

Predicción de la ocurrencia de depósitos minerales tipo pórfido usando técnicas de
aprendizaje automático

Ana Gabriela Mantilla Dulcey

Trabajo de Grado para Optar al Título de Geóloga

Director

Yesid Paul Goyes Peñafiel

Geólogo PhD (c)

Universidad Industrial de Santander

Facultad de Ingenierías Físicoquímicas

Escuela de Geología

Bucaramanga

2023

Cuando un científico distinguido pero anciano afirma que algo es posible, es casi seguro que tiene razón. Cuando afirma que algo es imposible, es muy probable que esté equivocado. La única manera de descubrir los límites de lo posible es aventurarse un poco más allá de ellos en lo imposible. Cualquier tecnología suficientemente avanzada es indistinguible de la magia.

Las tres leyes de Arthur C. Clarke

Dedicatoria

A mis padres, Amisadai Dulcey y Rafael Mantilla, quienes forjaron los cimientos de mi desarrollo y me apoyaron incondicionalmente en el inicio y culminación de mis estudios universitarios. En dedicación a su esfuerzo.

A mi familia, Omar, Isaias, Paula, Maité, Ivon, Yulieth, Nicolás, Samuel y Santiago, que me brindaron su amor, consejo y guía en mi desarrollo personal e intelectual. A mis nonitos, Maria Teresa Peña y Ramón Dulcey, por recargarme de energía y motivarme a ser mejor cada día.

Agradecimientos

A PhD (c) Yesid Paul Goyes Peñafiel, director de mi proyecto de grado, por su entrega incondicional en el desarrollo de este trabajo de investigación, por sus consejos y enseñanzas.

Al MsC Juan Martínez por sus valiosos comentarios que mejoraron sustancialmente el desarrollo de este proyecto.

Al Geólogo Luis Páez por sus sugerencias en el manuscrito.

A la Geóloga Rosana Báez por sus valiosos aportes en la presentación del trabajo.

A la Geóloga Alejandra Hernández por sus observaciones y correcciones que permitieron dar más claridad al libro.

Al Semillero de Investigación en Geofísica Aplicada y Computacional (SIGAC) por brindarme el apoyo y los recursos académicos necesarios para desarrollar este proyecto.

A la Universidad Industrial de Santander por contribuir a mi formación académica y profesional.

A los profesores de la Escuela de Geología que pusieron todo su esfuerzo y dedicación en apoyar la culminación de esta etapa de mi vida.

A mis amigos por su valiosa compañía y apoyo en esta etapa de aprendizaje.

Tabla de Contenido

Introducción	14
1 Planteamiento y justificación del problema.....	16
2 Objetivos.....	18
2.1 Objetivo general.....	18
2.2 Objetivos Específicos	18
3 Marco Teórico.....	19
3.1 Configuración geológica.....	19
3.1.1 Depósitos pórfidos de Yukón.....	24
3.2 Un enfoque para sistemas minerales tipo pórfido en Yukón	26
3.2.1 Escala de trabajo para los sistemas minerales tipo pórfido en Yukón	28
3.2.2 Criterios mapeables de exploración	29
3.3 Aprendizaje Automático.....	33
3.3.1 Redes neuronales artificiales	35
3.3.2 Bosques aleatorios.....	39
3.3.3 Máquinas de soporte vectorial	41
4 Metodología	44

4.1 FASE 1: Análisis de los insumos geológicos, geofísicos y geoquímicos relacionados con la ocurrencia de depósitos minerales tipo pórfido.....	45
4.2 FASE 2: Diseño modelos de aprendizaje automático usando redes neuronales artificiales, bosques aleatorios y máquinas de soporte vectorial en el lenguaje de programación Python. 46	46
4.3 FASE 3: Implementación de una metodología utilizando los modelos de aprendizaje automático y la información geológica y geofísica para predecir la probabilidad de ocurrencia de depósitos minerales tipo pórfido.	47
5 Resultados	48
5.1 Análisis de los insumos geológicos, geofísicos y geoquímicos relacionados con la ocurrencia de depósitos minerales tipo pórfido	48
5.1.1 Revisión del estado de arte en la predicción de depósitos minerales usando aprendizaje automático	48
5.1.2 Creación de la base de datos	53
5.1.3 Preprocesamiento de la base de datos	55
5.1.4 Análisis entre las variables de entrada	62
5.1.5 Creación de Ráster Virtual.....	71
5.2 Diseño modelos de aprendizaje automático usando redes neuronales artificiales, bosques aleatorios y máquinas de soporte vectorial en el lenguaje de programación <i>Python</i> . .	72
5.2.1 Redes Neuronales Artificiales (ANN).....	73

5.2.2	Bosques aleatorios (RF).....	77
5.2.3	Máquinas de soporte vectorial (SVM)	80
5.3	Implementación de una metodología utilizando los modelos de aprendizaje automático y la información geológica y geofísica para predecir la probabilidad de ocurrencia de depósitos minerales tipo pórfido.	82
5.3.1	Mapa de probabilidad con redes neuronales artificiales.....	83
5.3.2	Mapa de probabilidad con bosques aleatorios	85
5.3.3	Mapa de probabilidad con máquinas de soporte vectorial	86
5.3.4	Análisis descriptivo y comparativo de los mapas de probabilidad	87
5.3.5	Propuesta de metodología.....	88
6	Conclusiones	101
7	Recomendaciones y trabajo futuro	103
8	Repositorio y acceso a los datos.....	103

