Mapeo Prospectivo Mineral de Pórfidos-Skarn de Cobre en Franja Andahuaylas-Yauri con Redes Convolucionales de Grafos

(Geología y Exploraciones – Aplicaciones de Inteligencia Artificial en Exploración Mineral)

Alejandro Bueno[1], Fernando Saez[2] y Antoine Caté[3]

[1 2] SRK Consulting Perú, Av. La Paz 1227, Lima, Perú (abueno@srk.com.pe / +51952274023).

[3] SRK Consulting Canadá, 155 University Avenue, suite 1500, Toronto, Canadá (acate@srk.com / +1464548228)

RESUMEN

Este estudio evalúa la aplicación de redes convolucionales de grafos (GCN) para el mapeo prospectivo mineral (MPM) de pórfidos-skarn de cobre en la franja Andahuaylas-Yauri, zona metalogenética relevante del sur de Perú. Se integraron datos geocientíficos públicos (geoquímica, litología y contactos litológicos) en mapas de evidencia, los cuales nutrieron cuatro modelos de machine learning: Random Forest, perceptrón multicapa (MLP), red neuronal convolucional (CNN) y GCN. Se empleó validación cruzada espacial con k-means para evitar el sesgo espacial inherente al bajo número de ocurrencias conocidas. Los resultados muestran que tanto MLP como GCN alcanzan los mayores valores de AUC (0.910 y 0.907, respectivamente), superando a RF y CNN. Sin embargo, las predicciones de la GCN presentan mayor coherencia espacial y geológica, identificando zonas favorables que corresponden a los procesos mineralizantes esperados según el modelo de sistema mineral. El uso de redes de grafos permite incorporar de forma efectiva la estructura espacial de los datos, optimizando la delimitación de áreas prioritarias para la exploración de cobre. Este enfoque representa un avance relevante en la integración de inteligencia artificial aplicada a la exploración mineral.

1. Introducción

El mapeo prospectivo mineral (MPM) implica la integración y representación cartográfica de diversos conjuntos de datos geocientíficos, que actúan como proxies espaciales del sistema mineral que se desea modelar, con el objetivo de identificar y delimitar áreas de exploración que presenten una mayor probabilidad relativa de albergar una ocurrencia mineral de interés. Este proceso es especialmente adecuado en regiones que cuentan con un historial suficiente de exploración, lo cual permite definir datos de entrenamiento para cuantificar las asociaciones espaciales entre mineralizaciones conocidas y distintos criterios de orientación en la exploración, representados a través de mapas predictivos como proxies del sistema mineral considerado (Ford, 2019).

Dentro del MPM sustentado en enfoques basados en juicio experto, se emplean técnicas semi-automatizadas que requieren la intervención de profesionales, tales como pesos de evidencia y lógica difusa (Joly et al., 2012). Estos métodos permiten evaluar la relación entre la presencia de mineralización y la distribución de variables predictoras, facilitando la priorización de zonas favorables para la exploración. Asimismo, se han incorporado algoritmos de aprendizaje automático (ML), incluidas redes neuronales y bosques aleatorios, que posibilitan el análisis eficiente de grandes volúmenes de datos, identificando patrones complejos y sutiles asociaciones espaciales entre los distintos factores geológicos que influyen en la formación y localización de depósitos minerales (Sun et al., 2020).

Zuo & Xu (2023) publicaron un caso de estudio demostrando las ventajas de aproximarse a esta tarea con algoritmos basados en grafos, en comparación con los basados en imágenes. El algoritmo rescatado de dicho estudio es la red convolucional de grafos (GCN), ya que sugiere tener un mejor desempeño que la red neuronal convolucional (CNN). Esta investigación es un caso de estudio que compara el desempeño de la propuesta GCN con otros algoritmos de ML ampliamente usados en MPM.

2. Objetivos

Objetivo general:

- Comparar el desempeño del modelo GCN con otros modelos usados comúnmente en MPM.

Objetivos específicos:

- Procesar información geológica de distintas fuentes de libre acceso para generar mapas de evidencia.
- Entrenar un modelo de grafos para generar un mapa de probabilidad de ocurrencias minerales en base a los mapas de evidencia.
- Evaluar el desempeño de distintos modelos de Machine Learning al predecir depósitos de Cu.

3. Metodología

3.1. Área de estudio y datos

3.1.1. Descripción geográfica y geológica

El estudio se delimita a la franja metalogenética del sur de Perú de Andahuaylas Yauri (Perello et al., 2003), el cual está limitado al norte por la deflexión de Abancay y al sur por el altiplano. La franja tiene varios yacimientos de tipo Pórfido y Skarn de Cu emplazados en secuencias sedimentarias cretácicas afectadas por las diferentes deformaciones andinas (Fases Peruana e Inca). El origen de estos yacimientos se atribuye a las series calcoalcalinas Eocenas-Oligocenas del Batolito Andahuaylas Yauri (Perello et al., 2003).

3.1.2. Modelo del sistema mineral

Un sistema mineral se define por todos los factores geológicos que controlan la generación y preservación de la mineralización, y describe todos los procesos que abarcan; desde el origen, transporte y su posterior concentración y preservación (Knox-Robinson & Wyborn, 1997; McCuaig et al., 2010; Wyborn et al., 1994). Los factores geológicos más críticos del sistema mineral han ido cambiando y mejorando a través de los años dependiendo de la escala de estudio (Hagemann et al., 2016; Joly et al., 2012; McCuaig et al., 2010), resumiéndose actualmente en procesos de fuente, flujo, trampa y depósito. Estos, a su vez, se describen por cada tipo de yacimiento según los juicios de expertos. El MPM se llevó a cabo aplicando el modelo de sistema mineral planteado por Hagemann et al. (2016) integrando toda la información geológica disponible (geoquímica, litologías, y otros) dentro de procesos geológicos críticos (Fuente, Flujo y Trampa) por la escala regional del estudio (Figura 1).

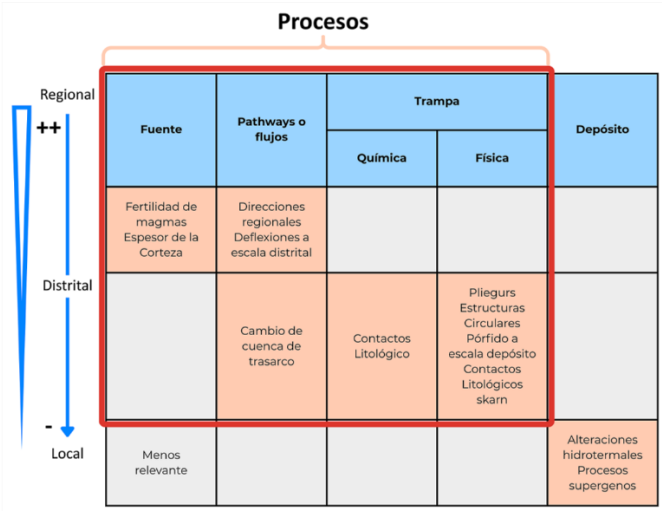

Figura 1. Sistema Mineral de Pórfido de Cobre en términos de procesos críticos (Fuente, flujo, trampa, depósito) a diferentes escalas (regional a local) (modificado de McCuaig et al., 2010).

