. These top models were different for different commodities and marked in Figures 7.6 to 7.10.
- **Consensus Filtering:** Prospectivity scores (prediction probabilities for the positive class) were generated across the entire study area using each of these top three models. A strict threshold (e.g. 0.85 for gold) was then applied, only locations where all three of the top models predicted a prospectivity score greater than this threshold were considered for the final combined map. This high-confidence threshold significantly reduces the risk of

false positive predictions, focusing on areas with strong, corroborating evidence from multiple algorithms.

- **Average Prospectivity Score:** For the locations that met this stringent consensus criterion, an average prospectivity score was calculated from the scores of the three selected models. This average score provides a consolidated measure of confidence for these highly prospective areas.
- **Spatial Representation:** The filtered high-confidence locations, along with their average prospectivity scores, were then used to generate a Combined Mineral Prospectivity Map. This map visually highlights areas where multiple advanced algorithms collectively indicate a strong likelihood of mineralization.

This combined map specifically delineates new prospective zones hitherto unexplored or unreported, identified purely through the data-driven approach of the machine learning models. These zones represent high-priority targets for follow-up ground-truthing and detailed exploration.

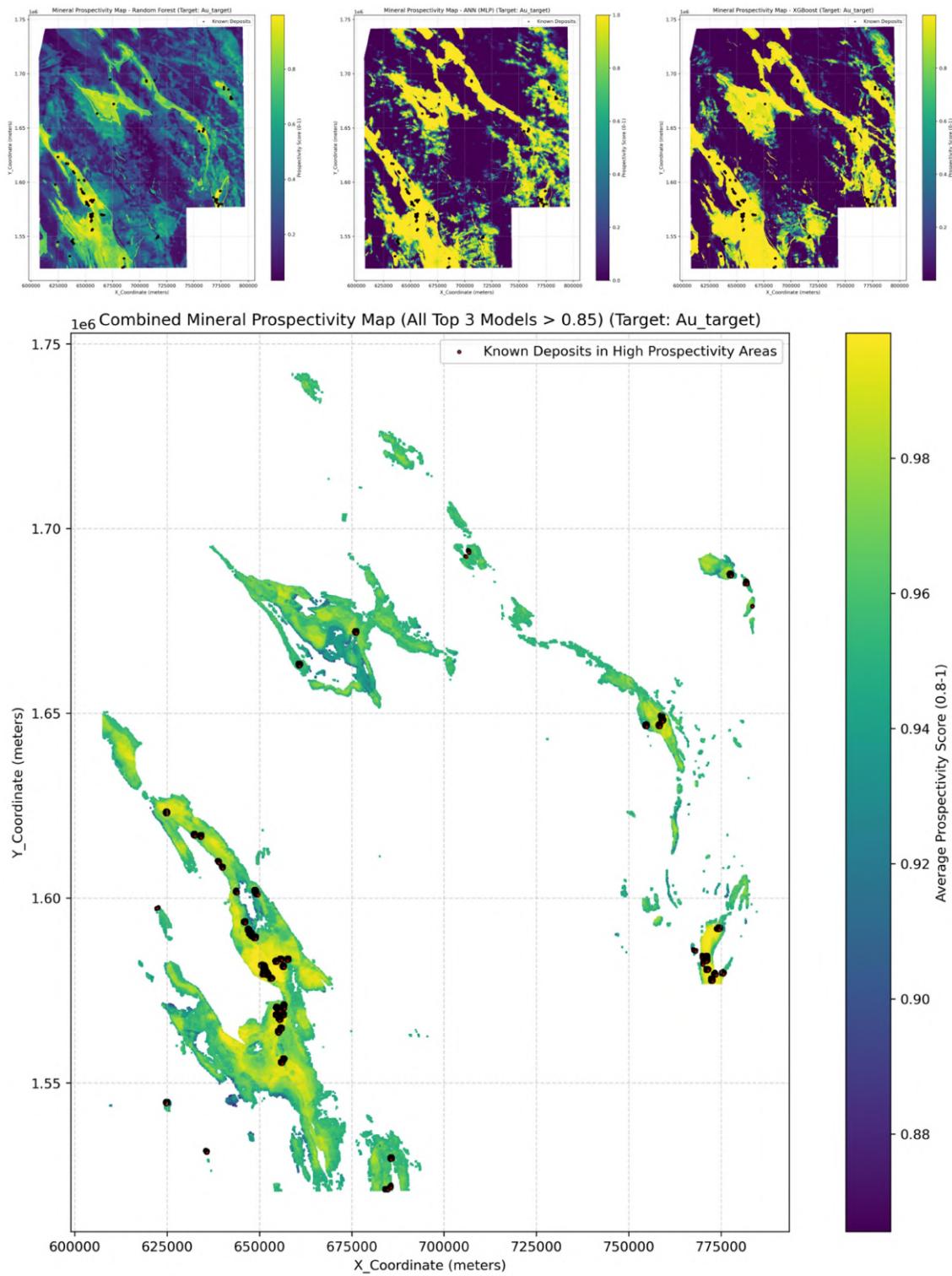


Figure 7.6: Combined prospectivity map, highlighting high-confidence target areas for gold (Au) mineralization. These zones represent locations where all top three performing machine learning models predicted a prospectivity score greater than 0.85.