Lista de Tablas

Tabla 1 Principales eventos tectónicos y metalogénicos de Yukón	22
Tabla 2 Criterios mapeables de los sistemas minerales tipo pórfido en Yukón (Canadá)	31
Tabla 3 Funciones de activación para modelos ANN	37
Tabla 4 Funciones de pérdida usadas en problemas de probabilidad.....	38
Tabla 5 Funciones de Kernel.....	44
Tabla 6 Capas de entrada usadas en el estado de arte para generar MPM de depósitos tipo pórfido usando técnicas de aprendizaje automático.....	51
Tabla 7 Capas de entrada para los algoritmos ANN, RF y SVM.....	53
Tabla 8 Métodos usados y límites de detección de los elementos geoquímicos de interés	54
Tabla 9 Nombre de las variables de entrada en los algoritmos computacionales	59
Tabla 10 Interpretación de los valores del coeficiente de correlación de Pearson.....	64
Tabla 11 Hiperparámetros definidos en el modelo de ANN.....	73
Tabla 12 Resumen del entrenamiento con ANN.....	74
Tabla 13 Hiperparámetros definidos en el modelo de RF	77
Tabla 14 Resumen del modelo RF	79
Tabla 15 Hiperparámetros definidos en el modelo de SVM.....	80
Tabla 16 Resumen del modelo SVM.....	81

Tabla 17 Precisión alcanzada por cada modelo con el aumento de datos a 289 puntos de entrenamiento	89
Tabla 18 Comparación de probabilidades con el aumento de datos de entrenamiento	90
Tabla 19 Estadística zonal con el mapa de promedio aritmético	93
Tabla 20 Estadística zonal con el mapa de promedio geométrico	93

Lista de Figuras

Figura 1	Localización del área de estudio.....	19
Figura 2	Configuración de la zona de subducción para la formación de depósitos de pórfido	30
Figura 3	Superficie de decisión para tareas de clasificación usando técnicas de aprendizaje automático	34
Figura 4	Estructura y elementos que conforman un modelo de las redes neuronales artificiales	36
Figura 5	Estructura de un modelo de bosques aleatorios.....	40
Figura 6	Estructura de un modelo de máquinas de soporte vectorial.....	42
Figura 7	Metodología establecida para el proyecto actual	45
Figura 8	Características de las capas de salida con el procesamiento de los datos.....	56
Figura 9	Histogramas.....	59
Figura 10	Mapas predictores de depósitos tipo pórfido en Yukón.....	61
Figura 11	Análisis de dispersión para las variables de entrada.....	64
Figura 12	Matriz de correlación para las variables predictoras	65
Figura 13	Matriz de dispersión para las variables fuertemente correlacionadas	66
Figura 14	Problema de la colinealidad en tareas de clasificación.....	67
Figura 15	Porcentaje de varianza explicada acumulada con diez componentes principales	69
Figura 16	Matriz de correlación para las diez variables predictoras PC1-PC11.....	70
Figura 17	Mapas predictores con PCA.....	71

Figura 18 Resumen estadístico de la evolución de la precisión y valor de función de pérdida en cada entrenamiento ANN.....	75
Figura 19 Métricas del modelo ANN	75
Figura 20 Métricas de validación para ANN	76
Figura 21 Modelo grafo del modelo RF con máxima profundidad de 2	78
Figura 22 Métricas del entrenamiento del modelo RF. evolución de a) la precisión y b) la función de pérdida respecto al número de árboles de decisión	78
Figura 23 Métricas de validación RF. a) matriz de confusión y b) curva ROC.....	79
Figura 24 Métricas de validación SVM.....	81
Figura 25 Mapa de probabilidad con ANN.....	83
Figura 26 Mapa de probabilidad con RF	85
Figura 27 Mapa de probabilidad con SVM.....	86
Figura 28 Mapas de probabilidad con el aumento de datos de entrenamiento para los modelos de: a) ANN, b) RF y c) SVM.....	89
Figura 29 Mapa del promedio aritmético	92
Figura 30 Mapa del promedio geométrico.....	92
Figura 31 Relación de las variables de entrada con los polígonos de probabilidad >80%. a) X3, b) X4, c) X5, d) X6, e) X7, f) X8, g) X9, h) X10, i) X11.....	95
Figura 32 Metodología propuesta para la creación de mapas de prospectividad mineral con técnicas de aprendizaje automático.....	99

Resumen

Título: Predicción de la ocurrencia de depósitos minerales tipo pórfido usando técnicas de aprendizaje automático*

Autor: Ana Gabriela Mantilla Dulcey**

Palabras Clave: mapas de prospectividad mineral, aprendizaje automático, redes neuronales artificiales, bosques aleatorios, máquinas soporte vectorial, pórfidos, aprendizaje automático, probabilidad.

Descripción: Los mapas de prospectividad mineral son un insumo esencial para delimitar zonas potenciales fuentes de materia prima en el nuevo paradigma de transición energética. Las técnicas usadas generalmente consideran la regresión logística, los pesos de evidencia, los métodos heurísticos y la estadística bivariada. No obstante, en los métodos heurísticos se deben asignar pesos basados en el criterio experto, en los bivariados se deben clasificar las capas de valores continuos a valores categóricos o nominales, y finalmente la regresión logística considera una clasificación lineal de los datos. Por otro lado, los métodos de aprendizaje automático proporcionan un enfoque flexible y preciso en el uso de variables continuas y permiten encontrar modelos que aproximan de manera adecuada un conjunto de datos o factores condicionantes con valores de probabilidad relacionados con la ocurrencia mineral. En este trabajo se presenta una aplicación detallada y sistemática de los métodos de aprendizaje automático redes neuronales artificiales, bosques aleatorios y máquinas de soporte vectorial. Adicionalmente, se realizó un flujo preprocesamiento (estandarización, análisis de componentes principales, análisis de distribución) de las variables de entrada con el objetivo de representar el problema de predicción mineral con un problema de clasificación con superficies de decisión. Para comprobar la efectividad y el flujo de trabajo propuesto se analizó una zona del territorio de Yukón (Canadá), donde se inició a partir de una revisión exhaustiva de las variables o criterios mapeables de exploración relacionados con ocurrencia minerales. Los mapas de probabilidad calculados tuvieron una precisión promedio de 98.96% y la evaluación del rendimiento de los modelos de aprendizaje automático se realizaron de manera cuantitativa usando matrices de confusión, curvas ROC y precisión. Finalmente, para mejorar los modelos basados en información o criterio geológico, se propuso la técnica de aumento de datos y la aritmética de bandas entre los mapas de probabilidad, lo que permitió delimitar un total de 4 zonas potenciales. Por esta razón, la incorporación del aprendizaje automático en el flujo de trabajo de exploración mineral supone una mejora considerable en la optimización de recursos y el grado de confiabilidad en los objetivos de exploración.