3.1.3. Mapas de evidencia de proceso de fuente

La información geológica fue descargada desde la aplicación Geocatmin, del Instituto Geológico, Minero y Metalúrgico (INGEMET, 2024) del Perú. Los procesos petrogénéticos que se desarrollan dentro de la evolución de un arco magmático derivan en la combinación de ciertas características fisicoquímicas (fugacidad de O_2, contenido de H_2O, volumen de magma, etc.) que han sido un foco de estudio importante en la formación de yacimientos de tipo pórfido durante los últimos años (Chiaradia, 2021; R. R. Loucks & Fiorentini, 2023; R. Loucks, 2014; Luffi & Ducea, 2022). Dentro del marco de fuente, la litogeoquímica de intrusivos permite interpretar las condiciones de formación y correlacionarlo con los parámetros químicos mencionados. Otro factor importante, y menos tratado, es el espesor cortical, el cual influye directamente en el estado de oxidación del magma y la cantidad de agua disuelta (Chiaradia, 2021). En este estudio se utilizaron ratios litogeoquímicos demostrados que correlacionan de manera general con los proxies magmáticos (fertilidad) y corticales para definir si un magma puede ser favorable en generar depósitos de tipo pórfido de Cu (Figura 2). Se eligió una grilla de 1 km x 1 km para discretizar el mapa y asignar valores a cada centro de píxel, y se incluyeron 35 depósitos de Cu de la zona, generando un mapa de evidencia binario de ocurrencias: 1 donde hay ocurrencia y 0 donde no.

3.1.4. Mapas de evidencia de proceso de flujo

En la franja Andahuaylas-Yauri existe un control estructural particular en la mayoría de los yacimientos de tipo pórfido-skarn de Cu reconocidos. Generalmente se les ha encontrado asociados a pliegues (anticlinales y sinclinales) de rumbos NWSE–EW. Sin embargo, también se destacan corredores de rumbos NESW-NS, que alinean algunos yacimientos (Ej. Los Chancas, Leticia, Cotabambas, Morosayhuas) y que, a escala distrital, representan fuertes zonas de fracturamiento importantes en la permeabilidad para el tránsito de fluidos. Se analizó la información pública de fallas y pliegues dentro de la zona de estudio mediante interpolaciones directas (densidad de estructuras, intersección de estructuras, buffering) y patrones estructurales fractales. Sin embargo, con la información disponible, no se encontró relación entre las estructuras y los depósitos conocidos, por lo que se decidió excluir las estructuras del mapa prospectivo.

3.1.5. Mapas de evidencia de proceso de trampa

En Andahuaylas-Yauri, las zonas de contacto entre intrusivos fértiles y paquetes sedimentarios (calizas, arenas, limolitas, lutitas) representan un

mecanismo importante para concentrar fluidos mineralizantes y desarrollar procesos geoquímicos que permitan la deposición de metales. A pesar de que las aureolas de contacto de los yacimientos en esta franja; a manera general, son bastante irregulares en tamaño y forma (Perello et al., 2003), sirven como trazadores para delimitar zonas favorables en la formación de estos yacimientos. Bajo el contexto mencionado, se utilizaron cinco interpolantes de distancia en intrusivos y rocas sedimentarias con el fin de resaltar estas zonas de contacto, y se generaron mapas de distancia a cada litología y zona de contacto.

3.1.6. Preprocesamiento y preparación

Los mapas de evidencia (o variables) constan del mismo número de píxeles. Estos fueron concatenados en un solo arreglo de datos (matriz), donde cada fila representa un píxel y cada columna, una variable (Figura 2). Las variables de fuente solo fueron generadas completamente para las litologías que tenían muestras geoquímicas. Ello deja 50% de píxeles sin variables de fuente. Para su imputación se revisó en la literatura los valores promedio de cada elemento, por litología; se calcularon los ratios correspondientes (por ejemplo, Sr/Y); y se asignó cada resultado a los píxeles vacíos de cada litología. Luego, cada variable fue normalizada entre 0 y 1, dividiéndola entre su valor máximo.

Criterio	Grupo	Subgrupo	Variable	Relación con ocurrencias
Fuente	Magmas favorables	Fertilidad	Sr/Y La/Yb Nb/Y	Si
		Cortical	V/Sc Ba/Nb Sr/Zr Al2O3/Zr Ancho de corteza*	
		Óxidos	Al2O3 pct Fe2O3 pct MgO pct SiO2 pct TiO2 pct	
		Geocronología	Edad	
Trampa	Contacto	Contacto intrusivo	Distancia a contactos entre rocas e intrusivos favorables en Km	Si
			Distancia a contactos entre rocas e intrusivos no favorables en Km	Si
		Contacto sedimentario	Distancia a roca sedimentaria clástica en Km	Si
			Distancia a roca sedimentaria calcárea en Km	Si
			Distancia a intrusivos máficos en Km	Si

Figura 2. Resumen de variables geológicas definidas en Fuente y Trampa.

3.2. Modelos predictivos

3.2.1. Random Forest (RF)

El modelo RF (Breiman, 2001) es un método de aprendizaje automático basado en el concepto de ensamble de árboles de decisión. Su idea central es combinar múltiples árboles para crear un predictor más fuerte y robusto. Cada árbol en el bosque se construye a partir de un subconjunto aleatorio de los datos originales (método conocido como bagging, abreviatura de bootstrap aggregating), y en cada nodo del árbol solo se considera un subconjunto aleatorio de variables para decidir la división. Este enfoque introduce diversidad entre los árboles y reduce la correlación entre ellos, lo que ayuda a evitar el sobreajuste, uno de los principales problemas de los árboles de decisión individuales. El modelo produce la predicción final combinando la salida de todos los árboles: mediante votación mayoritaria en clasificación o promedio en regresión. Su aplicación en MPM es vasta (Ford, 2019; Hariharan et al., 2017; Zheng et al., 2023), mostrando resultados positivos y sobresalientes con respecto a modelos más sofisticados como redes convolucionales (Sun et al., 2020). A pesar de ser un modelo publicado en 2001, RF y otros modelos basados en árboles, como XGBoost, suelen tener un mejor desempeño predictivo frente a modelos de aprendizaje profundo (Grinsztajn et al., 2022), principalmente, porque trabajan mejor con patrones irregulares, variables con ruido y resultados con límites de decisión no suavizados.