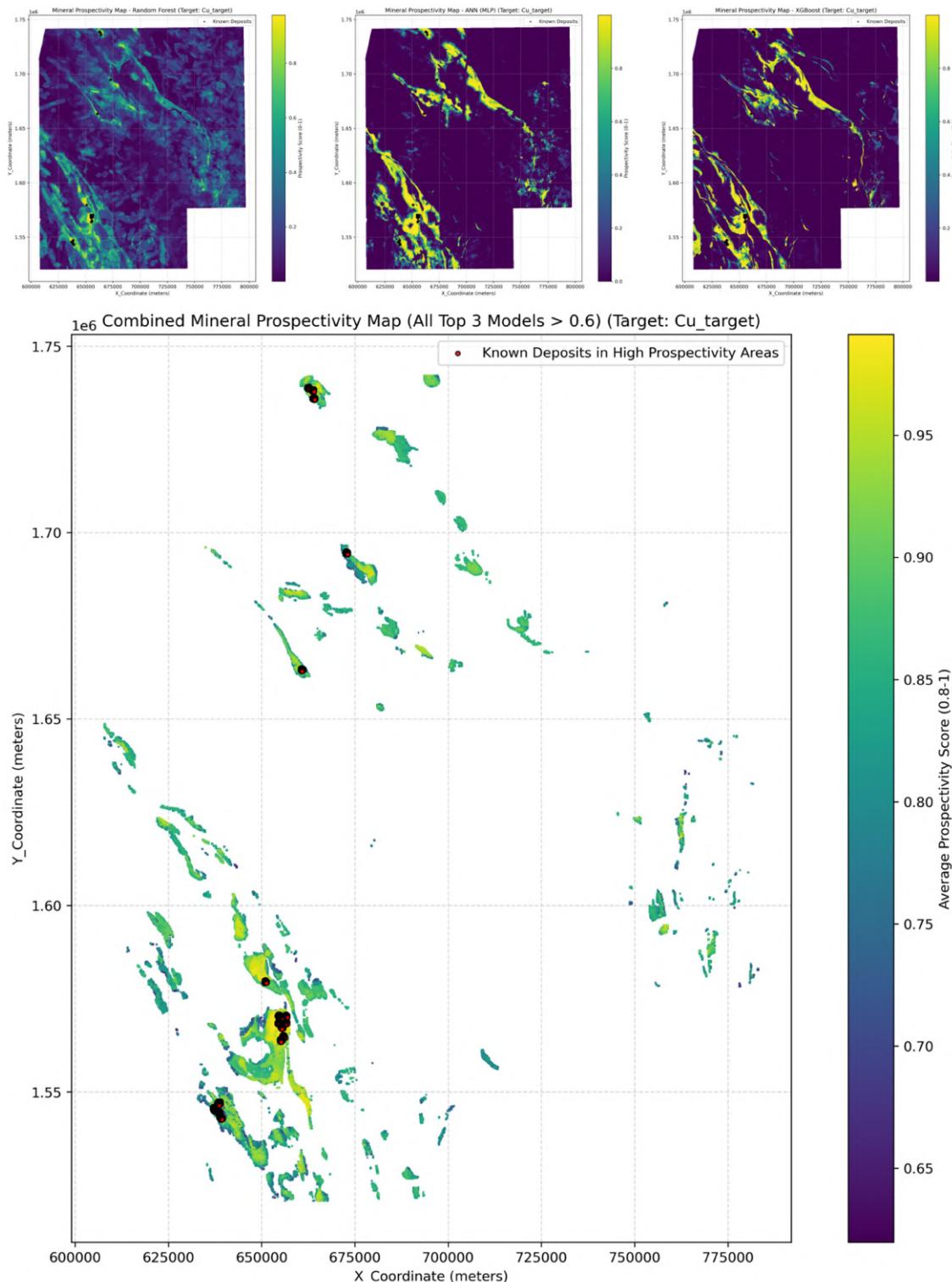


Figure 7.7: Combined prospectivity map, highlighting high-confidence target areas for copper (Cu) mineralization. These zones represent locations where all top three performing machine learning models predicted a prospectivity score greater than 0.60.

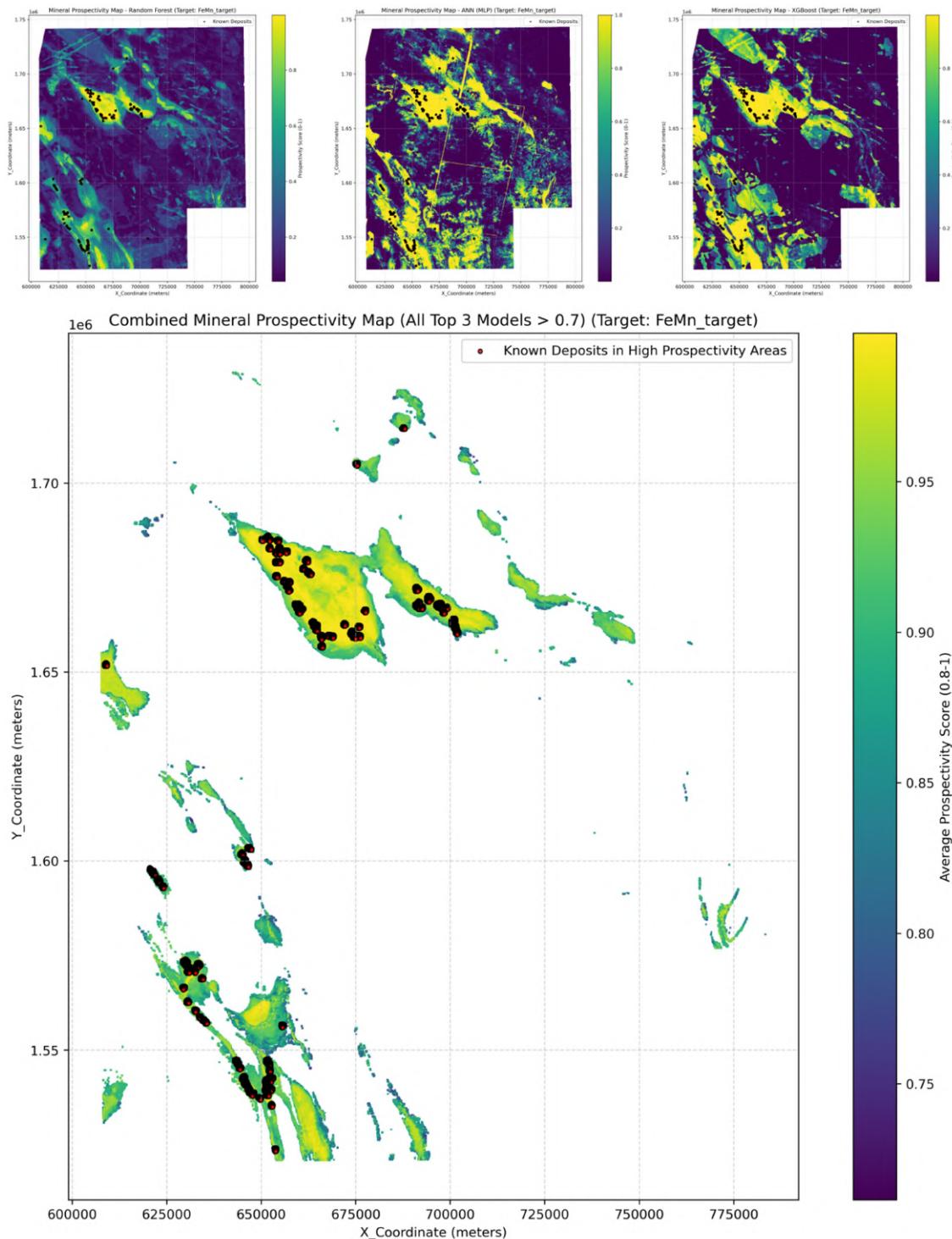


Figure 7.8: Combined prospectivity map, highlighting high-confidence target areas for iron-manganese (Fe-Mn) mineralization. These zones represent locations where all top three performing machine learning models predicted a prospectivity score greater than 0.70.