* Trabajo de Grado

** Facultad de Ingenierías Físicoquímicas. Escuela de Geología. Director: Yesid Paul Goyes Peñafiel. Candidato a Doctor en Ciencias de la Computación.

Abstract

Title: Prediction of the occurrence of porphyry-type mineral deposits using machine learning techniques *

Author(s): Ana Gabriela Mantilla Dulcey¹

Key Words: mineral prospectivity maps, machine learning, artificial neural networks, random forests, support vector machines, porphyries, machine learning, probability.

Description: Mineral prospectivity maps are an essential input to delimit potential source areas of raw material in the new paradigm of energy transition. The techniques used generally consider logistic regression, weights of evidence, heuristic methods, and bivariate statistics. However, in heuristic methods, weights must be assigned based on expert judgment, in bivariate methods, the layers of continuous values must be classified to categorical or nominal values, and finally, logistic regression considers a linear classification of the data. On the other hand, machine learning methods provide a flexible and precise approach in the use of continuous variables and allow finding models that adequately approximate a set of data or conditioning factors with probability values related to mineral occurrence. In this work is presented a detailed and systematic application of machine learning methods artificial neural networks, random forest, and support vector machine. Additionally, a preprocessing flow (standardization, principal component analysis, distribution analysis) of the input variables was performed with the objective of representing the mineral prediction problem with a classification problem with decision surfaces. To verify the effectiveness and the proposed workflow, an area of the Yukon Territory (Canada) was analyzed, where an exhaustive review of the exploration variables or mappable criteria related to mineral occurrences began. The calculated probability maps had an average precision of 98.96% and the evaluation of the performance of the machine learning models was performed quantitatively using confusion matrices, ROC curves and precision. Finally, to improve the models based on geological information or criteria, the data augmentation technique, and the arithmetic of bands between the probability maps were proposed, which allowed delimiting a total of 4 potential zones. For this reason, the incorporation of machine learning in the mineral exploration workflow represents a considerable improvement in the optimization of resources and the degree of confidence in exploration targets.

* Bachelor Thesis

¹Faculty of Physicochemical Engineering. Department of Geology. Advisor: Yesid Paul Goyes Peñafiel. PhD Candidate in Computer Science.

Introducción

La demanda de materias primas en las sociedades modernas y el agotamiento de los recursos ya conocidos resaltan la importancia de la exploración mineral para encontrar nuevos depósitos (Brandmeier et al., 2020). Por eso, se requiere identificar y delimitar áreas con alta probabilidad de ocurrencias minerales mediante métodos que disminuyan el riesgo de fracaso y optimicen los recursos tanto económicos como operacionales.

Por consiguiente, es necesario coleccionar y analizar conjuntos de datos relacionados con la formación de depósitos minerales, por ejemplo, aspectos geológicos, geofísicos y geoquímicos, que puedan ser integrados en un mapa de prospectividad mineral (MPM) que contribuya en el proceso de toma de decisiones a diferentes escalas. Por lo tanto, se necesitan modelos robustos que integren grandes volúmenes de datos y extraigan la información más representativa relacionada con la búsqueda de objetivos de exploración. Algunos esfuerzos en el estado de arte incluyen el uso de técnicas como la regresión logística, pesos de evidencia, métodos heurísticos y estadísticos bivariados que, generalmente, no son capaces de aproximar de manera precisa la relación entre las variables geológicas como una función que calcule la probabilidad de ocurrencia mineral.

Sin embargo, los algoritmos de aprendizaje automático sí cumplen estas condiciones, por lo cual en distintas partes del mundo han tenido éxito en la predicción de probabilidad de zonas con potencial mineral. Por ejemplo, en China (Fu et al., 2021; Li et al., 2020a; Sun et al., 2019, 2020; Zhang et al., 2018), Irán (Keykhay-Hosseinpour et al., 2020a), Australia (Hariharan et al., 2017), Canadá (Lachaud et al., 2021b; Maepa et al., 2021; Maepa & Smith, 2017) y Brasil (Prado et al., 2020).

Dado que la predicción de depósitos minerales se considera un problema de clasificación binaria, entonces para complementar el proceso de generación de MPM se requiere desarrollar una metodología basada en la integración de insumos geológicos, geofísicos y geoquímicos con las ciencias computacionales. Debido a que los algoritmos de aprendizaje automático permiten establecer una relación no lineal entre las variables geológicas y la probabilidad, entonces la descripción cuantitativa de la predicción de ocurrencias minerales permitirá predecir nuevas áreas a partir del entrenamiento de modelos no lineales como lo son redes neuronales artificiales, bosques aleatorios y máquinas de soporte vectorial.

1 Planteamiento y justificación del problema

Los recursos minerales son una fuente importante de materia prima para el sector energético, industrial y social, por lo cual es necesario delimitar zonas de prospección mineral. Sin embargo, los métodos convencionales suelen invertir tiempo y presupuestos muy elevados en proyectos donde suelen recolectarse grandes volúmenes de datos que no pueden analizarse eficazmente sin un sistema de gestión adecuado que permita relacionar variables geológicas y probabilidad.