3.2.2. Multilayer Perceptron (MLP)

El modelo MLP es una arquitectura clásica de redes neuronales artificiales ampliamente utilizada en tareas de clasificación y regresión. Consiste en una red de nodos organizados en capas (una capa de entrada, una o más ocultas y una capa de salida), donde cada nodo aplica una transformación no lineal sobre una combinación ponderada de sus entradas. El aprendizaje se realiza mediante el algoritmo de retropropagación (backpropagation) y optimización por descenso del gradiente.

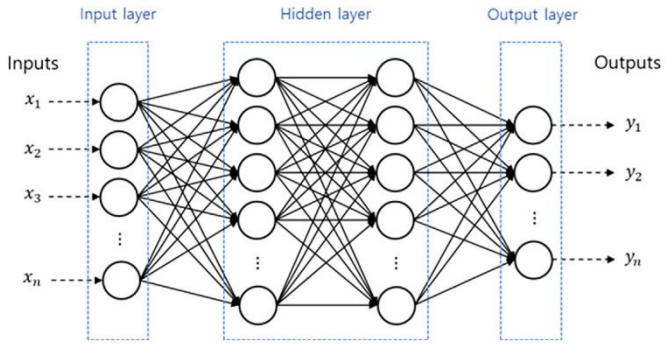

Figura 3. Arquitectura típica de un MLP.

A diferencia de modelos lineales o basados en reglas, el MLP es capaz de modelar relaciones no lineales complejas entre múltiples variables de entrada (Sun et al., 2020). En el contexto del MPM, los MLP han sido utilizados con éxito para identificar patrones sutiles y no lineales entre variables geoquímicas, geofísicas y geológicas, que pueden estar asociados con la presencia de mineralizaciones ocultas. Su capacidad para capturar estas relaciones permite superar las limitaciones de técnicas más tradicionales basadas en umbrales o análisis lineales. Diversos estudios

han demostrado su eficacia en regiones con geología compleja o información incompleta (Mantilla Dulcey, 2023; Sun et al., 2020; Yousefi & Carranza, 2015).

El desempeño de los MLP depende fuertemente de una correcta selección de hiperparámetros, arquitectura con regularización y del preprocesamiento de los datos (normalización, manejo del desbalance, etc.). Suelen tener mejor desempeño con respecto a los modelos basados en árboles cuando las relaciones entre variables son altamente no lineales y suaves, aunque pueden ser superados cuando los datos presentan ruido, valores atípicos o patrones discontinuos (Grinsztajn et al., 2022).

3.2.3. Convolutional Neural Network (CNN)

Las CNN son una clase de redes neuronales profundas diseñadas para capturar patrones espaciales en datos estructurados, como imágenes o mapas. A diferencia de modelos tradicionales, las CNN consideran de manera más compleja las relaciones entre elementos vecinos, permitiendo identificar estructuras, texturas y patrones espaciales relevantes en sus aplicaciones tradicionales (LeCun et al., 2015). El modelo se caracteriza por tener capas convolucionales, que operan mediante la aplicación de filtros (kernels) que recorren la matriz de entrada extrayendo características locales. Posteriormente, capas de reducción de dimensionalidad (pooling) conservan las características más relevantes. Finalmente, una capa de neuronas similar a una de un MLP combina estas características para emitir una predicción sobre la prospectividad mineral del área.

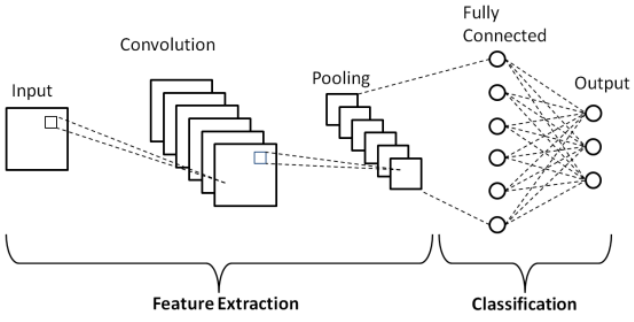

Figura 4. Arquitectura típica de una CNN.

En el mapeo de prospectividad mineral, las CNN han mostrado eficacia al integrar información geológica, alteraciones, firmas geoquímicas y texturas espectrales. En el caso de las CNN de 1 dimensión, las convoluciones se realizan entre las variables, a diferencia de las CNN 2D donde las operaciones se efectúan entre los registros o píxeles de la matriz. Por ejemplo, una CNN 1D aplicada a datos geoquímicos superó a modelos clásicos como Random Forest en la detección de zonas Pb-Zn en Changba, China, siendo eficiente sin requerir mapas espaciales (B. Li et al., 2023). Por su parte, el uso de CNN 2D con patches espectrales de imágenes satelitales (Farahbakhsh et al., 2025) y CNN 3D para identificar zonas alteradas y cuerpos minerales en profundidad también ha mostrado resultados prometedores (X. Li et al., 2023)

3.2.4. Graph Convolutional Network (GCN)

Las GCN (Kipf & Welling, 2017) son una extensión de las redes neuronales que trabajan específicamente sobre grafos. Esto las convierte en modelos muy atractivos para datos que presentan relaciones espaciales, topológicas o de conectividad, como en la geología o en la distribución de muestras minerales. En una GCN, cada punto de muestreo se modela como un nodo, y las conexiones entre ellos son llamadas aristas y representan relaciones como cercanía geográfica, pertenencia a una misma unidad litológica o correlación estructural.Este modelo aprende representaciones combinando la información de un nodo con la de sus vecinos a través de operaciones similares a convoluciones. Esta propagación se repite por capas, permitiendo que los nodos integren progresivamente información de su entorno geológico más amplio. Existen etapas clave al aplicar este modelo, como construir la matriz de adyacencia para indicar qué nodos están conectados entre sí.

Las GCN son especialmente adecuadas para MPM, ya que modelan no solo las propiedades locales, sino también el contexto espacial mediante estructuras de grafos, algo clave en la predicción de yacimientos minerales (Zuo & Xu, 2023). En la actualidad, el uso de este tipo de red ha demostrado ser efectivo y se puede ver en la aplicación de una GCN con atención (Atten-GCN) que alcanzó una precisión del 96.59% en la identificación de zonas prospectivas de oro en Yingfengjie, China (Rui et al., 2025).