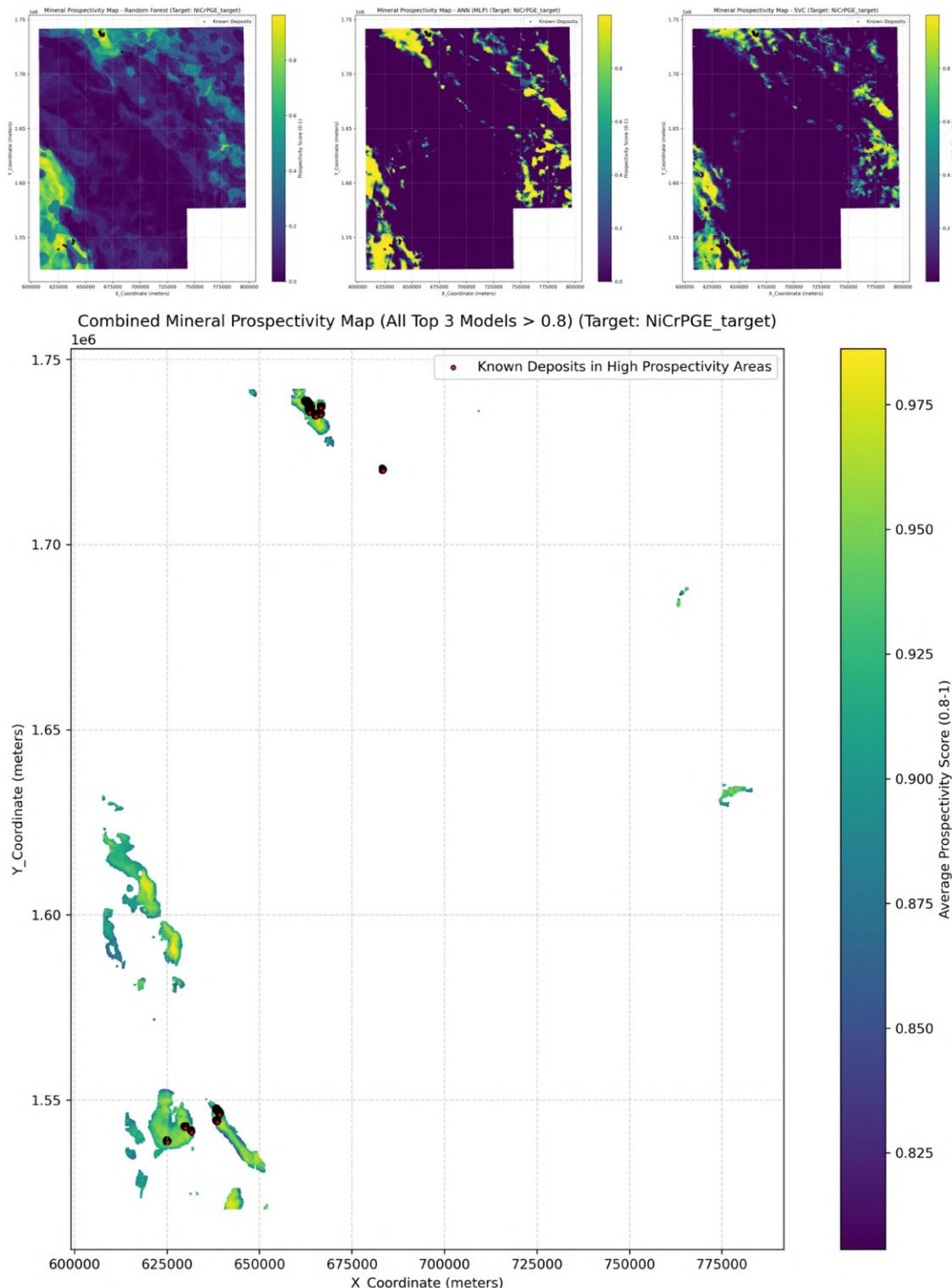


Figure 7.9: Combined prospectivity map, highlighting high-confidence target areas for nickel - PGE (Ni - PGE) mineralization. These zones represent locations where all top three performing machine learning models predicted a prospectivity score greater than 0.80.

