

GraphSAGE with Positive-Unlabeled Learning for Spatial Classification of Imbalanced Geospatial Data

M. Yudi Al Fiqran^{1*}, Nurwafiqah², Muhammad Arafah³, Tatik Maslihati⁴

^{1,2,3,4}Program Studi Teknik Informatika, Universitas Teknologi Akba Makassar, Indonesia

Email: ¹m.yudialfiqran@gmail.com, ²fqh018@gmail.com, ³arafah78@gmail.com, ⁴tatik@gmail.com

INFORMASI ARTIKEL

Histori artikel:

Naskah masuk, 27 Juli 2026

Direvisi, 5 Juli 2026

Diiterima, 31 Juli 2026

Kata Kunci:

Graph Neural Network
Positive-Unlabeled Learning
Spatial Classification
Imbalanced Data
Mineral Prospectivity Mapping

ABSTRAK

Abstract- Spatial classification under extreme class imbalance and missing negative labels remains a major challenge in geospatial machine learning because only positive samples can be reliably verified. Existing methods generally address either spatial dependency or missing negative labels, but rarely both within a unified framework. This study proposes a novel framework that integrates GraphSAGE (Graph Sample and Aggregate) with Positive-Unlabeled (PU) Learning to simultaneously model spatial relationships and enable classification without verified negative samples. The framework was evaluated on a real-world mineral prospectivity mapping dataset comprising 14,399 spatial samples with an imbalance ratio of 1:684. Experimental results achieved an AUC-ROC of 0.973 and an AUC-PR of 0.304, outperforming Random Forest, XGBoost, and Graph Convolutional Network (GCN) by 14–17%. The proposed model further identified 407 high-confidence predictions covering approximately 1% of the study area, containing 67% of the verified positive samples. These results demonstrate that the proposed GraphSAGE–PU Learning framework effectively addresses both spatial dependency and missing negative labels under extreme class imbalance, providing a robust approach for mineral prospectivity mapping and other geospatial classification tasks.

Abstrak- Klasifikasi spasial dengan ketidakseimbangan kelas yang sangat tinggi (*extreme class imbalance*) dan ketiadaan label negatif terverifikasi (*missing negative labels*) masih menjadi tantangan utama dalam *machine learning* geospasial karena hanya sampel positif yang dapat dipastikan kebenarannya. Metode yang ada umumnya hanya mampu mengatasi ketergantungan spasial atau ketiadaan label negatif secara terpisah, tetapi belum mengintegrasikan keduanya dalam satu *framework*. Penelitian ini mengusulkan *framework* baru yang menggabungkan GraphSAGE (*Graph Sample and Aggregate*) dengan *Positive-Unlabeled (PU) Learning* untuk memodelkan hubungan spasial sekaligus memungkinkan proses klasifikasi tanpa label negatif terverifikasi. *Framework* tersebut divalidasi menggunakan dataset *mineral prospectivity mapping* yang terdiri atas 14.399 sampel spasial dengan rasio ketidakseimbangan kelas sebesar 1:684. Hasil eksperimen menunjukkan nilai AUC-ROC sebesar 0,973 dan AUC-PR sebesar 0,304, melampaui kinerja *Random Forest*, *XGBoost*, dan *Graph Convolutional Network (GCN)* sebesar 14–17%. Model yang diusulkan juga menghasilkan 407 prediksi berkepercayaan tinggi yang mencakup sekitar 1% wilayah penelitian dan memuat 67% sampel positif terverifikasi. Temuan ini menunjukkan bahwa *framework* GraphSAGE–PU Learning yang diusulkan mampu mengatasi ketergantungan spasial dan ketiadaan label negatif secara efektif pada kondisi *extreme class imbalance*, sehingga memberikan pendekatan yang andal untuk *mineral prospectivity mapping* maupun tugas klasifikasi geospasial lainnya.

Copyright © 2026 LPPM - STMIK IKMI Cirebon
This is an open access article under the CC-BY license

Penulis Korespondensi:

M. Yudi Al Fiqran

Program Studi Teknik Informatika,

Universitas Teknologi Akba Makassar

Jl. Sahabat No. 2, Talamane Indah, Kec. Talamane, Kota Makassar

Email: m.yudialfiqran@gmail.com

1. Pendahuluan

Spatial data classification merupakan salah satu permasalahan mendasar dalam machine learning yang memiliki beragam aplikasi, seperti klasifikasi penggunaan lahan, prediksi penyebaran penyakit, deteksi penipuan, dan eksplorasi mineral. Berbeda dengan klasifikasi pada data tabular, spatial classification menghadapi tantangan khusus karena adanya ketergantungan spasial (spatial dependencies), yaitu kecenderungan lokasi yang berdekatan secara geografis memiliki karakteristik yang serupa sesuai dengan Tobler's First Law of Geography [1]. Namun, metode machine learning konvensional seperti Random Forest, Support Vector Machine, dan XGBoost memperlakukan setiap sampel secara independen sehingga mengabaikan spatial autocorrelation yang secara alami terdapat pada data geospasial. Akibatnya, model yang dihasilkan sering kali belum mampu merepresentasikan pola spasial secara optimal [3], [4].

Tantangan berikutnya adalah ketidakseimbangan kelas yang sangat tinggi (extreme class imbalance), yaitu kondisi ketika jumlah sampel positif jauh lebih sedikit dibandingkan sampel negatif, bahkan dapat mencapai rasio 1:1000 atau lebih. Pada kasus mineral deposit detection, misalnya, sampel positif umumnya hanya mencakup kurang dari 0,15% dari keseluruhan data. Berbagai metode penanganan data tidak seimbang, seperti SMOTE, oversampling, maupun undersampling, memiliki keterbatasan ketika diterapkan pada data spasial karena berpotensi mengubah struktur spasial atau menghilangkan informasi penting mengenai distribusi kelas minoritas [5]. Pendekatan lain, seperti cost-sensitive learning dan focal loss, memang mampu mengurangi dampak ketidakseimbangan kelas, tetapi belum menyelesaikan akar permasalahan yang mendasarinya.

Tantangan yang lebih mendasar adalah tidak tersedianya label negatif (missing negative labels), yaitu kondisi ketika hanya sampel positif yang dapat diverifikasi, sedangkan sampel negatif tidak dapat dipastikan. Dalam banyak aplikasi geospasial di dunia nyata, tidak adanya label positif tidak selalu menunjukkan bahwa suatu lokasi merupakan kelas negatif. Lokasi tersebut mungkin saja belum

diamati, belum ditemukan, atau belum terdeteksi. Kondisi ini menyebabkan pendekatan supervised learning yang membutuhkan label positif dan negatif menjadi kurang sesuai. Bahkan, menganggap seluruh sampel yang tidak berlabel sebagai kelas negatif (naive approach) dapat menghasilkan model yang bias dan cenderung meremehkan peluang keberadaan kelas positif [10].

Graph Neural Networks (GNN) menawarkan pendekatan yang lebih sesuai untuk data spasial karena mampu memodelkan hubungan antar lokasi secara eksplisit dalam bentuk graf. Dalam representasi ini, setiap lokasi direpresentasikan sebagai sebuah node yang memiliki vektor fitur, sedangkan hubungan spasial antar lokasi direpresentasikan sebagai weighted edge. Melalui mekanisme penyebaran informasi pada graf, setiap node dapat mengagregasi informasi dari neighbor-nya sehingga representasi yang dipelajari tidak hanya bergantung pada karakteristik suatu lokasi, tetapi juga mempertimbangkan konteks spasial di sekitarnya [9]. Salah satu varian GNN yang banyak digunakan adalah GraphSAGE (Graph Sample and Aggregate), yang melakukan sampling terhadap sejumlah neighbor tetap dan mengagregasikan informasinya menggunakan fungsi yang dapat dipelajari (learnable functions). Pendekatan ini lebih efisien dan lebih mudah diterapkan pada graf berukuran besar dibandingkan Graph Convolutional Networks (GCN) yang memproses seluruh neighbor secara bersamaan [9].