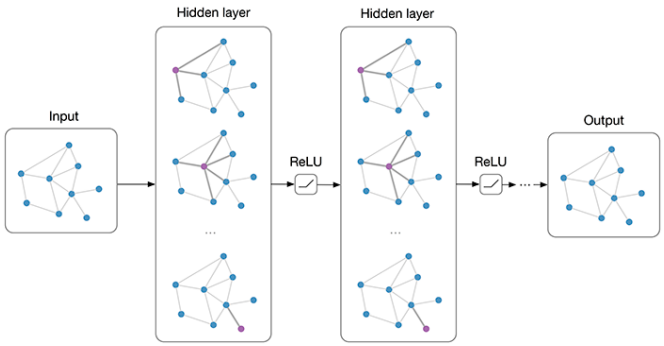

Figura 5. Arquitectura típica de una GCN.

3.3. Configuración de modelos

3.3.1. Validación cruzada geoespacial

Una de las principales dificultades en problemas referentes a mapeo prospectivo mineral (MPM) es la distribución altamente desbalanceada y

espacialmente sesgada de los datos de ocurrencias minerales. En este estudio, se parte de un conjunto de datos conformado por 35 puntos con ocurrencias de mineral conocidas de depósitos tipo pórfido-skarn de Cu y ~20,000 puntos sin conocimiento sobre su mineralización. Esta distribución representa un problema clásico de desbalance de clases y también introduce riesgos de fugas de información espacial si se aplica una validación cruzada aleatoria convencional (Roberts et al., 2017).

Debido a esto, se adoptó una estrategia de validación cruzada basada en clustering espacial, que consiste en aplicar el algoritmo k-means con k=5 sobre las coordenadas geográficas (X e Y) de las ocurrencias minerales (Figura 6). Este enfoque permite dividir el espacio geográfico en 5 regiones mutuamente excluyentes que luego son utilizadas como folds para validación cruzada (Valavi et al., 2019). Se utiliza k-means debido a que permite preservar la autocorrelación espacial inherente a los datos geológicos, mediante la agrupación de ocurrencias geográficamente cercanas. Asimismo, al aplicar k-means únicamente sobre las ocurrencias se garantiza que cada fold tenga un subconjunto representativo de los pocos puntos positivos. Lo que permite una partición balanceada del conjunto total. Este tipo de validación cruzada espacial, conocida como “spatial blocking”, ha sido recomendada en estudios de modelamiento ambiental y geológico debido a que permite la generalización del modelo de manera realista a nuevas regiones no antes vistas (Roberts et al., 2017; Valavi et al., 2019).

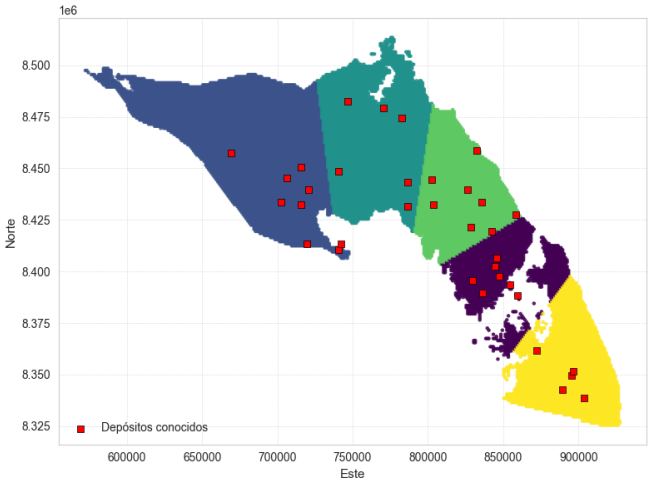

Figura 6. Mapa de 5 folds para validación cruzada

3.3.2. Funciones de activación

Las funciones de activación son componentes esenciales en redes neuronales, ya que introducen la no linealidad que permite a los modelos aprender representaciones complejas de los datos. En este trabajo, se han empleado principalmente 2 funciones de activación: Rectified Linear Unit (ReLU) en las capas ocultas y Softmax en las capas de salida.

1. **ReLU:** Se utilizó en todas las capas ocultas. Esta función es ampliamente adoptada gracias a su simplicidad computacional y por mitigar el problema de desvanecimiento de gradiente que afecta a funciones como sigmoid o tanh (Glorot et al., 2011). Además, ReLU promueve una activación más dispersa (sparse), lo que mejora la eficiencia del modelo y su capacidad de generalización (Nair & Hinton, 2010). Aunque, puede provocar neuronas “muertas” que dejan de aprender, propias de la naturaleza de la función. Se define matemáticamente como:

$$ReLU(x) = \max(x, 0) = \begin{cases} x \; if \; x > 0 \\ 0 \; if \; x \leq 0 \end{cases}$$

Esto significa que cualquier entrada negativa se convierte en 0, mientras que las positivas se mantienen sin cambiar.

2. **Softmax:** Para la capa de salida de los diferentes modelos se ha utilizado la función Softmax. Esta transforma un vector de valores arbitrarios (logits) en una distribución de probabilidad sobre las K clases. Se define como:

$$softmax(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

A pesar de que el problema es binario (ocurrencia de mineral o no), Softmax se empleó en lugar de la tradicional Sigmoid porque permite tener estabilidad numérica en ciertas implementaciones. En algunos frameworks, la función Softmax ofrece una implementación más estable de la mano de funciones de pérdida como la entropía cruzada (Goodfellow et al., 2016). Asimismo, esta función considera la relación entre clases. Puede generar decisiones más diferenciadas cuando las clases están altamente desbalanceadas, como es el caso. Esto es útil cuando se interpreta la salida como una distribución de probabilidad sobre hipótesis espaciales (Murphy, 2012).

3.3.3. Arquitectura de modelos y selección de hiperparámetros

En esta sección se describen las configuraciones generales utilizadas para los modelos predictivos evaluados en este estudio: RF, MLP, CNN y GCN. Todos los modelos fueron diseñados con el objetivo de estimar la probabilidad de ocurrencia de depósitos minerales tipo pórfido-skarn, a partir de información geoquímica, geofísica y espacial.Cada modelo fue entrenado para maximizar el valor promedio del Área Bajo la Curva ROC (AUC) obtenida luego de haber evaluado el modelo en el conjunto de prueba. La configuración específica de hiperparámetros seleccionados para cada modelo se resume en la Tabla 1.