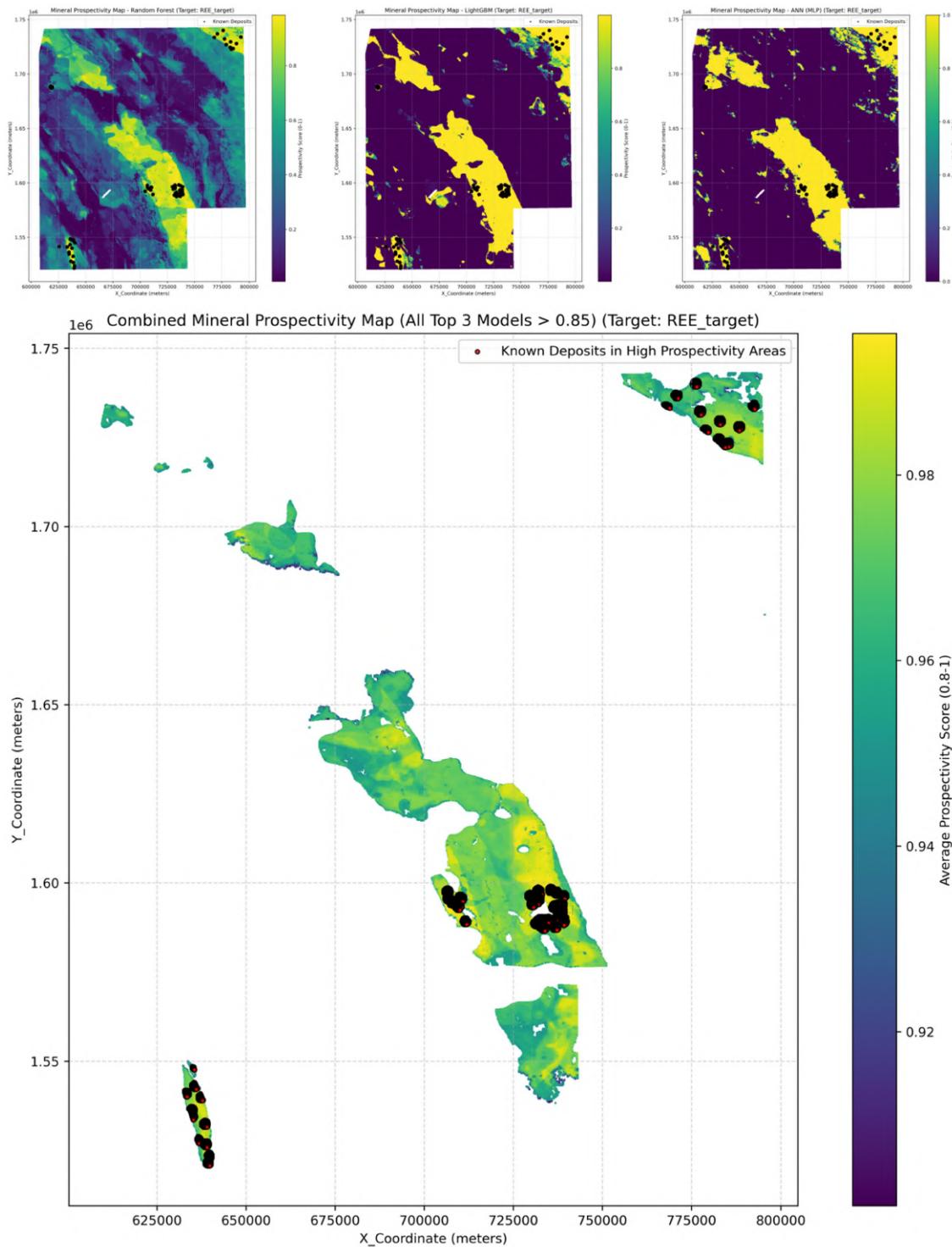


Figure 7.10: Combined prospectivity map, highlighting high-confidence target areas for rare earth elements (REE) mineralization. These zones represent locations where all top three performing machine learning models predicted a prospectivity score greater than 0.85.

8. Conclusions and Recommendations

8.1. Summary of Key Findings:

This project successfully leveraged a multi-thematic geoscientific dataset and advanced machine learning algorithms to delineate new mineral prospective zones within the approximately 39,000 sq. km study area across Karnataka and Andhra Pradesh, India. The most important outcomes and key findings are summarized as follows:

Successful Identification of New Prospective Zones: The machine learning models effectively identified and delineated a total of 67 new prospective blocks across the targeted critical and other high-value commodities. These identified zones represent areas with high geological favourability based on integrated geoscientific data, moving beyond previously known occurrences.

For Gold (Au), 26 new prospective blocks were identified, covering an area of approximately 1504 Sq. Km. For Copper (Cu), 23 new prospective blocks were delineated, spanning approximately 832 Sq. Km. For Iron-Manganese (Fe-Mn), 22 new prospective blocks were delineated, encompassing approximately 1056 Sq. Km. For Nickel-Platinum Group Elements (Ni-PGE), 6 new prospective blocks were found, totaling 240 Sq. Km. For Rare Earth Elements (REE), 8 new prospective blocks were identified, covering an area of approximately 1220 Sq. Km. These quantitative results highlight the project's capability to significantly expand the known potential for mineral exploration.

Robustness of the ML Approach: The application of multiple machine learning algorithms (especially Random Forest, LightGBM, XGBoost, SVC, K-Nearest Neighbors and ANN) demonstrated the versatility and robustness of the AI/ML approach for mineral prospectivity mapping. The comparative analysis of these models using metrics like AUC-ROC, F1-Score, and Precision-Recall curves (as discussed in Section 7.1) allowed for the selection of top-performing models and a high-confidence consensus approach (Section 7.3.1), which significantly reduced false positives and yielded reliable predictive maps. The systematic data processing, cleaning, and normalization procedures further ensured the quality and consistency of the input data for the models.

	Au	Cu	Ni-PGE	REE	Fe-Mn
No. Of Blocks	26	23	6	8	22
Prospective area Identified (in Sq. Km.)	1504	832	240	1220	1056

The prospectivity maps along with proposed block boundary are given in Figure 8.1. The block boundary has been marked over high prospectivity score zones.