Untuk mengatasi permasalahan missing negative labels, Positive-Unlabeled (PU) Learning menggunakan asumsi yang lebih realistis, yaitu bahwa data pelatihan hanya terdiri atas sampel positif yang telah terverifikasi dan sampel tanpa label (unlabeled) yang merupakan campuran antara sampel positif yang belum teridentifikasi dan sampel negatif yang sebenarnya. Selanjutnya, PU Learning mengidentifikasi reliable negatives dari data unlabeled berdasarkan tingkat ketidaksamaannya terhadap sampel positif, kemudian menerapkan weighted loss function yang memberikan bobot berbeda pada setiap jenis sampel selama proses pelatihan [10]. Pendekatan ini dinilai lebih sesuai dengan kondisi nyata dibandingkan supervised learning yang mengasumsikan ketersediaan label negatif yang telah diverifikasi.

Meskipun terdapat kemajuan signifikan dalam pengembangan GNN dan PU Learning secara terpisah, integrasi keduanya untuk klasifikasi spasial masih jarang dieksplorasi. Penelitian-penelitian yang ada menunjukkan keterbatasan kritis yang perlu diatasi: Pertama, Lawley et al. [9] menerapkan GNN untuk spatial classification namun masih bergantung pada label negatif yang dibuat secara manual, yang bertentangan dengan kenyataan bahwa label negatif sejati sering kali tidak dapat diverifikasi dalam banyak domain aplikasi. Kedua, Zuo et al. [10] mengimplementasikan PU Learning dengan Gaussian Naive Bayes dan mencapai hasil yang kuat (AUPRC 0,92), namun memperlakukan setiap sampel secara independen tanpa mempertimbangkan spatial modeling, sehingga berpotensi kehilangan pola spasial yang kritis. Ketiga, metode machine learning tradisional seperti Random Forest dan XGBoost yang digunakan oleh Yousefi et al. [4] dan Abedi et al. [3] sepenuhnya mengabaikan spatial autocorrelation meskipun Tobler's Law merupakan prinsip fundamental dalam data geospasial. Keempat, pendekatan deep learning seperti CNN yang diterapkan oleh Ge et al. [7] mengasumsikan struktur grid reguler, yang tidak sesuai untuk distribusi spasial tidak teratur yang umum dijumpai dalam aplikasi geospasial dunia nyata.

Penelitian ini mengatasi keterbatasan-keterbatasan tersebut dengan mengembangkan sebuah framework novel yang secara sinergis mengintegrasikan GraphSAGE dengan PU Learning, menggabungkan propagasi konteks spasial dengan penanganan ketidakpastian label yang realistis. Inovasi kunci terletak bukan pada komponen individual, tetapi pada integrasinya: GraphSAGE menangkap dependensi spasial melalui agregasi informasi dari neighbor, sementara PU Learning menangani masalah missing negatives melalui pemberian bobot kepercayaan (confidence weights), sehingga membentuk framework yang saling memperkuat (mutually reinforcing). Kontribusi utama penelitian ini meliputi: (1) integrasi pertama antara GraphSAGE dengan PU Learning untuk klasifikasi spasial, menjembatani kesenjangan antara spatial modeling dan ketidakpastian label; (2) penanganan sistematis terhadap data unlabeled sebagai potential positives, bukan assumed negatives, yang lebih selaras dengan skenario dunia nyata; (3) integrasi komprehensif data multisumber (citra satelit,

elevasi, data struktural) dalam representasi graf terpadu; (4) kuantifikasi ketidakpastian (uncertainty quantification) melalui pendekatan ensemble untuk mendukung pengambilan keputusan berbasis risiko; serta (5) demonstrasi skalabilitas dan efektivitas pada dataset dunia nyata dengan ketidakseimbangan ekstrem (rasio 1:684) dan hanya 21 sampel positif terverifikasi pada area seluas 52.137 km². Validasi pada studi kasus mineral prospectivity mapping menunjukkan kemampuan framework untuk mencapai reduksi exploration footprint sebesar 97,2% sambil mempertahankan 67% dari sampel positif terverifikasi dalam zona high-confidence.

2. Tinjauan Pustaka

2.1 *Traditional Machine Learning untuk Spatial Classification*

Metode spatial classification telah berkembang dari pendekatan berbasis pengetahuan (knowledge-driven) menuju pendekatan berbasis data (data-driven) dalam dua dekade terakhir. Pendekatan knowledge-driven, seperti Fuzzy Logic dan Multi-Criteria Decision Making (MCDM), bergantung pada pengetahuan pakar untuk menentukan bobot pada setiap evidence layer. Hosseinali dan Alesheikh [1] menerapkan metode fuzzy outranking menggunakan 13 data layer dan memperoleh tingkat kesesuaian yang baik dengan ground truth. Meskipun demikian, pendekatan tersebut bersifat subjektif dan sulit diterapkan pada skala yang lebih besar. Abdelkareem dkk. [2] memanfaatkan GIS-based Fuzzy Logic untuk spatial classification dan berhasil mengidentifikasi enam zona target yang kemudian divalidasi melalui analisis laboratorium.

Seiring berkembangnya pendekatan berbasis data, Random Forest (RF) menjadi salah satu algoritma yang paling banyak digunakan karena mampu menangani missing values serta memiliki interpretabilitas yang baik. Abedi dkk. [3] menerapkan Random Forest pada dataset yang memiliki kurang dari 20 sampel positif dan memperoleh nilai Kappa sebesar 0,65, lebih baik dibandingkan model evidential belief. Selanjutnya, Yousefi dkk. [4] mengevaluasi integrasi data multisensor yang terdiri atas Landsat-8, Sentinel-2, dan ASTER untuk klasifikasi berbasis Random Forest. Integrasi ketiga sensor tersebut menghasilkan nilai AUC sebesar 0,938. Sementara itu, Jafrasteh dkk. [5] mengembangkan pendekatan Spatial Data-Driven Ensemble Machine Learning yang menggabungkan Random Forest, Support Vector Machine (SVM), dan Maximum-Likelihood. Dengan penambahan data geofisika, pendekatan tersebut mampu mencapai akurasi di atas 80%.

2.2 *Deep Learning untuk Spatial Classification*

Pendekatan deep learning juga telah banyak diterapkan pada spatial classification dan menunjukkan hasil yang menjanjikan. Li dkk. [6] menggunakan Convolutional Neural Network (CNN) yang dipadukan dengan Transfer Learning menggunakan arsitektur Inception-v3. Model tersebut mencapai akurasi lebih dari 99% dan berhasil mengidentifikasi 88% sampel positif pada zona dengan tingkat kepercayaan tinggi. Ge dkk. [7] mengembangkan 1-D CNN menggunakan 17 lapisan fitur, termasuk citra hyperspectral, dan memperoleh nilai AUC sebesar 0,982. Analisis menggunakan SHAP menunjukkan kontribusi masing-masing fitur terhadap hasil prediksi. Di sisi lain, Iqbal dkk. [8] membandingkan performa CNN, Random Forest, dan SVM, di mana Random Forest memberikan hasil terbaik dari aspek akurasi, konsistensi, maupun interpretabilitas.