Tabla 1. Resumen de hiperparámetros seleccionados para cada modelo

Modelo	Hiperparámetros seleccionados
Random Forest	Compuesta por 50 árboles y el criterio de entropía para medir la calidad de las divisiones. Se estableció una profundidad máxima de 10 niveles, una selección aleatoria de atributos usando log2, un mínimo de 5 muestras para dividir un nodo y al menos 2 muestras por hoja. El muestreo se realizó con reemplazo (bootstrap = True).
MLP	Compuesta con 4 capas ocultas con 64, 64, 16 y 32 neuronas respectivamente. Se utilizó una tasa de aprendizaje de 0.001 y un término de regularización L2 de 0.0032. El tamaño del lote fue de 64. Se aplicó una capa de dropout con una tasa de 29.3% antes de la capa final, como mecanismo para mitigar el sobreajuste.
CNN	Compuesta por 2 capas convolucionales con 31 y 25 filtros respectivamente, utilizando un tamaño de kernel de 2×2. Cada capa convolucional fue seguida por una activación ReLU, y dropout con una tasa de 15.7%. El tamaño del lote fue de 128 y el aprendizaje se realizó con una tasa de 0.0001.
GCN	Se implementó con 2 capas convolucionales de grafos con 123 y 38 canales ocultos respectivamente. Se aplicó dropout con una tasa de 35.1% después de cada capa convolucional. El entrenamiento utilizó una tasa de aprendizaje de 0.009 y un peso de regularización L2 de 0.006.

*Los hiperparámetros fueron obtenidos mediante optimización bayesiana utilizando la librería Optuna en Python.
** Los hiperparámetros fueron seleccionados mediante búsqueda en malla utilizando GridSearchCV en Python.

Para todos los modelos basados en redes neuronales (MLP, CNN y GCN), se utilizó la función de pérdida Cross-Entropy, ampliamente empleada en tareas de clasificación, especialmente cuando se interpretan las salidas como distribuciones de probabilidad (Goodfellow et al., 2016).

En clasificación multiclase con K clases y una única etiqueta correcta, el valor de pérdida se define como:

$$L_{CE}(p,q) = -\sum_{i=1}^{k} p_i \log(q_i) = -\log(q_y)$$

Donde p es la distribución verdadera (etiquetas one-hot), q es la predicha (softmax(x)), y es la clase correcta y qy su probabilidad predicha.

Debido al gran desbalance de clases, se optó por darle pesos a cada una durante el entrenamiento dentro de la función de pérdida.

Estos pesos fueron obtenidos con la siguiente fórmula:

$$w_i = \left[1, \frac{n_i}{m_i}\right]$$

Donde ni es la cantidad de puntos sin ocurrencia de mineral y mi es la cantidad de ocurrencias en el fold i.

Por último, como optimizador se empleó Adam (Adaptive Moment Estimation), conocido por su eficiencia computacional, bajo requerimiento de memoria y buen desempeño en problemas con gradientes escasos o ruidosos (Kingma & Ba, 2017)

3.3.4. Métricas de evaluación

1. AUC: La curva ROC representa de manera bidimensional el rendimiento de un clasificador. Para poder comparar diferentes clasificadores, se reduce el ROC a un único valor escalar calculando el área bajo la curva ROC, abreviada como AUC (Bradley, 1997; Hanley & McNeil, 1982), que representa el rendimiento esperado del clasificador.

El AUC es una parte del área del cuadrado unitario, por lo que su valor siempre estará entre 0 y 1. La estimación aleatoria produce una línea diagonal entre (0,0) y (1,1) y tiene un área de 0.5. Un clasificador siempre debería ser mejor que el azar, es decir, su AUC debería ser mayor a 0.5.

Esta métrica tiene una propiedad estadística muy relevante: el AUC de un modelo de clasificación es igual a la probabilidad de que este asigne un puntaje mayor a una muestra positiva que a una negativa (Fawcett, 2006), esto se define como:

$$AUC = P(score(X_1) > score(X_0))$$

Es útil en MPM porque evalúa la capacidad del modelo para distinguir entre zonas prospectivas y no prospectivas, sin depender de un umbral de decisión fijo.Todo esto se logra gracias a que se utiliza cada región (fold) previamente obtenida con el algoritmo k-means como conjunto de prueba y las 4 regiones restantes como conjunto de entrenamiento. Finalmente, se reporta el valor promedio del AUC, métrica robusta para el desbalance de clases (Bradley, 1997), para medir el rendimiento de cada modelo.

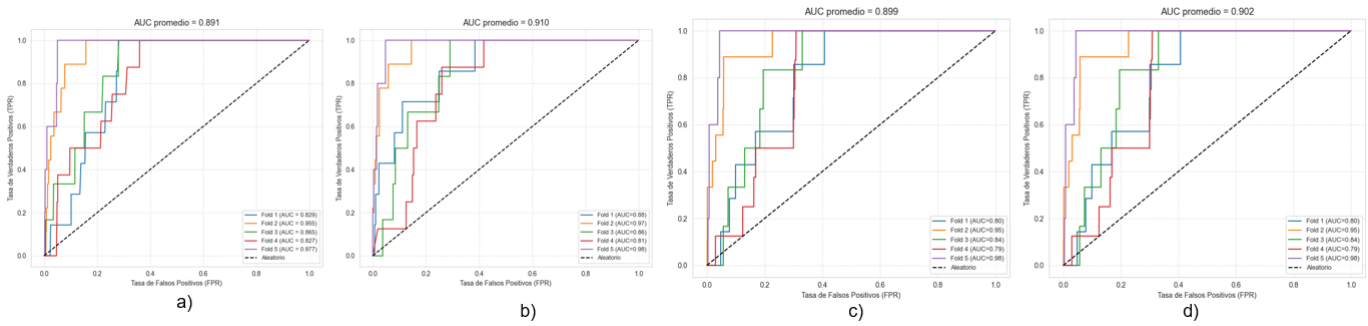

Figura 7. AUC promedio de cada modelo obtenido de la validación cruzada con 5 folds: (a) RF, (b) MLP, (c) CNN y (d) GCN.

4. Resultados y discusión

La Tabla 2 presenta el rendimiento promedio de los modelos evaluados (Random Forest, MLP, CNN y GCN) en la validación cruzada basada en clusters espaciales (5 folds). La métrica principal de comparación fue el Área Bajo la Curva ROC (AUC).