Los modelos de aprendizaje automático tienen la capacidad de organizar, transformar y analizar información en un tiempo corto, generando resultados que permiten delimitar zonas de prospección mineral con alto nivel de confiabilidad. Por lo tanto, el reto actual radica en desarrollar una metodología basada en la integración de información geológica y modelos de aprendizaje automático. La aplicación de estos modelos en los flujos de trabajo típicos de la industria minera permite realizar interpretaciones cuantitativas y elaborar mapas de prospectividad mineral basados en modelos matemáticos de alta precisión.

Sin embargo, el uso práctico de los algoritmos computacionales requiere de un análisis cuidadoso y detallado de las variables geológicas, geofísicas y geoquímicas. Pues, debido a que representan aspectos diferentes de la naturaleza que están relacionados con ocurrencias minerales, las variables deben ser tratadas de tal manera que sea posible establecer correlaciones directas con la probabilidad, ya que estos factores pueden impactar en el rendimiento de los modelos de aprendizaje automático y conducir a una interpretación errónea.

En este orden de ideas, el problema abordado es la predicción de zonas prospectivas para la exploración mineral, lo cual se considera un problema de clasificación en el cual se requiere identificar a qué clase pertenece el conjunto de datos, es decir, depósito (1) o no-depósito (0). La solución está dada por una superficie de decisión entre las dos clases objetivo, la cual puede ser estimada por modelos de aprendizaje automático tales como redes neuronales artificiales, máquinas de soporte vectorial y bosques aleatorios, los cuales por excelencia han sido usados en el estado de arte para clasificar, pues tienen la capacidad de representar correlaciones no lineales.

Este trabajo de investigación surge de la necesidad de integrar los algoritmos computacionales con el análisis geológico para predecir la ocurrencia de depósitos minerales tipo pórfido. Para esto se tuvieron en cuenta factores que generalmente afectan el entrenamiento adecuado de los modelos como, por ejemplo, la selección adecuada de las variables geológicas, el preprocesamiento de los datos (estandarización, interpolación, filtrado, transformación), la escala de trabajo, los puntos de no-depósito, el análisis estadístico de los factores condicionantes, el ajuste de hiperparámetros, la comparación entre modelos computacionales y la delimitación de objetivos de exploración.

2 Objetivos

2.1 Objetivo general

Predecir la ocurrencia de depósitos minerales tipo pórfido a partir de la integración de información geológica, geoquímica y geofísica usando técnicas de aprendizaje automático.

2.2 Objetivos Específicos

1. Analizar los diferentes insumos geológicos, geoquímicos y geofísicos que se relacionen con la ocurrencia de depósitos minerales pórfido (sección 5.1).
2. Comparar el rendimiento y capacidad de predicción de depósitos minerales tipo pórfido de los modelos de aprendizaje automático usando redes neuronales artificiales, bosques aleatorios y máquinas de soporte vectorial (sección 5.2).
3. Implementar una metodología utilizando los modelos de aprendizaje automático y la información geológica, geoquímica y geofísica para predecir la ocurrencia de depósitos minerales tipo pórfido (sección 5.3).

3 Marco Teórico

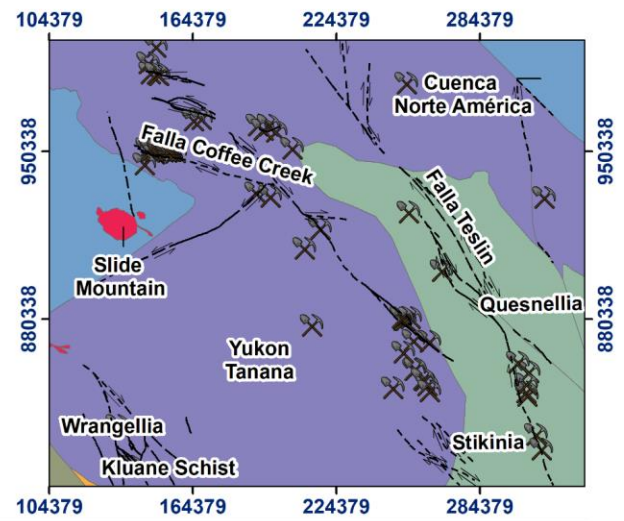
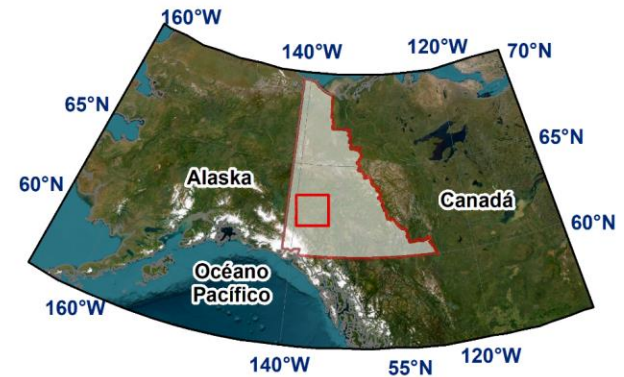
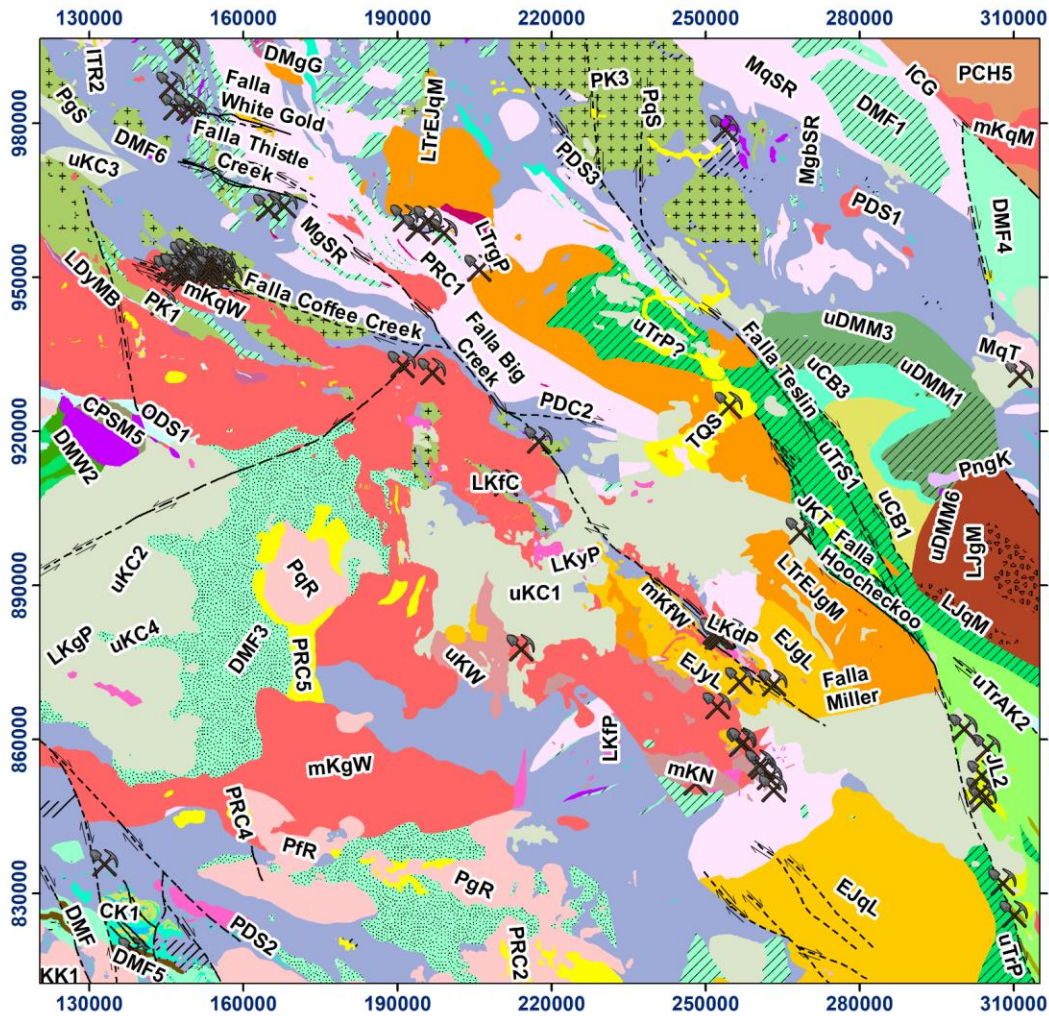
3.1 Configuración geológica