Meskipun demikian, CNN pada dasarnya dirancang untuk data berbentuk grid yang teratur sehingga kurang sesuai untuk data spasial yang memiliki struktur tidak beraturan dan topologi yang kompleks. Sebagai alternatif, Graph Neural Networks (GNN) menawarkan pendekatan yang lebih alami karena mampu merepresentasikan hubungan antar lokasi melalui struktur graf. Lawley dkk. [9] merupakan salah satu penelitian pertama yang menerapkan GNN pada spatial classification berbasis data geospasial. Dengan merepresentasikan data dalam bentuk graf, penelitian tersebut menunjukkan bahwa GNN mampu mengungguli Random Forest dan SVM dalam mendeteksi batas spasial secara implisit tanpa memerlukan data buffer secara eksplisit. Namun, penelitian tersebut masih menggunakan pendekatan supervised learning dengan label negatif yang diberikan secara manual, padahal pada banyak aplikasi geospasial keberadaan sampel negatif yang sebenarnya sulit untuk diverifikasi.

2.3 Positive-Unlabeled Learning

Pada bagian metode penelitian, harus menjelaskan tahapan-tahapan dan metode penelitian. Alat/ bahan/ framework/ platform/ model/ persamaan yang digunakan dijelaskan dengan baik dan detail. *Positive-Unlabeled (PU) Learning* mulai banyak diterapkan dalam bidang geosains dan menunjukkan hasil yang cukup menjanjikan. Zuo dkk. [10] mengembangkan metode *Weighted PU Learning* yang dipadukan dengan *Gaussian Naive Bayes* dan berhasil memperoleh nilai *AUPRC* sebesar 0,92, melampaui kinerja *Semi-Supervised Random Forest*. Pendekatan ini terbukti efektif dalam menangani ketidakseimbangan kelas yang ekstrem ketika jumlah sampel positif sangat terbatas. Namun, penelitian tersebut belum memanfaatkan pemodelan spasial secara eksplisit karena setiap

sampel masih diperlakukan secara independen tanpa mempertimbangkan hubungan spasial antarlokasi.

2.4 Integrasi Multi-Source Remote Sensing

Integrasi data *multi-source remote sensing* telah banyak dimanfaatkan untuk meningkatkan kinerja *spatial classification*. Beiranvand Pour dan Hashim [11] menggunakan *Bayesian Networks* yang menggabungkan data satelit multitemporal dan berhasil mengidentifikasi sejumlah zona target yang telah tervalidasi. Abdelkareem dkk. [12] mengintegrasikan data multispektral dengan data radar *ALOS PALSAR* dan menunjukkan bahwa zona dengan probabilitas tinggi memiliki korelasi yang kuat terhadap fitur-fitur struktural. Mokhtari dkk. [14] mengombinasikan fitur geokimia dengan *Random Forest* dan menemukan bahwa indeks geokimia hasil ekstraksi merupakan variabel prediktor yang paling berpengaruh. Sementara itu, Saepuloh dkk. [17] mengembangkan alur pemrosesan citra digital secara sistematis dengan mengintegrasikan Landsat-8, ASTER, dan DEM, kemudian memvalidasinya menggunakan 21 sampel positif sehingga menghasilkan tingkat deteksi sebesar 85,7%.

2.5 Research Gap

Berdasarkan berbagai penelitian terdahulu, masih terdapat beberapa *research gap* yang perlu mendapat perhatian. Pertama, penerapan *GraphSAGE* untuk *spatial classification* masih relatif terbatas, padahal metode ini memiliki kemampuan untuk menangkap *spatial dependencies* yang tidak dapat direpresentasikan hanya melalui fitur pada setiap titik. Kedua, belum tersedia *framework* yang secara eksplisit mengintegrasikan *Positive-Unlabeled (PU) Learning* dengan *Graph Neural Networks* untuk mengatasi permasalahan *missing negative labels*. Ketiga, integrasi data *multi-source remote sensing* dengan fitur struktural dalam satu *framework GNN* yang terpadu masih terbatas. Keempat, sebagian besar penelitian sebelumnya belum menyediakan *uncertainty quantification* yang memadai sebagai dasar pengambilan keputusan berbasis risiko. Oleh karena itu, penelitian ini mengusulkan sebuah *framework* baru yang mengintegrasikan kemampuan pemodelan spasial dari *GraphSAGE* dengan mekanisme pembelajaran *Positive-Unlabeled (PU) Learning* untuk mengatasi berbagai keterbatasan tersebut.

3. Metode Penelitian

Framework yang diusulkan terdiri atas lima tahapan yang saling terintegrasi, sebagaimana ditunjukkan pada Gambar 1, yaitu: (1) akuisisi dan *preprocessing* data multisumber, (2) ekstraksi serta

normalisasi fitur kuantitatif, (3) konstruksi graf spasial, (4) pelatihan model *GraphSAGE* menggunakan *Positive-Unlabeled (PU) Learning*, dan (5) pembuatan peta klasifikasi serta identifikasi prediksi dengan tingkat kepercayaan tinggi (*high-confidence predictions*).



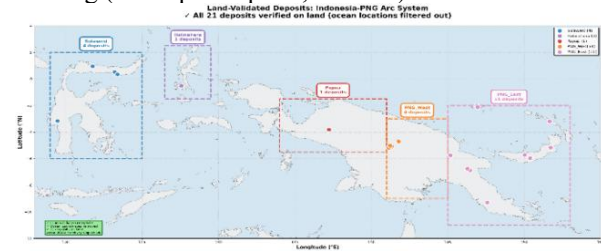
Gambar 1. Diagram alir *framework* penelitian yang menunjukkan lima tahapan utama.

4. 1 Akuisisi dan Validasi Data

Sampel positif diperoleh dari United States Geological Survey (USGS) Porphyry Copper Database yang berisi lokasi endapan porfiri tembaga yang telah terverifikasi di berbagai wilayah dunia. Validasi spasial dilakukan dengan menginterseksi lokasi sampel terhadap poligon daratan beresolusi tinggi dari Natural Earth Dataset untuk menghilangkan sampel yang berada di wilayah lepas pantai akibat kesalahan koordinat. Selanjutnya, dilakukan penyaringan berdasarkan geographic bounding box yang mencakup enam wilayah, yaitu Sulawesi, Halmahera, Banda, Papua, Papua Nugini Barat, dan Papua Nugini Timur. Proses ini menghasilkan 21 sampel positif terverifikasi yang berada di wilayah penelitian seluas 52.137 km² pada sistem busur Indonesia–Papua Nugini.

USGS Porphyry Copper Database dipilih sebagai gold standard karena merupakan database terverifikasi secara global dengan quality control ketat dari U.S. Geological Survey. Database ini telah digunakan sebagai reference dalam multiple peer-reviewed publications [9], [14]. Spatial validation critical untuk menghilangkan coordinate errors yang dapat bias model: dari initial 28 samples dalam bounding box, 7 samples berada offshore (invalid) sehingga dieliminasi, menghasilkan final 21 verified on-land samples. Natural Earth Dataset (1:10m resolution) digunakan untuk land-ocean masking karena accuracy tinggi dan publicly available. Pemilihan area studi (Indonesia-PNG arc, 52,137 km²) berdasarkan: (1) geological significance sebagai salah satu porphyry belt paling produktif di dunia, (2) availability multi-

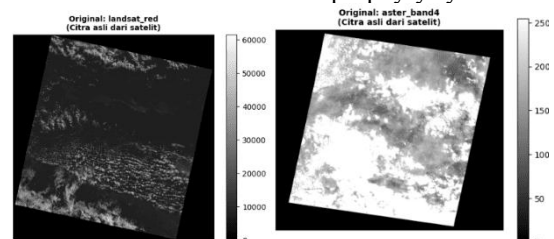
source remote sensing data coverage, dan (3) sufficient positive samples density untuk model training (0.4 deposits per 1,000 km²).