Tabla 2. Resultados promedio de los modelos (5 regiones- Validación cruzada)

Modelo	AUC promedio
Balanced Random Forest	0.89
MLP	0.912
CNN	0.901
GCN	0.907

Los modelos que obtuvieron un mejor desempeño en la tarea de mapeo prospectivo fueron el MLP y la GCN, alcanzando valores promedio de AUC de 0.910 y 0.907, respectivamente. Este resultado sugiere que ambos modelos fueron altamente efectivos para discriminar zonas con alta y baja probabilidad de ocurrencia mineral, incluso en un contexto de datos muy desbalanceados. El MLP logró capturar relaciones no lineales complejas entre variables geoquímicas y geofísicas, mientras que la GCN demostró ser útil al incorporar la estructura espacial del entorno a través de grafos. Por su parte, los modelos Random Forest y CNN también mostraron un rendimiento competitivo, con AUC de 0.893 y 0.899 respectivamente, aunque ligeramente inferior.

La Figura 7 muestra la comparación entre las curvas ROC de cada fold, para cada modelo. En ella, se aprecia que todos los modelos tienen dificultad distinguiendo entre clases en los folds 1, 3 y 4, y muy buena distinción en los 2 y 5. Esto, presuntamente, se debe a la distribución espacial de los datos en tales regiones o a la calidad o naturaleza de los datos. Teniendo zonas que no brindan mucha información. Sin embargo, ello no es un inconveniente para asegurar que los 4 modelos tienen un buen desempeño separando ocurrencias de no ocurrencias, ya que sus curvas son similares para cada fold. MLP y GCN muestran los mejores puntajes de AUC ponderado: 0.91 y 0.907, respectivamente.

4.1. Distribución de probabilidades

La Figura 8 muestra la distribución de probabilidad de ocurrencia predicha de los píxeles con ocurrencia conocida (dorado) y de los que no tienen ocurrencia conocida (gris), de cada uno de los modelos. En general, las predicciones para las “no ocurrencias” en todos los modelos están fuertemente agrupadas cerca de 0. Lo que indica que los modelos tienen alta confianza de que esos puntos son negativos. Por otro lado, las predicciones para ocurrencias están desplazadas hacia la derecha, especialmente en el modelo Random Forest (a) y GCN (d). Los modelos son capaces de asignar altas probabilidades a la clase positiva.

Los modelos MLP (b) y CNN (c), aunque tienen un buen rendimiento, muestran mayor solapamiento entre clases, no son capaces de discriminar en algunos puntos del conjunto de datos. Tanto RF como GCN muestran distribuciones más separadas, y una buena porción de píxeles sin ocurrencia conocida con alta probabilidad predicha, que pueden ser zonas de interés para exploración.

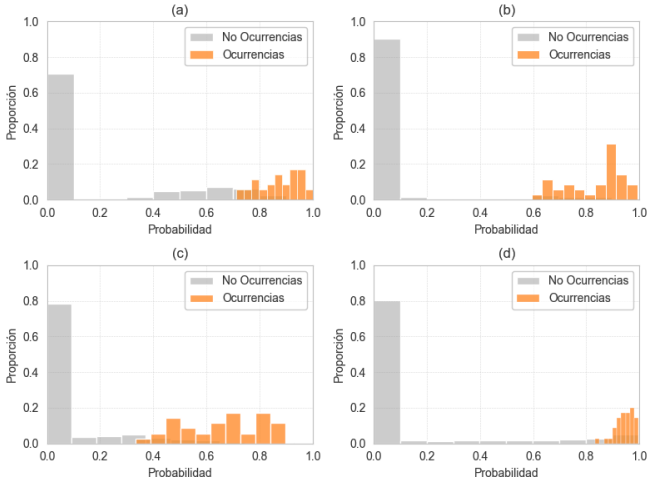

Figura 8. Distribución de probabilidad por existencia de ocurrencias y modelo: RF (a), MLP (b), CNN (c) y GCN (d)

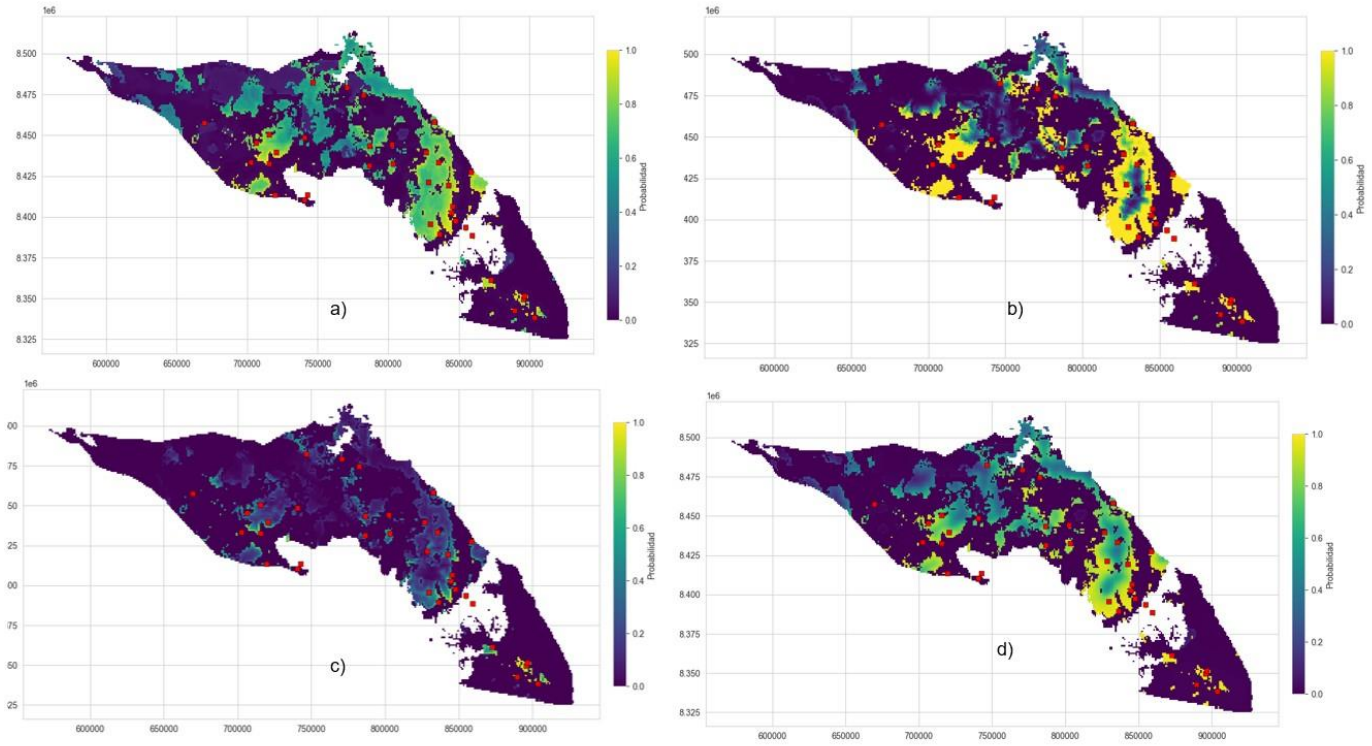

Figura 9. Mapas de prospectividad generados por cada modelo: (a) RF, (b) MLP, (c) CNN, (d) GCN.