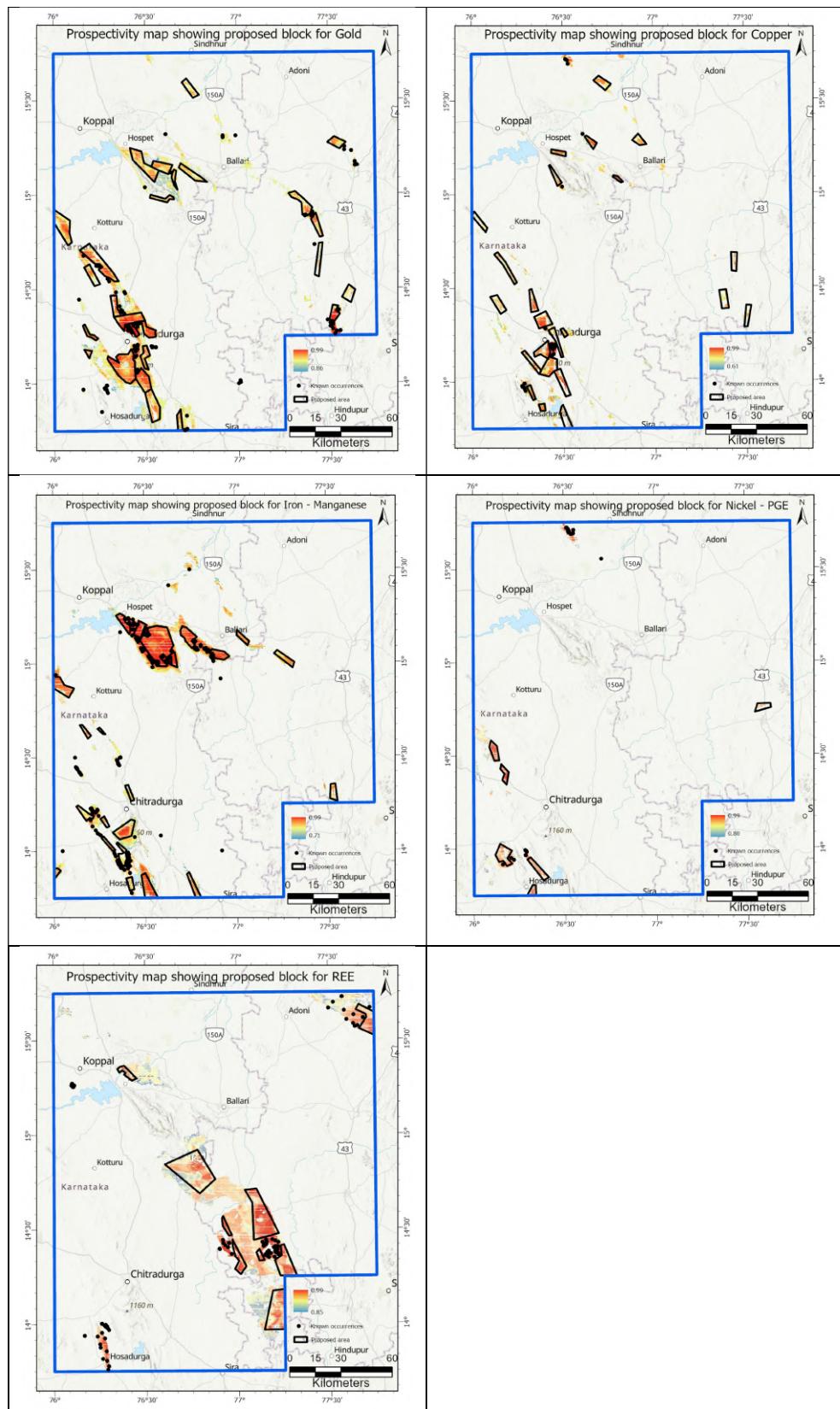


Figure 8.1 Prospectivity map showing proposed blocks for gold, copper, iron-manganese, Ni-PGE and REE

8.2. Implications for Mineral Exploration:

The predictive maps and identified prospective zones generated through this machine learning-driven mineral prospectivity mapping project have significant implications for guiding future mineral exploration activities within the Karnataka and Andhra Pradesh study area. These results provide a data-driven and spatially explicit framework for optimizing exploration strategies, moving beyond traditional subjective methods. Specifically, the outcomes of this project can:

- **Prioritize Areas for Follow-up Ground Surveys:** The high-confidence prospective zones delineated by the combined model approach serve as immediate priority areas for detailed ground-based geological and geochemical surveys. Resources can be efficiently allocated to these specific areas, reducing the vast search space of the 39,000 sq. km region. This targeted approach increases the likelihood of discovering new mineralized bodies by focusing efforts on areas with the most favorable predicted signatures.
- **Identify Potential Drilling Targets:** Within the highly prospective zones, localized high-score anomalies, especially those corroborated by multiple models and potentially by conceptual 3D models (if available), can be ranked as potential drilling targets. The predictive maps provide the spatial context for strategically locating drill holes, particularly for concealed and deep-seated mineralization, minimizing blind drilling and maximizing the chances of intersecting mineralization.
- **De-risk Exploration Decisions:** By leveraging multi-thematic geoscientific data and advanced machine learning algorithms, the project significantly de-risks early-stage exploration decisions. The quantitative nature of the prospectivity scores provides a robust basis for investment and planning, enabling more informed decision-making regarding where to commit further exploration capital.
- **Highlight Concealed Potential:** The methodology's emphasis on integrating diverse datasets and machine learning algorithms (e.g., those capable of handling complex patterns) has successfully identified concealed and potentially deep-seated mineralized bodies. This is crucial in mature exploration terrains where near-surface deposits may already have been discovered, opening new frontiers for exploration.
- **Optimize Data Acquisition:** Insights from feature importance analysis (Section 7.2.1) can guide future data acquisition strategies. Geoscientific datasets and derived features identified as highly influential in predicting mineralization can be prioritized for more detailed collection or higher resolution surveys in new areas. This ensures that future data collection efforts are focused on variables with proven predictive power.

Ultimately, this project demonstrates the transformative potential of AI/ML in accelerating mineral discovery by providing an objective, reproducible, and spatially comprehensive assessment of mineral prospectivity, leading to more efficient and successful exploration campaigns.

8.3. Future Work and Enhancements:

While this project has successfully demonstrated the application of machine learning for mineral prospectivity mapping, several avenues for future work and enhancements can further improve the model's robustness, predictive power, and practical utility:

8.3.1 Incorporating New and Higher-Resolution Data Types:

- **Drill Core Assays and Lithology:** Integration of comprehensive quantitative drill hole assay data and detailed lithological logs (beyond just presence/absence) could significantly improve the resolution and accuracy of prospectivity predictions, particularly for depth modelling.
- **High-Resolution Geophysical Surveys:** Incorporating high-resolution ground or airborne geophysical surveys (e.g., detailed magnetics, gravity, IP/resistivity) over prioritized areas would provide more localized and precise information, refining the spatial prediction of targets.
- **Advanced Remote Sensing Products:** Utilizing hyperspectral or LiDAR data could provide more nuanced information on alteration patterns, structural features, and subtle surface expressions of mineralization.