El área de estudio se encuentra ubicada al oeste del territorio de Yukón, Canadá (Figura 1). Hace parte del dominio de terrenos acrecionados de la Cordillera Canadiense, la cual tiene una tendencia noroeste a lo largo del margen occidental de Norte América (Mihalasky et al., 2011). Ha evolucionado como un cinturón activo a través de diversos estilos tectónicos y configuraciones paleogeográficas que proporcionan un excelente estudio de la forma en que la tectónica gobierna la metalogénesis (Nelson & Colpron, 2007) dando lugar a un entorno altamente fértil que aloja cuerpos minerales de clase mundial como pórfidos, epitermales, skarn, MVT, VMS, entre otros (Nelson & Colpron, 2007). Por tal razón, su configuración geológica ha sido ampliamente estudiada (Mortensen, 1992; Mortensen & Friend, 2021; J. L. Nelson et al., 2006).

Existen principalmente cinco dominios tectónicos que conforman la Cordillera y son divididos por Nelson & Colpron (2007) de la siguiente manera: 1) América del Norte Ancestral (Laurentia), 2) terrenos Intermontanos, 3) terrenos Insular y Farewell, 4) Alaska Ártica y 5) terrenos de acreción tardía (Figura 1). En este trabajo se hizo énfasis en los depósitos tipo pórfido de cobre que se distribuyen en los dominios Intermontanos donde se encuentra el terreno acrecionado de Yukón-Tanana (YTT) y Stikinia, América del Norte Ancestral que contiene la cuenca Selwyn y el terreno Windy- McKinley que hace parte del dominio Ártico según la configuración litotectónica de Colpron (2006), Silberling et al., (1984) y Wheeler et al., (1991).

Figura 1




























Localización del área de estudio



CONVENCIONES	FALLAS		TERRENOS	
OCURENCIAS MINERALES	- - - normal, cubierta	—> rumbo, sinistral, definida	Slide Mountain	Stikinia
⌘ Ocurrencia mineral tipo pórfido	— normal, definida	—> cabalgamiento, definido	Kluane Schist	Wrangellia
	— inversa, definida	- - - aproximada	Cuenca América del Norte	Yukon Tanana
	— rumbo, dextral, definida	- - - cubierta	Quesnellia	
		— definida		