4.2. Mapas de prospectividad generados

Se generaron mapas de prospectividad para cada modelo, a partir de sus predicciones de probabilidad sobre el conjunto total de datos. En la Figura 9 se muestran los mapas obtenidos para RF, MLP, CNN y GCN, respectivamente. Los puntos en rojo representan las ocurrencias de Cu conocidas, mientras que los demás puntos en cada mapa representan una probabilidad generada por uno de los modelos.

El modelo RF se ajusta más a toda la roca intrusiva y le asigna una probabilidad promedio (60%) en zonas alejadas de las ocurrencias, mientras que los 3 modelos de redes tienen una variación gradual clara desde los bordes del intrusivo hacia el centro de este. Esto último resulta coherente geológicamente, ya que, según el modelo de sistema mineral, las ocurrencias deberían venir del intrusivo, seguir estructuras y depositarse en zonas cercanas a los contactos con otras rocas. La falta de variación gradual hacia el centro del intrusivo en RF evidencia que el modelo no identifica correctamente este fenómeno.

Todos los resultados apuntan a que los cuatro modelos son capaces de separar las ocurrencias de no ocurrencias de manera eficiente. No obstante, GCN muestra un rendimiento mayor, ya que tiene un AUC promedio alto, cubre los píxeles de ocurrencia conocidos con probabilidades altas y propone nuevas zonas con contexto geológico más coherente que los otros modelos. Por su parte, CNN logra zonas más regulares y definidas espacialmente, lo cual puede ser útil en etapas iniciales de exploración.

Con el modelo GCN entrenado con la data completa, se generó un mapa de probabilidad y se discretizó en 3 niveles de prospectividad: alto, medio y bajo. Los umbrales entre cada categoría se definieron observando la distribución acumulada de predicción del modelo, e identificando los cambios más abruptos. Ello resulta en el mapa prospectivo final. Finalmente, se cruzó el mapa con las zonas no denunciadas (FUENTE), resultando en X km² de área altamente prospectiva sin denuncios mineros.

5. Conclusiones y trabajos futuros

En base a lo expuesto en esta investigación, se concluye lo siguiente:

- El modelo GCN tiene un mejor desempeño general en el MPM de Cu al compararlo con los modelos usados comúnmente, convergiendo con lo sugerido por Zuo and Xu (2023). CNN se ajusta más a zonas similares a las ocurrencias y es útil para exploración temprana, MLP tiende a sobreajustar y no dar probabilidades altas en píxeles de ocurrencia conocida, y RF sigue siendo una buena alternativa para evaluaciones rápidas, a pesar de tener casi 24 años al momento de escribir esta investigación. Esto último se alinea pertinentemente con lo expuesto por Grinsztajn et al. (2022).

- El procesamiento de la data geológica sigue siendo el proceso más importante en el MPM. En este trabajo, el 80% del tiempo se invirtió en la correcta interpretación de patrones geológicos, estimación de leyes de elementos y compuestos, imputación de data y compilación de mapas de evidencia; y que los modelos tengan resultados similares en validación cruzada es un indicador de que la data es consistente. Sin embargo, es conveniente profundizar en la generación de variables del proceso de flujo (fallas y pliegues), ya que la información pública disponible no tuvo correlación con la variable objetivo del estudio.
- Las zonas de contacto entre intrusivos y otras litologías, excluyendo la zona del nor-oeste del mapa, tienen el mejor potencial prospectivo. Se considera que el modelo distingue correctamente entre contactos con intrusivo prospectables y no prospectables.
- La aplicación de algoritmos de Inteligencia Artificial en MPM es un campo de estudio muy entusiasmante, ya que hay herramientas unificadoras de información y predictoras de libre acceso, como las utilizadas en esta investigación. Al disponer y combinar data pública y privada (como estudios locales cercanos a depósitos minerales, muestras de minas, reservas naturales, etc.), los resultados pueden ser prometedores a una escala local.

6. Referencias Bibliográficas

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms [Publisher: Elsevier BV]. *Pattern Recognition, 30*(7), 1145-1159. https://doi.org/10.1016/s0031-3203(96)00142-2

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324

Chiaradia, M. (2021). Magmatic Controls on Metal Endowments of Porphyry Cu-Au Deposits. In *Tectonomagmatic Influences on Metallogeny and Hydrothermal Ore Deposits: A Tribute to Jeremy P. Richards (Volume I)* (pp. 1–16). Society of Economic Geologists. https://doi.org/10.5382/sp.24.01

Farahbakhsh, E., Goel, D., Pimparkar, D., Muller, R. D., & Chandra, R. (2025). Convolutional neural networks for mineral prospecting through alteration mapping with remote sensing data [Version Number: 1]. https://doi.org/10.48550/ARXIV.2502.18533

Fawcett, T. (2006). An introduction to ROC analysis [Publisher: Elsevier BV]. *Pattern Recognition Letters, 27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Ford, A. (2019). Practical Implementation of Random Forest-Based Mineral Potential Mapping for Porphyry Cu–Au Mineralization in the Eastern Lachlan Orogen, NSW, Australia. *Natural Resources Research, 29*, 267-283. https://doi.org/10.1007/s11053-019-09598-y

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, 315-323.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning* [http://www.deeplearningbook.org]. MIT Press.

Grinsztajn, L., Oyallon, E., & Varoquax, G. (2022). Why do tree-based models still outperform deep learning on tabular data? *Neural Information Processing Systems Datasets and Benchmarks Track*.