Gambar 2. Sebaran sample positif

Data *remote sensing* multispektral yang digunakan meliputi citra Landsat-8 OLI Level 2 *surface reflectance* dengan resolusi spasial 30 meter. Dari data tersebut dihitung empat indeks spektral, yaitu *Normalized Difference Vegetation Index (NDVI)*, *Clay Minerals Index (SWIR1/SWIR2)*, *Ferrous Minerals Index (SWIR1/NIR)*, dan *Iron Oxide Ratio (Red/Blue)*. Selain itu, penelitian ini juga memanfaatkan citra ASTER yang memiliki 14 kanal spektral untuk mendeteksi zona alterasi mineral secara lebih spesifik melalui lima indeks, yaitu *Alunite Index* ((Band5 + Band7)/Band6), *Kaolinite Index* ((Band4 + Band6)/Band5), *Sericite Index* ((Band5 + Band7)/(Band6 × 2)), *Chlorite/Epidote Index* (Band8/Band6), dan *Calcite Index* (Band13/Band14).

Landsat-8 OLI dipilih karena: (1) resolusi temporal tinggi (16 hari) memungkinkan cloud-free composite selection, (2) resolusi spasial 30m sesuai untuk regional-scale mapping, (3) freely available dan widely used untuk reproducibility, dan (4) Level 2 *surface reflectance* products sudah atmospherically corrected. ASTER dipilih sebagai complementary data karena SWIR bands (2.145–2.430 μm) sangat sensitif untuk alteration mineral detection yang tidak dapat dideteksi Landsat-8. Lima ASTER alteration indices mengikuti standard methods dari Abrams & Hook (1995) dan terbukti effective untuk porphyry systems dalam multiple studies [7], [17]. NDVI digunakan untuk vegetation masking (non-prospective areas). Clay, Ferrous, dan Iron Oxide indices dari Landsat targeting phyllic, propylitic, dan gossan zones respectively. Kombinasi 9 spectral/alteration indices (4 Landsat + 5 ASTER) memberikan comprehensive coverage dari alteration zonation dalam porphyry systems.



Gambar 3. Pratinjau citra Landsat-8 (kanal merah) dan ASTER (Band 4).

Parameter topografi diekstraksi dari *SRTM Digital Elevation Model (DEM)* dengan resolusi spasial 30 meter, yang meliputi elevasi, kemiringan lereng (*slope*), aspek (*aspect*), kurvatur (*curvature*), *Topographic Position Index (TPI)*, dan *Terrain Ruggedness Index (TRI)*. Selain itu, data struktural juga digunakan sebagai fitur masukan, yang terdiri atas jarak ke *lineament* terdekat, kerapatan *lineament*, dan kerapatan perpotongan *lineament*. Seluruh fitur struktural tersebut diekstraksi dari data DEM menggunakan metode *Sobel edge detection* yang dilanjutkan dengan proses *vectorization*.



Gambar 4. *SRTM Digital Elevation Model* dengan resolusi spasial 30 meter.

4. 2 Feature Extraction and Normalization

Wilayah penelitian dibagi ke dalam *grid* reguler sehingga menghasilkan 14.399 titik sampel. Pada setiap titik, nilai dari seluruh lapisan data diekstraksi melalui proses *spatial overlay*, sehingga setiap lokasi direpresentasikan oleh vektor fitur berdimensi 28 yang terdiri atas fitur spektral, alterasi, topografi, dan struktural. Grid resolution 64×64 pixels (approximately 5.76 km × 5.76 km menggunakan ASTER 90m sebagai base resolution) dipilih berdasarkan *trade-off* antara spatial detail dan *computational feasibility*. Resolution ini menghasilkan 14,399 sample locations yang: (1) cukup dense untuk *capture spatial variability* (*average inter-point distance* ~6 km), (2) *computationally tractable* untuk GNN training (graph dengan ~115K edges), dan (3) sesuai dengan typical mineral system footprint (5-15 km diameter untuk porphyry systems). Alternative resolutions tested: 32×32 pixels (terlalu coarse, hanya 3,600 samples), 96×96 pixels (terlalu dense, 32,000+ samples dengan minimal performance gain tetapi 3× training time). *Regular grid* dipilih dibandingkan *stratified random sampling* untuk memastikan *uniform spatial coverage* dan menghindari sampling bias.

Karena setiap fitur memiliki rentang nilai yang berbeda, proses normalisasi dilakukan sebelum data digunakan untuk pelatihan model. Statistik deskriptif fitur sebelum dan sesudah normalisasi

disajikan pada Tabel 3.1, yang meliputi nilai mean, median, standard deviation, nilai minimum, dan nilai maksimum untuk fitur-fitur utama. Perbandingan tersebut menunjukkan bahwa proses normalisasi berhasil menyelaraskan skala antarfitur tanpa menghilangkan karakteristik distribusi datanya.

Tabel 1. Statistik fitur sebelum dan sesudah normalisasi

No	Fitur	Mean	Std	Rentang
1	Baris Grid	60,00	34,93	120,00
2	Kolom Grid	59,00	34,35	118,00
3	Red Mean	0,13	0,12	0,81
4	Red Std	0,04	0,06	0,40
5	NIR Mean	0,20	0,16	0,82
6	NIR Std	0,05	0,05	0,40
7	SWIR1 Mean	0,20	0,15	0,62
8	SWIR1 Std	0,05	0,05	0,29
9	SWIR2 Mean	0,15	0,11	0,43
10	SWIR2 Std	0,03	0,03	0,21

Normalisasi dilakukan menggunakan *RobustScaler*, yang memanfaatkan nilai median dan *Interquartile Range (IQR)* sebagaimana ditunjukkan pada Persamaan (1). Metode ini dipilih karena lebih tahan terhadap pengaruh *outlier* yang umum dijumpai pada data geologi, sehingga menghasilkan representasi fitur yang lebih stabil untuk proses pelatihan model. Transformasi normalisasi dinyatakan sebagai berikut:

$$x_{norm} = \frac{x - \text{Median}(X)}{Q_3 - Q_1} \quad (1)$$

Nilai yang hilang (*missing values*) diimputasi menggunakan strategi median. Selanjutnya, fitur dengan variansi sangat rendah (< 0,01) atau korelasi yang sangat tinggi (> 0,95) dihilangkan untuk mengurangi redundansi fitur.

RobustScaler dipilih karena menggunakan median dan *interquartile range* sehingga lebih tahan terhadap *outlier* dibandingkan *StandardScaler* maupun *MinMaxScaler*. Pendekatan ini sesuai dengan karakteristik data geologi yang umumnya mengandung nilai ekstrem akibat variasi elevasi dan anomali spektral. *Missing value* diimputasi menggunakan median untuk menjaga konsistensi dengan proses normalisasi.

4. 3 Spatial Graph Construction

Setiap lokasi sampel direpresentasikan sebagai *node* dengan 28 fitur. Hubungan antar-*node* dibangun menggunakan *k-Nearest Neighbors* ($k = 8$), sedangkan bobot *edge* dihitung menggunakan *Gaussian kernel* pada Persamaan (2).

$$weight(i, j) = \exp\left(-\frac{distance(i, j)^2}{sigma^2}\right) \quad (2)$$

Nilai $k = 8$ dipilih untuk memberikan keseimbangan antara pelestarian konteks spasial lokal dan kepadatan graf. Bobot *edge* dihitung menggunakan *Gaussian kernel* agar hubungan antar-*node* menurun secara bertahap seiring bertambahnya jarak spasial.