Escala unidades geológicas 1:250.000

UNIDADES GEOLÓGICAS DE OCURRENCIAS MINERALES TIPO PÓRFIDO		
<p>CUATERNARIO</p> <p> TQS: SELKIRK. Flujos basálticos</p>	<p>JURÁSICO SUPERIOR Y CRETÁCICO INFERIOR</p> <p> JKT: TANTALUS Conglomerado de guijarros de pedernal y arenisca arenosa de cuarzo, pedernal y feldespatos</p>	<p> LTrD: DOGHEAD. Peridotita serpentinizada, gabro y plagiogranito</p> <p> uTrPm: POVOAS?. Anfíbolita variablemente migmatizada y esquisto qtz-pl-bt; balsas en plutones jurásicos</p>
<p>PALEOCENO A EOCENO INFERIOR</p> <p> PgR: SUITE RUBY RANGE. Granodiorita Bt-Hbl</p>	<p>JURÁSICO INFERIOR Y MEDIO, HETTANGIANO A BAJOCIANO</p> <p> JL2: TANGLEFOOT. Arenisca arcósica y conglomerado de lutitas, guijarros y cantos rodados</p> <p> IJN: NORDENSKIOLD. Toba de dacita verde caqui y arenisca volcanoclástica</p>	<p>PÉRMICO MEDIO A SUPERIOR</p> <p> PqS: SUITE SULPHUR CREEK. Granito de augen de feldespato K, metapórfido</p> <p> PK1: ESQUISTO KLONDIKE. Esquisto de cuarzo, moscovita y clorita</p>
<p>CRETÁCICO SUPERIOR</p> <p> LKdP: SUITE PROSPECTOR MOUNTAIN. Gabro cristalino grueso y diorita</p> <p> LKgP: SUITE PROSPECTOR MOUNTAIN. Granodiorita Hbl-Bt, diorita Hbl, diorita de cuarzo, monzonita de cuarzo</p>	<p>JURÁSICO INFERIOR</p> <p> EJqL: LONG LAKE. Cuarzo monzonita a granito de Bt, Bt-Ms y Bt-Hbl</p> <p> EJgL: SUITE LONG LAKE Granodiorita de Bt-Hbl masiva a débilmente foliada</p> <p> EJyL: SUITE LONG LAKE Sienita de Hbl porfídica</p>	<p>MISSISSIPPIANO</p> <p> MgSR: SUITE SIMPSON RANGE. Metagranodiorita, metadiorita y metatonalita que contienen Hbl</p>
<p>CRETÁCICO SUPERIOR</p> <p> uKc1: CARMACKS Basalto de augita-olivino y brecha</p>	<p>TRIÁSICO SUPERIOR A JURÁSICO INFERIOR</p> <p> LTrEJgM: SUITE MINTO Granodiorita de Bt-Hbl foliada; Pantallas ricas en Bt y schlieren gneisicos</p>	<p>DEVÓNICO, MISSISSIPPIANO Y MAYOR (?)</p> <p> DMF6: FINLAYSON. Rocas ultramáficas, serpentinita; metagabro</p> <p> DMF3: FINLAYSON. Rocas metasedimentarias carbonáceas</p> <p> DMF1: FINLAYSON. Rocas volcánicas y volcanoclásticas intermedias a máficas</p>
<p>CRETÁCICO MEDIO</p> <p> mKfW: SUITE WHITEHORSE Pórfido de cuarzo-feldespatos, pórfido de feldespatos y hornblenda</p> <p> mKgW: SUITE WHITEHORSE Granodiorita de Bt-Hbl, cuarzo diorita Hbl y diorita Hbl</p> <p> LKfC: SUITE CASINO Pórfido de cuarzo-feldespatos</p> <p> mKqW: SUITE WHITEHORSE Cuarzo monzonita de Bt, granito de Bt y leucogranito</p> <p> mKN: MOUNT NANSEN Flujos masivos de andesita afírica o feldespatos-fírica a dacita</p>	<p>TRIÁSICO SUPERIOR, CARNIANO Y MAYOR (?)</p> <p> uTrP?: POVOAS. Flujos de basalto andesítico</p>	<p>NEOPROTEROZOICO A PALEOZOICO</p> <p> PDS1: SNOWCAP. Cuarzita, psammita, pelita y mármol; anfíbolita</p>

Nota. En la leyenda se incluyen solamente las unidades geológicas relacionadas con la ocurrencia de depósitos minerales tipo pórfido en la zona de estudio. El número de depósitos en cada terreno se distribuye así: Cuenca Norte América (5), Yukon Tanana (98), Stikinia (45). La información sobre las unidades geológicas, fallas, depósitos y terrenos fueron tomados de la página oficial del Servicio Geológico de Yukón (<https://yukon.ca/en/>).

La evolución de los estilos tectónicos de la Cordillera y la formación de los terrenos acrecionados hacen parte de la historia metalogénica de Yukón que abarca desde la cratonización de Laurentia en el Proterozoico hasta el modelo de subducción actual que ha dado lugar a unidades geológicas que hospedan depósitos de gran interés económico (Nelson & Colpron, 2007). En la Tabla 1 se presenta una síntesis de la historia tectónica y metalogénica de Yukón.

Tabla 1

Principales eventos tectónicos y metalogénicos de Yukón

Período / Época	Edad (Ma)	Historia metalogénica	Configuración tectónica
Paleoceno-Holoceno	66.0 a la actualidad	- Extinción de los cinturones magmáticos. La subducción frente a la costa Oeste de Canadá cesó y la de la placa Juan de Fuca y el Pacífico continuó. - Los pórfidos de Cu-Au-Mo y Mo se concentraron cerca de las fallas transcurrentes y entre los campos volcánicos. También se formaron epitermales de Au-Ag en Stikinia. - Extensión de la corteza y procesos de exhumación. Movimiento dextral de 430-490 km de la falla Tintina e inicio de actividad de la falla Denali cuyo movimiento fue de 370 km. - Ruptura de configuraciones tectónicas y magmáticas previas. Se generaron unidades félsicas de afinidad shoshonítica.	Arco continental
Cretácico Superior	100.5 a 66.0	-Durante el Cretácico el terreno Alaska Ártica rotó 30° en sentido horario. -Formación de pórfidos importantes de Mo y Cu-Mo-Au como El Casino (73 ma) en Yukón. -El régimen predominante de YTT fue extensivo. Se emplazó el batolito Cassiar (110 ma) y las unidades Mayo, Tombstone y Tungsten producto de la fundición litosférica. -Se desarrolló el Complejo Coast como el principal cinturón en la zona.	Arco de isla