Hagemann, S., Lisitsin, V., & Huston, D. (2016). Mineral system analysis: Quo vadis [Publisher: Elsevier BV]. *Ore Geology Reviews, 76*, 504–522. https://doi.org/10.1016/j.oregeorev.2015.12.012

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. [Publisher: Radiological Society of North America (RSNA)]. *Radiology, 143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Hariharan, S., Tirodkar, S., Porwal, A., Bhattacharya, A., & Joly, A. (2017). Random Forest-Based Prospectivity Modelling of Greenfield Terrains Using Sparse Deposit Data: An Example from the Tanami Region, Western Australia. *Natural Resources Research, 26*, 489–507. https://doi.org/10.1007/s11053-017-9335-6

Joly, A., Porwal, A., & McCuaig, T. C. (2012). Exploration targeting for orogenic gold deposits in the Granites-Tanami Orogen: Mineral system analysis, targeting model and prospectivity analysis [Publisher: Elsevier BV]. *Ore Geology Reviews, 48*, 349–383. https://doi.org/10.1016/j.oregeorev.2012.05.004

Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. https://arxiv.org/abs/1412.6980

Kipf, T., & Welling, M. (2017). Semi-Supervised Classification with Graph Convolutional Networks. *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.1609.02907

Knox-Robinson, C. M., & Wyborn, L. A. I. (1997). Towards a holistic exploration strategy: Using Geographic Information Systems as a tool to enhance exploration [Publisher: Informa UK Limited]. *Australian Journal of Earth Sciences,*

44(4), 453–463. https://doi.org/10.1080/08120099708728326

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning [Publisher: Springer Science and Business Media LLC]. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Li, B., Yu, Z., & Ke, X. (2023). One-dimensional convolutional neural network for mapping mineral prospectivity: A case study in Changba ore concentration area, Gansu Province [Publisher: Elsevier BV]. *Ore Geology Reviews, 160*, 105573. https://doi.org/10.1016/j.oregeorev.2023.105573

Li, X., Xue, C., Chen, Y., Yuan, F., Li, Y., Zheng, C., Zhang, M., Ge, C., Guo, D., Lan, X., Tang, M., & Lu, S. (2023). 3D convolutional neural network-based 3D mineral prospectivity modeling for targeting concealed mineralization within Chating area, middle-lower Yangtze River metallogenic Belt, China [Publisher: Elsevier BV]. *Ore Geology Reviews, 157*, 105444. https://doi.org/10.1016/j.oregeorev.2023.105444

Loucks, R. R., & Fiorentini, M. L. (2023). Early zircon saturation in adakitic magmatic differentiation series and low Zr content of porphyry copper magmas [Publisher: Springer Science and Business Media LLC]. *Mineralium Deposita, 58*(8), 1381–1393. https://doi.org/10.1007/s00126-023-01208-5

Loucks, R. (2014). Distinctive composition of copper ore-forming arc magmas [Publisher: Informa UK Limited]. *Australian Journal of Earth Sciences, 61*(1), 5–16. https://doi.org/10.1080/08120099.2013.865676

Luffi, P., & Ducea, M. N. (2022). Chemical Mohometry: Assessing Crustal Thickness of Ancient Orogens Using Geochemical and Isotopic Data. *Reviews of Geophysics (Washington, D.C.: 1985), 60*(2), e2021RG000753. https://doi.org/10.1029/2021RG000753

Mantilla Dulcey, A. G. (2023). Predicción de la ocurrencia de depósitos minerales tipo pórfido usando técnicas de aprendizaje automático [Doctoral dissertation, Universidad Industrial de Santander].

McCuaig, T. C., Beresford, S., & Hronsky, J. (2010). Translating the mineral systems approach into an effective exploration targeting system [Publisher: Elsevier BV]. *Ore Geology Reviews, 38*(3), 128–138. https://doi.org/10.1016/j.oregeorev.2010.05.008

Murphy, K. (2012, January). *Machine learning: A probabilistic perspective* (Vol. 58).

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th International Conference on Machine Learning (ICML)*.

Perello, J., Carlotto, V., Zarate, A., Ramos, P., Posso, H., Neyra, C., Caballero, A., Fuster, N., & Muhr, R. (2003). Porphyry-Style Alteration and Mineralization of the Middle Eocene to Early Oligocene Andahuaylas-Yauri Belt, Cuzco Region, Peru [Publisher: Society of Economic Geologists, Inc.]. *Economic Geology, 98*(8), 1575–1605. https://doi.org/10.2113/gsecongeo.98.8.1575

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure [Publisher: Wiley]. *Ecography, 40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Rui, W., Linfu, X., Yongsheng, L., Jianbang, W., Qun, Y., & Xiangjin, R. (2025). Mineral prospectivity prediction based on the dynamic relation model Atten-GCN: A case study of gold prospecting in the Yingfengjie area, Shaanxi province (northern China) [Publisher: Elsevier BV]. *Ore Geology Reviews, 176*, 106399. https://doi.org/10.1016/j.oregeorev.2024.106399

Sun, T., Li, H., Wu, K., Chen, F., Zhu, Z., & Hu, Z. (2020). Data-Driven Predictive Modelling of Mineral Prospectivity Using Machine Learning and Deep Learning Methods: A Case Study from Southern Jiangxi Province, China [Publisher: MDPI AG]. *Minerals, 10*(2), 102. https://doi.org/10.3390/min10020102

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2019). CV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models (D. Warton, Ed.) [Publisher: Wiley]. *Methods in Ecology and Evolution, 10*(2), 225–232. https://doi.org/10.1111/2041-210x.13107

Wyborn, L. A. I., Heinrich, C., & Jaques, A. (1994). Australian Proterozoic mineral systems: Essential ingredients and mappable criteria, 109–115.

Yousefi, M., & Carranza, E. J. M. (2015). Fuzzification of continuous-value spatial evidence for mineral prospectivity mapping [Publisher: Elsevier BV]. *Computers & Geosciences, 74*, 97–109. https://doi.org/10.1016/j.cageo.2014.10.014

Zheng, C., Yuan, F., Luo, X., Li, X., Liu, P., Wen, M., Chen, Z., & Albanese, S. (2023). Mineral prospectivity mapping based on Support vector machine and Random Forest algorithm– A case study from Ashele copper–zinc deposit, Xinjiang,

NW China. *Ore Geology Reviews, 159*. https://doi.org/10.1016/j.oregeorev.2023.105567

Zuo, R., & Xu, Y. (2023). Graph Deep Learning Model for Mapping Mineral Prospectivity [Publisher: Springer Science and Business Media LLC]. *Mathematical Geosciences, 55*(1), 1–21. https://doi.org/10.1007/s11004-022-10015-z

Alejandro Iván Bueno Vera

Especialista en Ciencia de Datos con más de 5 años de experiencia en desarrollo de aplicaciones de inteligencia artificial y web.

Ingeniero Industrial.

Fernando Saez Rivera

Especialista en modelamiento y geoestadística con más de 23 años de experiencia en estimación de diferentes commodities en América.

Magister Ingeniero Geólogo.

Antoine Vincent Caté

Especialista en geología económica, geología estructural, geoquímica y ciencia de datos con más de 12 años de experiencia.

Geólogo.