8.3.2. Testing Advanced ML Algorithms and Architectures:

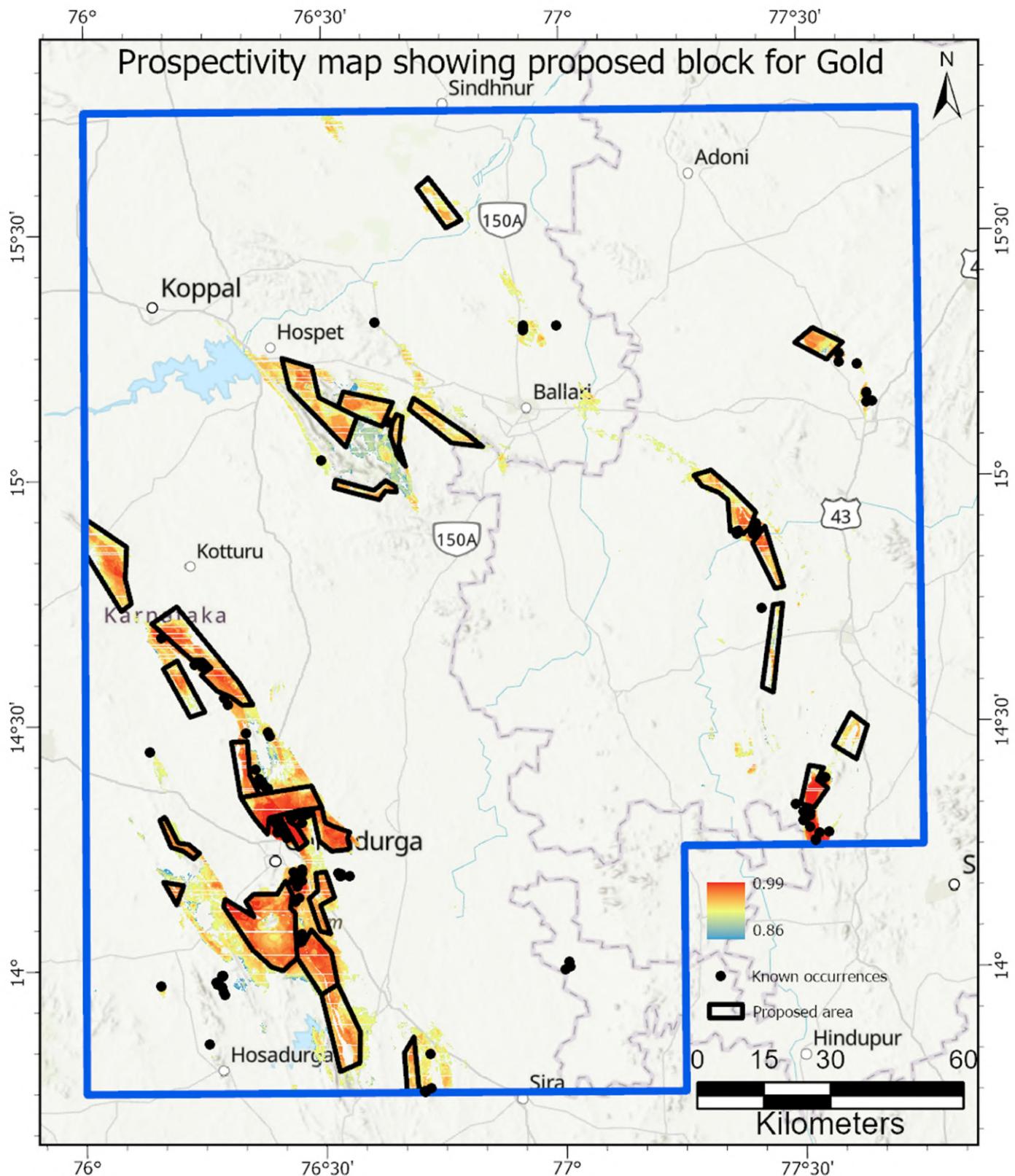
- **Deep Learning (DL) for Geospatial Data:** Exploring more advanced DL architectures, such as Convolutional Neural Networks (CNNs) for direct image-based feature extraction from raster data, or Graph Neural Networks (GNNs) for modelling geological networks (faults, veins), could potentially capture more intricate spatial relationships.
- **Ensemble Learning Refinement:** Investigating more sophisticated ensemble methods (e.g., stacking, boosting with custom loss functions) or active learning strategies for optimal model combination.