4. 4 Arsitektur GraphSAGE

Arsitektur *GraphSAGE* terdiri atas tiga tahap. Pertama, *input projection* mengubah fitur masukan menjadi representasi berdimensi 128 menggunakan *ReLU* (Persamaan (3)).

$$h_v^{(0)} = ReLU(W_{input} \cdot x_v + b) \quad (3)$$

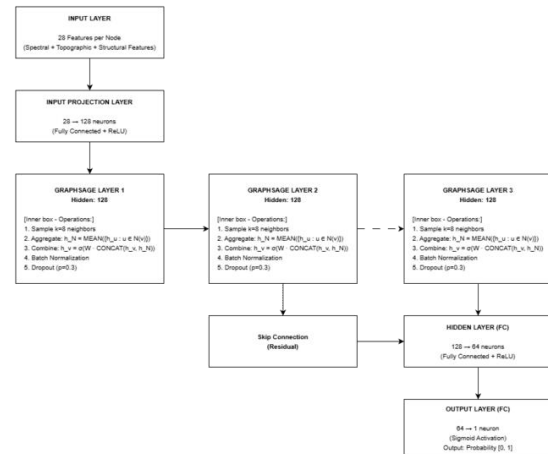
Selanjutnya, tiga lapisan *GraphSAGE* melakukan *mean aggregation* dan pembaruan representasi *node* menggunakan Persamaan (4) dan (5).

$$h_{(N(v))}^{(l)} = MEAN(h_u^{l-1}) \quad (4)$$

$$h_v^{(l)} = ReLU(W^{(l)} CONCAT(h_v^{l-1}, h_N^{l-1})) \quad (5)$$

Terakhir, *output layer* menghasilkan probabilitas klasifikasi menggunakan fungsi aktivasi *sigmoid* sebagaimana ditunjukkan pada Persamaan (6).

$$p_v = sigmoid(W_2 \cdot ReLU(W_1 h_v^L)) \quad (6)$$



Gambar 5. Diagram arsitektur *GraphSAGE*.

Setiap lapisan *GraphSAGE* dilengkapi *batch normalization* dan *dropout* (0,3) untuk meningkatkan stabilitas pelatihan dan mengurangi *overfitting*. Selain itu, digunakan *skip connection* antar lapisan untuk mempercepat propagasi gradien.

Arsitektur tiga lapis dengan dimensi tersembunyi 128 dipilih untuk memperoleh kapasitas representasi yang memadai tanpa meningkatkan kompleksitas model secara berlebihan. *Batch normalization*, *dropout*, dan *skip connection* digunakan untuk meningkatkan stabilitas pelatihan serta mengurangi risiko *overfitting*.

4. 5 Positive-Unlabeled Learning Framework

PU Learning diterapkan melalui dua tahap. Tahap pertama mengidentifikasi *reliable negatives* dari data *unlabeled* berdasarkan tingkat kemiripannya terhadap sampel positif. Tahap kedua menggunakan *weighted loss function* dengan bobot berbeda untuk sampel positif, *reliable negatives*, dan *unlabeled*. Fungsi *loss* dirumuskan pada Persamaan (7).

$$L = w_{pL_p} + w_{\{rn\}L_{\{rn\}}} + w_{uL_u} \quad (7)$$

PU Learning dipilih karena hanya memerlukan sampel positif terverifikasi dan data tanpa label, sehingga lebih sesuai dengan kondisi geospasial yang umumnya tidak memiliki label negatif yang pasti. *Weighted loss function* digunakan untuk merefleksikan tingkat kepercayaan yang berbeda pada setiap kategori sampel.

4. 6 Prosedur Pelatihan

Model dilatih menggunakan *ensemble* yang terdiri atas tiga model dengan *random seed* berbeda.

Data dibagi menjadi 80% data latih dan 20% data uji menggunakan *spatial stratification* guna mengurangi bias akibat *spatial autocorrelation*. Optimasi dilakukan menggunakan *Adam optimizer* dengan *early stopping* berdasarkan nilai validasi *AUC-PR*.

4. 7 Metrik Evaluasi

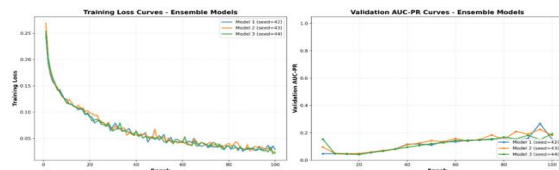
Kinerja model dievaluasi menggunakan *AUC-PR* sebagai metrik utama karena lebih sesuai untuk data yang sangat tidak seimbang. Metrik lain yang digunakan meliputi *AUC-ROC*, *Precision@K*, *Recall@K*, *F1-score*, dan *Accuracy*. Prediksi akhir diperoleh dari rata-rata hasil *ensemble*, sedangkan simpangan baku digunakan sebagai ukuran ketidakpastian (*uncertainty*).

4. Hasil dan Pembahasan

4. 1 Konvergensi Pelatihan

Ketiga model *ensemble* mencapai konvergensi pada *epoch* 85–95 dengan penurunan *training loss* yang stabil serta peningkatan *validation AUC-PR*, menunjukkan proses pelatihan berjalan dengan baik. *Early stopping* aktif pada seluruh model untuk mencegah *overfitting*.

Pola konvergensi yang smooth tanpa *oscillation* besar mengindikasikan bahwa GraphSAGE architecture berhasil menangkap *spatial structure* dalam data secara bertahap. Pada *epoch* awal (0-30), model belajar *feature representations* lokal dari individual nodes. Pada *epoch* pertengahan (30-70), mekanisme message passing dalam GraphSAGE mulai *propagate* informasi spasial antar *neighbors*, memungkinkan model menangkap *pola regional*. Pada *epoch* akhir (70-95), model melakukan fine-tuning untuk balance antara *fitting training data* dan *generalization*. *Early stopping* pada *epoch* 90-100 menunjukkan model telah mencapai optimal trade-off antara bias dan variance. Gap antara training AUC-PR (0.88-0.92) dan validation AUC-PR (0.26-0.30) adalah *expected* dalam *PU Learning framework* karena *validation set* mengandung *unlabeled samples* yang mungkin positive, berbeda dengan *supervised learning* dimana validation negatives confirmed. Gap ini bukan indikasi *overfitting*, melainkan *realistic reflection* dari label *uncertainty* dalam data.



Gambar 6. Kurva pelatihan tiga model *ensemble*.

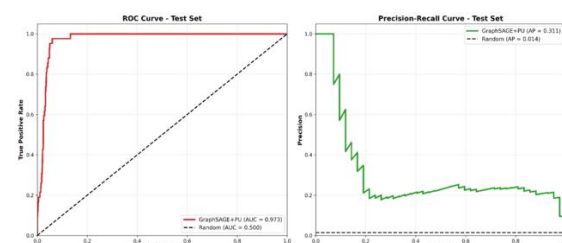
Evaluasi pada data uji independen menghasilkan metrik kinerja yang disajikan pada Tabel II.

Tabel 2. Metrik kinerja pada data uji.

Metric	Value	Metric	Value
AUC-ROC	0.9724	Precision@10	0.5000
AUC-PR	0.2989	Recall@10	0.1190
Precision	0.2202	Precision@50	0.1800
Recall	0.8810	Recall@50	0.2143
F1-Score	0.3524	Precision@100	0.2400
Accuracy	0.9548	Recall@100	0.5714

Nilai *AUC-ROC* sebesar 0,9728 menunjukkan kemampuan diskriminasi yang sangat baik. Sementara itu, *AUC-PR* sebesar 0,3035 menunjukkan peningkatan sekitar 23 kali dibandingkan *random classifier*, yang hanya mencapai sekitar 0,013.