Jurásico Superior- Cretácico Inferior	163.5 ±1.0 a 100.5	-Se generó el primer episodio magmático después de la acreción, el cual se caracterizó por vulcanismo e intrusiones de expansión limitada, lo que dio lugar a depósitos de Mo.
Jurásico Inferior- Jurásico Medio	174.1 ±1.0 a 163.5 ±1.0	-Stikinia choca con Quesnel, el cual ya había acrecionado con América del Norte. Desaparecieron los arcos asociados con estos dos terrenos y el prisma de acreción Cache Creek fue atrapado.
Pérmico -Jurásico Inferior	273.01 ±0.14 a 174.1 ±1.0	-Formación de depósitos pórfidos de Cu-Au y Cu-Mo en Stikine, Quesnel y YTT, hospedados en plutones de afinidad alcalina a calcoalcalina (shoshonítica). La variación se debe probablemente a la fusión del manto metasomatizado durante eventos repetitivos de subducción en el Devónico-Triásico. -Acreción de terrenos pericratónicos (Slide Mountain, Quesnel y YTT). Se produce la migración de los cinturones magmáticos. -Ruptura tectónica, magmática y deposicional
Pennsylvaniano- Pérmico	323.2 ±0.4 a 273.01 ±0.14	-Predominaron secuencias de depósitos volcánicos – intermedios que se depositaron en YTT, Stikine y Quesnel. -El margen tectónico continental fue moderado, dando lugar a estratos más delgados. Hubo mineralización epigenética y tipo pórfido.
Devónico- Mississippiano	419.2 ±3.2 a 323.2 ±0.4	-Durante el Mississippiano temprano se generó un período de extensión que generó que el océano Slide Mountain se abriera, separando una parte del arco frontal del margen continental de Laurentia y formando el terreno Yukón-Tanana (YTT). - Del Devónico medio-tardío surgió magmatismo de arco en la Cordillera. Presentó magmatismo bimodal félsico y máfico. El primero asociado con la corteza continental y el segundo con magmas tipo MORB, OIB, calcoalcalinos, toleíticos, entre otros.

Cámbrico- Silúrico	538.8 ±0.2 a 419.2 ±3.2	<ul style="list-style-type: none"> - Ocurrieron tres pulsos magmáticos (Cámbrico, Ordovícico medio y tardío) que dieron lugar a depósitos tipo SEDEX, diatremas, rocas volcánicas ultrapotásicas, mineralizaciones cupríferas (BIF). - Se generaron dos episodios de rift y magmatismo alcalino intermitente debido a procesos de extensión. - Del Neoproterozoico al Cámbrico se formó el margen occidental de Laurentia al separarse de Rodinia. Durante este evento un fragmento de corteza de Siberia se desprendió y formó el terreno Ártico. 	Rifting
--------------------	-------------------------	---	---------

Nota. La información referente a los eventos tectónicos y metalogénicos fue adaptada de Nelson & Colpron (2007).

3.1.1 Depósitos pórfidos de Yukón

Los depósitos tipo pórfido de Yukón se formaron durante dos períodos de tiempo: del Triásico Superior-Jurásico Medio y el Cretácico Superior-Eoceno. El primer grupo conforma los pórfidos alcalinos emplazados durante los eventos de acreción al margen occidental de Norte América en una configuración tectónica de arco de isla. Ocurren mayoritariamente en los terrenos de Wrangellia, Stikinia y Quesnellia (McMillan, 1991). También se encuentran depósitos calcoalcalinos, pero predominan las configuraciones de Cu-Au que ocurren en intrusiones porfiríticas que consisten en zonas de rocas fracturadas y brechas de las cuales se extraen principalmente cobre y oro (McMillan, 1991).

El segundo grupo contiene los pórfidos calcoalcalinos post-acrecionarios que se emplazaron durante una configuración de arco continental. Dentro de este grupo se encuentran tres subtipos de pórfidos: Cu-Au, Cu-Au-Mo y Cu-Mo cuya distribución y temporalidad reflejan las principales estructuras de la corteza y geometría de la subducción (McMillan, 1991). Se encuentran en vetas y venillas mineralizadas, con fracturas estrechas, stockworks y cuerpos de brecha económicamente explotables por su contenido en cobre, molibdeno y oro (McMillan, 1991).

Debido a la configuración tectónica de arco de isla y arco continental que permitió la formación de pórfidos alcalinos y calcoalcalinos desde el Triásico hasta el Eoceno, el emplazamiento de los depósitos se vio favorecido por el movimiento de las fallas Tintina y Denali cuyo carácter regional dio paso a la generación de fallas secundarias que juegan un papel importante en la disposición de las ocurrencias minerales en la Cordillera Canadiense.

En resumen, el desarrollo metalogénico de la Cordillera data desde el desprendimiento de Laurentia del supercontinente Rodinia a finales del Neoproterozoico y principios del Cámbrico. A partir de entonces se desarrollaron múltiples episodios de rift y pulsos magmáticos que generaron yacimientos de gran interés económico. Según Nelson & Colpron (2007), el período que abarca desde mediados del Triásico se denomina "la edad de los pórfidos", asociado a un régimen tectónico de compresión y acreción de los arcos Intermontanos que continuaron hasta el Cretácico tardío.

Las unidades que hospedan estos depósitos en la zona de estudio están en la leyenda del mapa de la Figura 1 las cuales corresponden principalmente a rocas ígneas y metamórficas como granodioritas, dioritas, monzodioritas, esquistos, ortoneis, cuarzomonzonitas, granitos y pórfidos, así como basaltos, anfibolitas y peridotitas en las que se encuentran los elementos de interés económico. La mayoría son de edad Mesozoica, pero se distinguen algunos que pertenecen al Paleozoico.

Dado que la mayor parte de los depósitos tipo pórfido se hospedan en rocas plutónicas y, en menor medida, volcánicas, en Mihalasky et al., (2011) se proponen dos zonas de influencia de 10 km y 2 km para las rocas intrusivas y volcánicas, respectivamente. Los criterios para la selección de estas distancias son:

A) Incertidumbre relacionada con la verdadera ubicación de los contactos de rocas ígneas mapeadas.