9. References

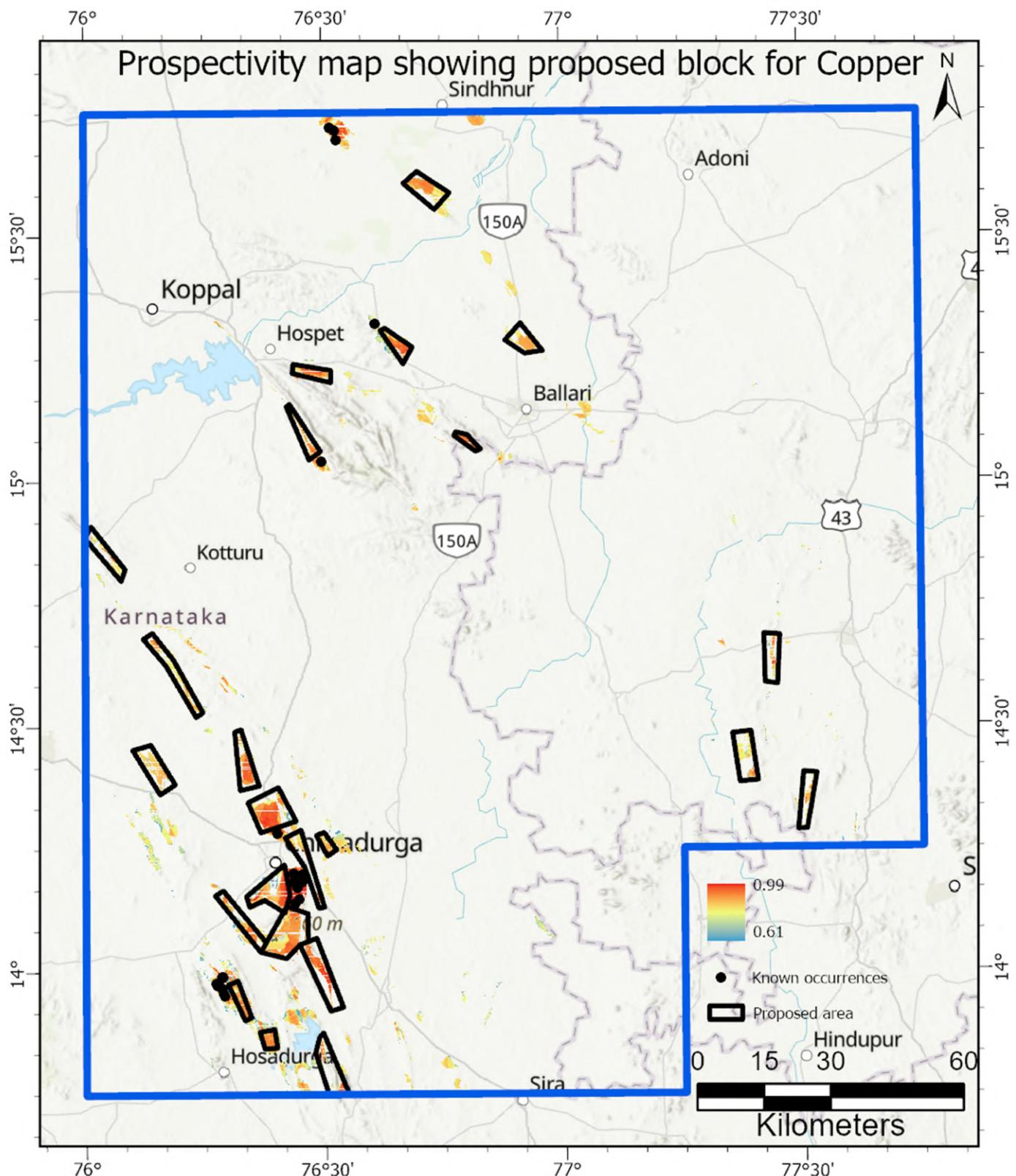
- Aggarwal, S. (2021). Litho-structural and geochemical characteristics of uraniferous basement rocks and gulcheru quartzite in and around veldurthi area along western margin of the cuddapah basin, kurnool district, a.p. [Master's Thesis, Homi Bhabha National Institute]. Bhabha Atomic Research Centre, Mumbai.
- Bhattacharyya, B. K. (1967). Some general properties of potential fields in space and frequency domain. *Geoexploration*, 5, 127-143.
- Bhaskar Rao, Y. J., Chetty, T. R. K., & Narayana, B. L. (2003). Geochronology of the Sargur Group, Dharwar Craton, southern India: constraints on its antiquity and correlation. *Journal of Asian Earth Sciences*, 22(3), 295-309.
- Bhaskar Rao, Y. J., Narayana, B. L., Chetty, T. R. K., & Narayana, S. (2000). Geochronology of the Closepet granite, Dharwar Craton, southern India: implications for its tectonic setting. *Journal of Asian Earth Sciences*, 18(6), 683-698.
- Chadwick, B., Ramakrishnan, M., & Vasudev, V. N. (2000). The Dharwar Craton. Geological Society, London, Special Publications, 173(1), 329-351.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417–441. <https://doi.org/10.1037/h0071325>
- Hsu, S. K. (2002). Imaging Magnetic sources using Euler's equation. *Geophysical Prospecting*, 50, 15–25.
- Jacobsen, B. H. (1987). A case for upward continuation as a standard separation filter for potential-field maps. *Geophysics*, 52(8), 1138–1148. <https://doi.org/10.1190/1.1442378>
- Maus, S., & Dimri, V. P. (1996). Depth estimation from the scaling power spectrum of potential fields? *Geophysical Journal International*, 124(1), 113–120.
- Michael, N. A., & Craigie, N. W. (2021). Automations in chemostratigraphy: Toward robust chemical data analysis and interpretation. In SPE Middle East Oil and Gas Show and Conference (MEOS), Virtual. D041S049R004.
- Mishra, B. (2015). Chapter 22 Precambrian metallic mineralization in India. In Geological Society, London, Memoirs (Vol. 43, pp. 327-337). Geological Society of London. <https://doi.org/10.1144/M43.22>
- Nabighian, M. N. (1972). The analytic signal of two-dimensional magnetic bodies with polygonal cross-section: its properties and use for automated anomaly interpretation. *Geophysics*, 37(5), 891-909.
- Naqvi, S. M., & Rogers, J. J. W. (1987). Precambrian Geology of India. Oxford University Press.
- Peucat, J. J., Pin, C., Vidal, P., Bernard-Griffiths, J., & Krogstad, E. J. (1993). U-Pb zircon geochronology of the Peninsular Gneiss Complex, southern India: an overview. *Journal of the Geological Society of India*, 42(3), 253-268.
- Ramakrishnan, M., & Vaidyanadhan, R. (2010). Geology of India (Vol. 1). Geological Society of India.

- Reid, A. B., Allsop, J. M., Granser, H., Millet, A. J., & Somerton, I. W. (1990). Magnetic interpretation in three dimensions using Euler deconvolution. *Geophysics*, 55(1), 173-181.
- Salem, A., & Ravat, D. (2003). A combined analytic signal and Euler method (AN-EUL) for automatic interpretation of Magnetic (TF) data. *Geophysics*, 68(6), 1952–1961.
- Scheibe, C., & Michael, N. A. (2023). The Aid of Principal Components and Discriminant Functions Logs in Chemostratigraphy. 84th EAGE Annual Conference & Exhibition 2023, 1, 1-5.
- Spector, A., & Grant, F. S. (1970). Statistical models for interpreting aeromagnetic data. *Geophysics*, 35(2), 293-302.