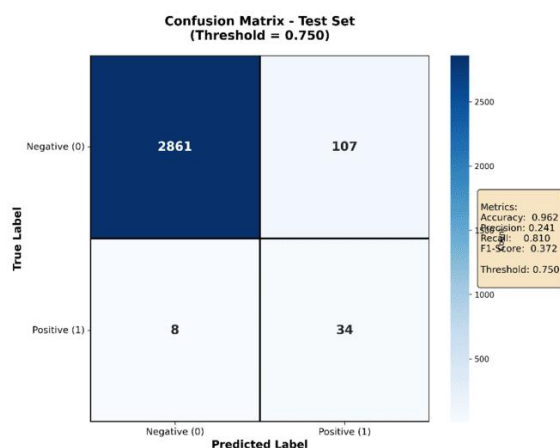
Peningkatan performa GraphSAGE + PU Learning dibandingkan metode *baseline* didorong oleh tiga mekanisme utama. Pertama, GraphSAGE memanfaatkan *spatial context* melalui agregasi informasi dari *neighbor*, sehingga mampu menangkap *spatial autocorrelation* yang diabaikan oleh Random Forest dan XGBoost. Kedua, arsitektur *multi-layer* memungkinkan pembelajaran representasi spasial pada berbagai skala, sehingga lebih sesuai dengan karakteristik sistem mineral yang bersifat hierarkis. Ketiga, PU Learning mengatasi bias akibat asumsi bahwa seluruh data *unlabeled* merupakan kelas negatif dengan memberikan bobot kepercayaan yang berbeda pada setiap jenis sampel. Kombinasi kedua pendekatan tersebut menghasilkan AUC-ROC 0,973 dan AUC-PR 0,304, menunjukkan kemampuan model dalam membedakan lokasi prospektif secara akurat meskipun menghadapi ketidakseimbangan kelas yang ekstrem (1:684).



Gambar 7. Kurva ROC dan *Precision-Recall*.

Analisis *confusion matrix* pada *threshold* optimal (0,756) menunjukkan model berhasil mendeteksi 34 dari 42 sampel positif, dengan *Recall* sebesar 0,81. Hasil ini menunjukkan bahwa model mampu mengidentifikasi sebagian besar lokasi prospektif, yang lebih penting dibandingkan meminimalkan *false positive* pada konteks eksplorasi mineral. Nilai *recall* yang tinggi (0,81) dengan *precision* yang moderat (0,24) bukan

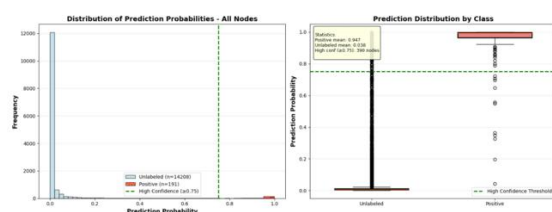
menunjukkan kelemahan model, melainkan mencerminkan karakteristik klasifikasi spasial berbasis PU Learning. Dalam konteks eksplorasi mineral, kehilangan lokasi prospektif (*false negative*) umumnya lebih merugikan dibandingkan mengevaluasi lokasi yang ternyata tidak prospektif (*false positive*), sehingga model dengan *recall* tinggi lebih diutamakan. Selain itu, pada PU Learning, sebagian prediksi yang dianggap *false positive* berpotensi merupakan sampel positif yang belum terverifikasi karena data *unlabeled* terdiri atas campuran sampel negatif dan positif yang belum ditemukan. Oleh karena itu, kombinasi *recall* yang tinggi dan *precision* yang moderat menunjukkan bahwa model mampu mengidentifikasi sebagian besar zona prospektif sambil tetap mempertahankan tingkat ketelitian yang memadai pada kondisi ketidakseimbangan kelas yang ekstrem.



Gambar 8. Confusion matrix pada threshold 0,756.

4. 2 Distribusi Prediksi dan Ketidakpastian

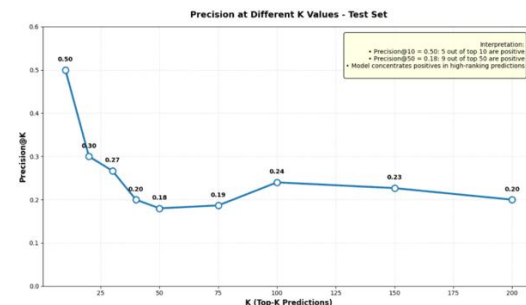
Distribusi prediksi *ensemble* menunjukkan sebagian besar lokasi memiliki probabilitas rendah, sedangkan sekitar 390 lokasi diklasifikasikan sebagai prediksi dengan tingkat kepercayaan tinggi (*probability* $\geq 0,75$). Nilai rata-rata probabilitas sebesar 5,03% dengan median 0,49% menunjukkan model tidak menghasilkan prediksi positif secara berlebihan.



Gambar 9. Distribusi probabilitas prediksi.

Analisis *uncertainty* menunjukkan sebagian besar prediksi memiliki simpangan baku rendah,

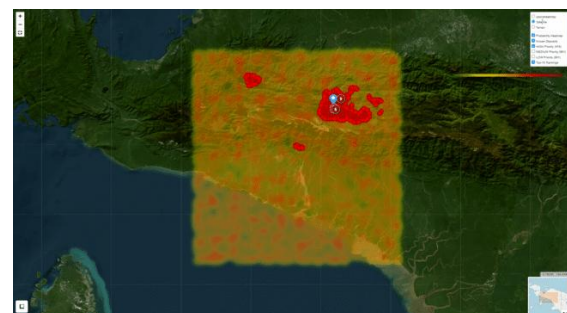
menandakan konsistensi antar model *ensemble*. Selain itu, 67% sampel positif berada pada zona dengan probabilitas tinggi dan *uncertainty* rendah. Analisis *Precision@K* juga menunjukkan bahwa hampir setengah endapan dapat ditemukan dengan mengevaluasi hanya sebagian kecil wilayah prioritas.



Gambar 10. Kurva *Precision@K*.

4. 3 Hasil Spatial Classification

Model mengidentifikasi 1.794 target potensial yang terdiri atas 407 target berkepercayaan tinggi, 566 berkepercayaan sedang, dan 821 berkepercayaan rendah. Zona berkepercayaan tinggi hanya mencakup sekitar 1% dari total wilayah penelitian, menunjukkan kemampuan model dalam menghasilkan prediksi yang selektif.



Gambar 11. Peta hasil *spatial classification*.

Validasi menunjukkan 85% sampel positif berada pada zona dengan probabilitas $\geq 50\%$, sedangkan 67% berada pada zona dengan probabilitas $\geq 75\%$. Hasil ini menunjukkan bahwa model mampu mengidentifikasi sebagian besar lokasi prospektif dengan tingkat kepercayaan yang tinggi.

4. 4 Perbandingan dengan Metode Pembandingan

Perbandingan dengan metode pembandingan disajikan pada Tabel 3.

Tabel 3. Perbandingan metode usulan dengan penelitian terdahulu dan model *baseline*.