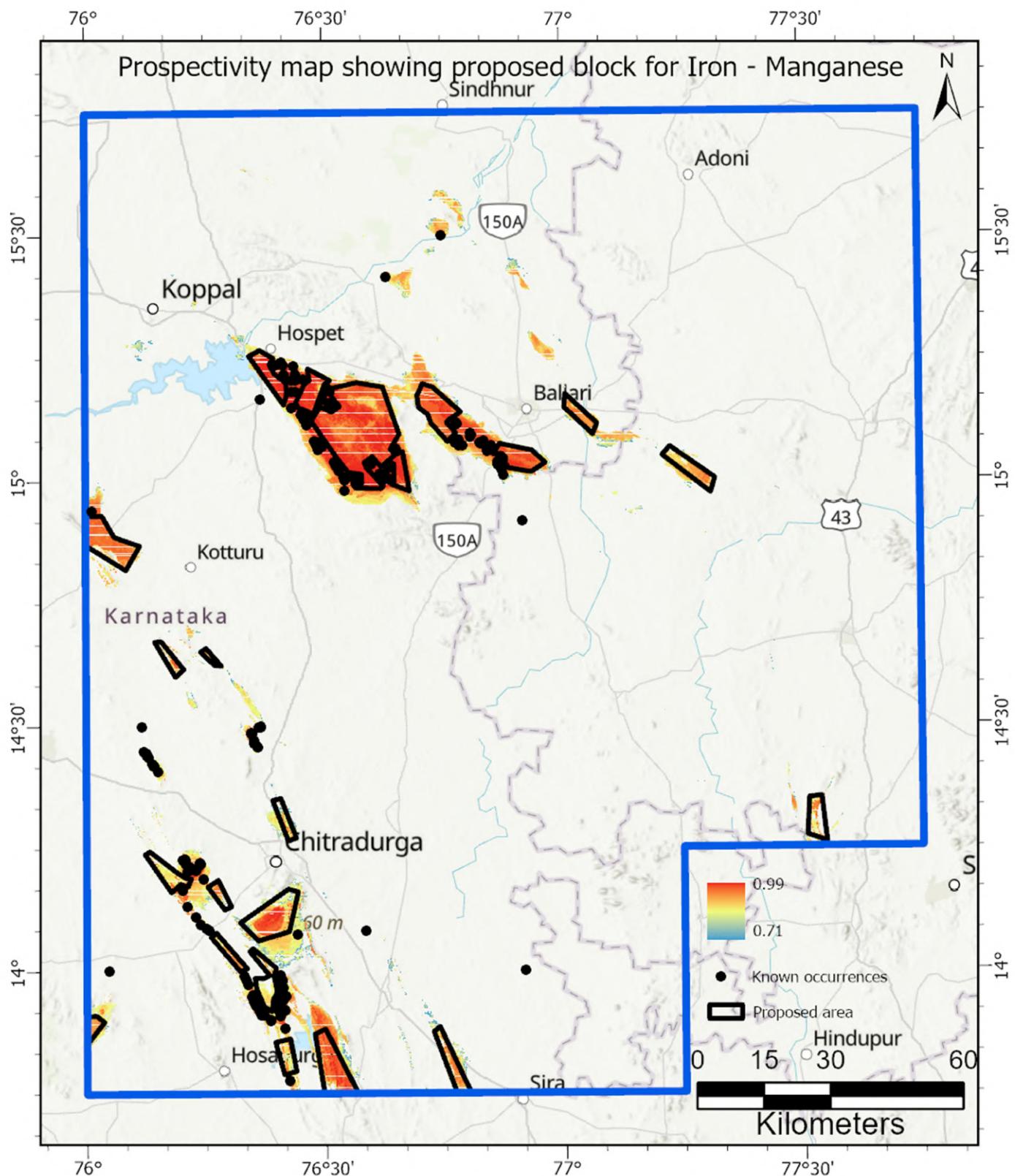
ANNEXURE - I



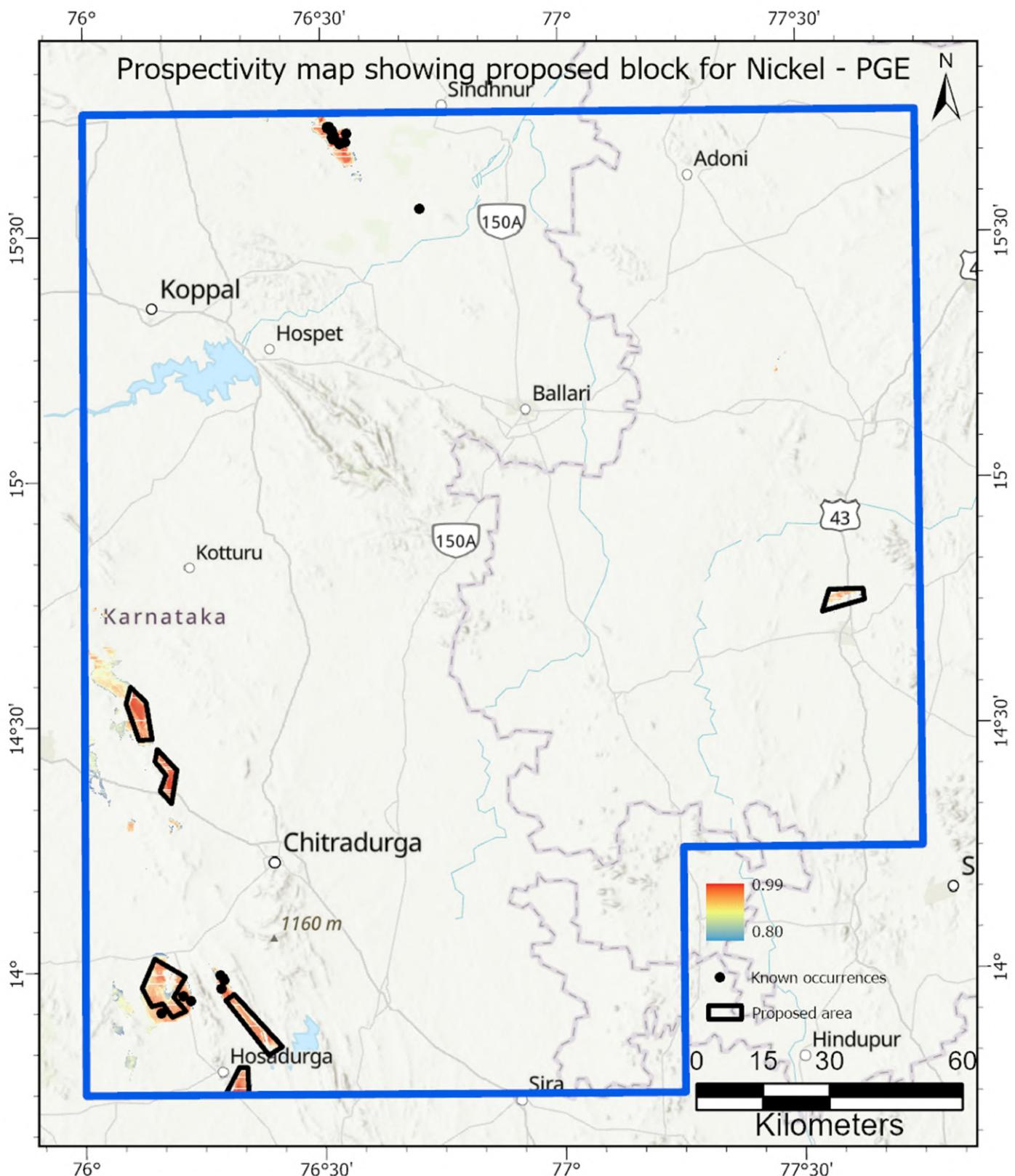
ANNEXURE - I



ANNEXURE - I



ANNEXURE - I



ANNEXURE - I