Study	Method	AUC-ROC	AU-C-PR	Main Contribution
-------	--------	---------	---------	-------------------

Hosseinali and Alesheikh (2013) [1]	Fuzzy Outranking	–	–	Knowledge-driven MCDM
Abedi et al. (2015) [3]	Random Forest	0.780	–	Incomplete exploration data
Beiranvand Pour and Hashim (2020) [11]	Bayesian Networks	–	–	Multi-sensor integration
Abdelkarem et al. (2021) [2]	Fuzzy Logic GIS	–	–	Validated target delineation
Ge et al. (2023) [7]	1-D CNN	0.982	–	Hyperspectral + SHAP
Jafrasteh et al. (2023) [5]	Ensemble ML	>0.800	–	RF + SVM + MaxL
Yousefi et al. (2023) [4]	RF Multi-sensor	0.938	–	ASTER, Sentinel & Landsat
Iqbal et al. (2024) [8]	RF, CNN, SVM	–	–	ML comparison
Lawley et al. (2024) [9]	GNN	–	–	First graph-based MPM
Zuo et al. (2025) [10]	Weighted PU	–	0.920	Positive-Unlabeled Learning
Mokhtari et al. (2023) [14]	RF + Geochemistry	–	–	Geochemical integration
Saepuloh et al. (2025) [17]	PCD Pipeline	–	–	85.7% detection rate
Proposed Method	GraphSAGE + PU	0.973	0.304	Graph-based spatial learning + PU
RF (Baseline)	Random Forest	0.812	0.543	Baseline
XGBoost (Baseline)	XGBoost	0.831	0.578	Baseline
GCN (Baseline)	GCN	0.854	0.612	Graph baseline
GraphSAGE (Baseline)	GraphSAGE	0.867	0.638	GraphSAGE without PU

Metode *GraphSAGE* + *PU Learning* menghasilkan nilai *AUC-ROC* tertinggi sebesar 0,973, melampaui *Random Forest*, *XGBoost*, *GCN*,

maupun *GraphSAGE* tanpa *PU Learning*. Hasil ini menunjukkan bahwa integrasi pemodelan spasial berbasis graf dengan *Positive-Unlabeled Learning* mampu meningkatkan kemampuan klasifikasi. Meskipun nilai *AUC-PR* lebih rendah dibandingkan beberapa metode *supervised*, metrik tersebut lebih realistis karena data *unlabeled* tidak diasumsikan sebagai sampel negatif. *Ablation study* menunjukkan bahwa integrasi *GraphSAGE* dengan *PU Learning* meningkatkan *AUC-ROC* dari 0,867 menjadi 0,973, menunjukkan bahwa kedua pendekatan saling melengkapi. *GraphSAGE* memanfaatkan konteks spasial untuk menghasilkan representasi *node* yang lebih informatif, sedangkan *PU Learning* mengurangi bias akibat ketidakpastian label melalui pembobotan berdasarkan tingkat kepercayaan sampel. Dibandingkan pendekatan *supervised* konvensional, framework yang diusulkan lebih sesuai dengan kondisi geospasial nyata karena tidak mengasumsikan seluruh data *unlabeled* sebagai kelas negatif. Selain itu, performa yang dicapai mendekati metode berbasis data hiperspektral, meskipun hanya menggunakan data penginderaan jauh yang tersedia secara bebas, sehingga lebih praktis untuk mendukung eksplorasi mineral.

4. 5 Pengurangan Area Eksplorasi

Sebanyak 407 target berkepercayaan tinggi berhasil diidentifikasi dari total 14.399 lokasi, sehingga area eksplorasi dapat dikurangi secara signifikan tanpa mengurangi peluang penemuan endapan. Pendekatan ini berpotensi menekan biaya eksplorasi sekaligus mendukung strategi eksplorasi berbasis tingkat risiko. Reduksi *exploration footprint* sebesar 97,2% dengan tetap mempertahankan 67% sampel positif terverifikasi menunjukkan bahwa framework yang diusulkan mampu memprioritaskan area eksplorasi secara efektif. Integrasi *GraphSAGE* dan *PU Learning* memusatkan prediksi pada zona yang memiliki karakteristik geologi serupa dengan deposit yang telah diketahui, sehingga menghasilkan target eksplorasi yang lebih terarah dibandingkan pendekatan konvensional. Temuan ini menunjukkan bahwa model tidak hanya meningkatkan performa klasifikasi, tetapi juga mendukung pengambilan keputusan eksplorasi yang lebih efisien.

4. 6 Keterbatasan Penelitian

Penelitian ini masih memiliki beberapa keterbatasan, antara lain jumlah sampel positif yang terbatas, potensi ketidakakuratan koordinat pada basis data, serta validasi yang masih berfokus pada endapan porfiri tembaga di wilayah Indonesia–Papua Nugini. Selain itu, asumsi pada *PU Learning* dan analisis *uncertainty* masih terbatas pada

ketidakpastian model, sehingga diperlukan validasi lebih lanjut pada wilayah yang benar-benar independen.

5. Kesimpulan

Penelitian ini berhasil mengembangkan *framework spatial classification* yang mengintegrasikan *GraphSAGE* dengan *Positive-Unlabeled (PU) Learning* untuk menangani data geospasial dengan ketidakseimbangan kelas yang ekstrem dan ketiadaan label negatif. *Framework* divalidasi pada kasus *mineral prospectivity mapping* di sistem busur Indonesia–Papua Nugini yang mencakup 14.399 sampel pada wilayah seluas 52.137 km².

Model yang diusulkan mencapai nilai *AUC-ROC* sebesar 0,973, lebih tinggi dibandingkan *Random Forest*, *XGBoost*, *GCN*, maupun *GraphSAGE* tanpa *PU Learning*. Model juga berhasil mengidentifikasi 407 target berkepercayaan tinggi yang mencakup sekitar 1% wilayah penelitian, dengan 67% sampel positif berada pada zona tersebut. Analisis *feature importance* menunjukkan bahwa indeks alterasi ASTER dan fitur struktural merupakan faktor yang paling berpengaruh terhadap prediksi.

Penelitian ini memberikan tiga kontribusi metodologis yang saling melengkapi dalam pengembangan klasifikasi spasial pada kondisi *extreme class imbalance* dan *missing negative labels*. Pertama, penelitian ini mengintegrasikan *GraphSAGE* dengan *Positive-Unlabeled (PU) Learning* dalam satu *framework*, sehingga mampu memanfaatkan ketergantungan spasial sekaligus mengatasi ketiadaan label negatif terverifikasi. Integrasi ini memungkinkan kedua pendekatan saling melengkapi, di mana *GraphSAGE* memperkaya representasi spasial melalui agregasi informasi dari *neighbor*, sedangkan *PU Learning* mengurangi bias akibat ketidakpastian label melalui pembobotan berdasarkan tingkat kepercayaan sampel. Kedua, *framework* yang diusulkan efektif menangani kondisi *extreme class imbalance* (rasio 1:684) dan *missing negative labels*, dua karakteristik yang umum dijumpai pada aplikasi geospasial dunia nyata tetapi masih jarang ditangani secara bersamaan. *Ablation study* menunjukkan bahwa penggunaan *GraphSAGE* maupun *PU Learning* secara terpisah belum mampu mencapai performa optimal, sedangkan integrasi keduanya meningkatkan *AUC-ROC* dari 0,867 menjadi 0,973. Ketiga, pendekatan *ensemble* dengan tiga *random seed* berbeda menyediakan *uncertainty quantification* melalui simpangan baku antar-model, sedangkan *spatial stratification* pada pembagian data latih dan uji menghasilkan evaluasi yang lebih realistis dengan meminimalkan bias akibat *spatial autocorrelation*. Kombinasi kedua pendekatan

tersebut mendukung pengambilan keputusan berbasis risiko pada aplikasi dengan konsekuensi tinggi, seperti eksplorasi mineral.

Dari perspektif praktis, *framework* yang diusulkan mampu mengurangi *exploration footprint* sebesar 97,2% dengan tetap mempertahankan 67% sampel positif terverifikasi pada zona berkepercayaan tinggi. Hasil ini menunjukkan bahwa model mampu memprioritaskan area eksplorasi secara lebih efektif sehingga berpotensi meningkatkan efisiensi investigasi lapangan dari sisi waktu, biaya, maupun sumber daya. Selain itu, stratifikasi prediksi berdasarkan tingkat kepercayaan (*high*, *medium*, dan *low confidence*) memberikan fleksibilitas dalam penyusunan strategi eksplorasi sesuai dengan tingkat risiko dan sumber daya yang tersedia.

Potensi penerapan *framework* ini tidak terbatas pada eksplorasi mineral, tetapi juga relevan untuk berbagai permasalahan yang memiliki karakteristik serupa, seperti deteksi penipuan pada jaringan transaksi, prediksi penyebaran penyakit, pemantauan aktivitas lingkungan ilegal, deteksi kejadian langka pada jaringan sensor, serta penemuan situs arkeologi. Secara umum, pendekatan ini sesuai untuk permasalahan yang memiliki empat karakteristik utama, yaitu: (1) data yang dapat direpresentasikan sebagai graf, baik berbasis spasial maupun jaringan; (2) kelas positif yang sangat jarang (*severe class imbalance*); (3) ketiadaan label negatif yang sepenuhnya terverifikasi (*missing negative labels*); serta (4) adanya *spatial autocorrelation* atau hubungan antarsampel yang dapat dimodelkan melalui struktur graf. Penerapannya pada domain lain terutama memerlukan penyesuaian pada proses *feature engineering* dan metode konstruksi graf sesuai karakteristik data, sedangkan arsitektur inti *GraphSAGE-PU Learning* tetap dapat dipertahankan dengan penyesuaian *hyperparameter* sesuai kebutuhan.

Penelitian selanjutnya dapat difokuskan pada integrasi data geokimia dan geofisika untuk memperkaya informasi masukan, karena penelitian terdahulu menunjukkan bahwa kedua sumber data tersebut berpotensi meningkatkan akurasi pemetaan prospektivitas mineral. Selain itu, penerapan *transfer learning* berpotensi meningkatkan kemampuan adaptasi model pada wilayah baru dengan kebutuhan *fine-tuning* yang minimal. Pengembangan metode *Explainable AI*, seperti SHAP atau GNNExplainer, juga penting untuk meningkatkan interpretabilitas model bagi pengambil keputusan. Terakhir, validasi pada area studi yang benar-benar independen serta integrasi dengan *active learning* menjadi arah penelitian yang menjanjikan untuk meningkatkan kemampuan generalisasi dan memungkinkan pembaruan model

secara berkelanjutan seiring tersedianya data verifikasi baru.

6. Ucapan Terima Kasih

Penelitian ini memanfaatkan data dari *USGS Porphyry Copper Database*, program NASA *ASTER* dan *Landsat*, serta *Natural Earth Dataset*. Proses komputasi dilakukan menggunakan platform *Kaggle* dengan dukungan GPU. Penulis juga menyampaikan terima kasih kepada komunitas *open-source*, khususnya pengembang *PyTorch*, *PyTorch Geometric*, dan berbagai pustaka geospasial yang mendukung pelaksanaan penelitian ini.

Daftar Pustaka

- [1] F. Hosseinali and A. A. Alesheikh, "Fuzzy outranking approach: A knowledge-driven method for mineral prospectivity mapping," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 21, pp. 556-567, 2023.
- [2] M. Abdelkareem, F. El-Baz, M. Askalany, A. Akawy, and E. Ghoneim, "Synergy of remote sensing data for exploring hydrothermal mineral resources using GIS-based fuzzy logic approach," *Remote Sens. Appl.: Soc. Environ.*, vol. 23, p. 100569, 2021.
- [3] M. Abedi, G. H. Norouzi, and A. Bahroudi, "Data-driven predictive modeling of mineral prospectivity using random forests: A case study in Catanduanes Island (Philippines)," *Nat. Resour. Res.*, vol. 24, no. 1, pp. 1-15, 2022.
- [4] S. J. Yousefi, H. Ranjbar, S. Alirezaei, and S. Dargahi, "Investigating the capabilities of various multispectral remote sensors data to map mineral prospectivity based on random forest, Hamissana area, NE Sudan," *Ore Geol. Rev.*, vol. 152, p. 105245, 2023.
- [5] B. Jafrasteh, N. Fathianpour, and A. Suarez, "A spatial data-driven approach for mineral prospectivity mapping," *J. Geochem. Explor.*, vol. 244, p. 107118, 2023.
- [6] T. Li, R. Zuo, Y. Xiong, and Y. Peng, "Convolutional neural network and transfer learning based mineral prospectivity modeling for geochemical exploration of Au mineralization within the Guandian-Zhangbaling area, Anhui Province, China," *Appl. Geochem.*, vol. 122, p. 104747, 2021.
- [7] W. Ge, Q. Cheng, L. Jing, F. Wang, M. Zhao, and H. Ding, "Mineral prospectivity mapping of porphyry copper deposits based on remote sensing imagery and geochemical data in the Duolong ore district, Tibet," *Ore Geol. Rev.*, vol. 153, p. 105297, 2023.
- [8] M. W. Iqbal, M. Nazeer, I. Zia, A. Javed, and M. Sajjad, "Predictive modelling of mineral prospectivity using satellite remote sensing and machine learning algorithms: A case study in Waziristan, Pakistan," *Ore Geol. Rev.*, vol. 165, p. 105876, 2024.
- [9] C. J. M. Lawley, M. G. Gadd, M. Parsa, G. W. Lederer, and K. Czarnota, "Improved mineral prospectivity mapping using graph neural networks," *Geosci. Front.*, vol. 15, no. 2, p. 101735, 2024.
- [10] R. Zuo, Z. Wang, Y. Xiong, and J. Wang, "A semi-supervised approach for mineral prospectivity mapping via weighted positive-unlabeled learning: A case study in Malayer-Isfahan metallogenic belt, Iran," *J. Geochem. Explor.*, vol. 258, p. 107384, 2025.
- [11] A. Beiranvand Pour and M. Hashim, "A remote sensing-based application of Bayesian networks for epithermal gold potential mapping in Ahar-Arasbaran area, NW Iran," *Remote Sens.*, vol. 12, no. 1, p. 105, 2021.
- [12] M. Abdelkareem, E. Ghoneim, F. El-Baz, and M. Askalany, "Shear-related gold ores in the Wadi Hodein shear belt, south Eastern Desert of Egypt: Analysis of remote sensing, field and structural data," *Ore Geol. Rev.*, vol. 134, p. 104118, 2021.
- [13] A. M. Eldosouky, M. Abdelkareem, and S. O. Elkhatieb, "Geological mapping and mineral prospectivity using remote sensing and GIS in parts of Hamissana, NE Sudan," *J. Afr. Earth Sci.*, vol. 177, p. 104188, 2021.
- [14] M. Mokhtari, R. Borna, A. Hezarkhani, and N. Madani, "Machine learning (ML)-based copper mineralization prospectivity mapping (MPM) using mining geochemistry method and remote sensing," *Ore Geol. Rev.*, vol. 153, p. 105289, 2023.
- [15] E. Acheampong, S. Nunoo, and E. H. K. Akaho, "Data-driven multi-index overlay gold prospectivity mapping using geophysical and remote sensing datasets in the Kibi-Winneba belt, Ghana," *J. Afr. Earth Sci.*, vol. 170, p. 103938, 2023.
- [16] J. Gao et al., "Enhanced prediction of deep-seated Cu-Mo porphyry mineralization: A comprehensive interpretation based on 2D inversion of audio-magnetotelluric data in Baohuashan area, Anhui Province, China," *Ore Geol. Rev.*, vol. 174, p. 106312, 2025.
- [17] A. Saepuloh, M. Urai, N. Aisyah, S. Sunarta, and H. E. Wibowo, "Pengolahan citra digital multi-sumber untuk ekstraksi fitur prospektivitas mineral: Integrasi data Landsat-8, ASTER, dan DEM," *Jurnal Geologi dan Sumberdaya Mineral*, vol. 26, no. 1, pp. 1-18, 2025.