B) Los contactos intrusivos generalmente se inclinan hacia el exterior y los depósitos de pórfido pueden formarse alrededor de cuerpos intrusivos y tener zonas de alteración de hasta 10 km (Singer et al., 2008).

C) Los cuerpos de rocas volcánicas pueden tener bordes relativamente delgados que pueden ser discontinuos, estar cubiertos o no mapearse a la escala de los mapas geológicos.

D) Las ocurrencias significativas de depósitos minerales alojados en rocas volcánicas similares a los pórfidos se encuentran alojados en rocas volcánicas dentro de un radio de 10 km de un cuerpo plutónico.

E) Un radio de 10 km alrededor de un plutón abarca el sistema potencialmente mineralizante, es decir, la circulación hidrotermal a escala local o de distrito.

F) Un radio de 2 km alrededor de las unidades volcánicas representa que las intrusiones son pequeñas y de extensión limitada.

Las zonas de influencia desarrolladas por Mihalasky et al., (2011) fueron hechas con base en mapas geológicos a escalas no mayores de 1:250.000, por lo cual los diques y alteraciones hidrotermales no están representadas.

3.2 Un enfoque para sistemas minerales tipo pórfido en Yukón

El enfoque de sistemas minerales se centra en que las ocurrencias minerales son la respuesta a procesos geológicos que operan a una escala espacial y un marco de tiempo más grandes que un depósito en sí mismo (Hagemann et al., 2016) por lo cual es de suma importancia reconocer y comprender los criterios que dan lugar a la formación de depósitos minerales, pues los

mapas de prospectividad generados partir de algoritmos computacionales pueden estar sesgados si el conjunto de datos de entrada es inadecuado. Con base en lo anterior, se presenta la metodología propuesta por Mccuaig et al., (2010) para traducir un modelo de sistema mineral conceptual en un sistema de objetivos de exploración. Para esto se tienen en cuenta cuatro pasos que implican definir: a) procesos críticos, b) procesos constituyentes, c) elementos objetivo y d) criterios mapeables.

En primer lugar, los procesos críticos son aquellos esenciales para que se genere un depósito mineral, de tal modo que tienen un enfoque probabilístico en el cual se afirma que, si la probabilidad de ocurrencia de cualquiera de los procesos críticos se vuelve cero, entonces no habrá depósito mineral (Mccuaig et al., 2010). En segundo lugar, los procesos constituyentes dependen del tipo de mineralización de interés y pueden conducir a que ocurra un proceso crítico específico. En tercer lugar, los elementos objetivo son aquellas características geológicas producto de los procesos constituyentes.

Debido a que los procesos no pueden mapearse sino sus resultados, entonces los parámetros que condicionan un sistema mineral de interés pueden reconocerse mediante proxis. Por último, los criterios mapeables corresponden al conjunto de datos de exploración que están disponibles ya sea directa o indirectamente. A estos se les llama “mapas predictores”.

La selección de los mapas predictores involucra tener en cuenta ocho procesos relacionados con la ocurrencia de pórfidos en la zona de estudio que son descritos por Hagemann et al., (2016): 1) eventos tectónicos que desencadenan y definen ventanas temporales para eventos de mineralización, 2) procesos geoquímicos y tectónicos que producen “regiones fuente” enriquecidas con metales, fluidos hidrotermales y magmas capaces de movilizar estos metales, 3)

procesos tectónicos, sedimentarios, diagenéticos y estructurales que producen conductos que pueden actuar como vías para el flujo de fluidos/magma en escala litosférica, cortical, provincial y distrital, 4) procesos que impulsan el flujo de fluidos, 5) mecanismos para concentrar los fluidos en zonas “trampa”, 6) procesos físicos y químicos que generen la deposición de los metales en la zona “trampa”, 7) procesos geoquímicos que permitan la detección de estas concentraciones y 8) procesos que favorezcan la exhumación, preservación y el mejoramiento del material mineralizado.

Debido a que los factores geológicos nombrados anteriormente pueden operar en una variedad de escalas, ya sea regional, de distrito o de prospecto, Mccuaig et al., (2010) establece que la relevancia de los parámetros del sistema mineral cambiará debido a: 1) la escala a la cual operan los procesos críticos del sistema mineral y 2) la disponibilidad de datos geocientíficos a partir de los cuales se pueden identificar los criterios mapeables. Por consiguiente, a continuación, se establece la escala de trabajo para la zona de estudio.

3.2.1 Escala de trabajo para los sistemas minerales tipo pórfido en Yukón

El polígono de estudio tiene un área de 36524.07 km² y se seleccionó teniendo en cuenta uno de los criterios propuestos por Hronsky & Groves (2008) en el cual las distribuciones de frecuencia de tamaño de las provincias metalogénicas del tipo de depósito objetivo deben ser agrupadas. En este caso, las ocurrencias minerales tipo pórfido se concentran sobre todo en los terrenos acrecionados de Yukón, específicamente sobre Yukón-Tanana (YTT), Stikinia, cuenca Selwyn y el terreno Windy- McKinley (Figura 1).

En este orden de ideas, la escala de trabajo para la zona de interés corresponde a 1:250.000 y, por lo tanto, para definir las capas de entrada se analizaron los procesos críticos y constituyentes

a escala regional, los cuales implican ambiente geodinámico, orógenos, procesos de subducción, evidencias de mezcla de magmas y metasomatismo, series magmáticas, entre otros.

En resumen, teniendo en cuenta los ocho criterios que involucran los sistemas minerales propuestos por Hagemann et al., (2016) y la importancia de comprender los elementos que hacen parte de estos, se generaron los objetivos mapeables de exploración con base en la metodología de Mccuaig et al., (2010) para resaltar los procesos que operan a escala regional 1:250.000 y que son factores representativos para la formación de este tipo de ocurrencias minerales.

3.2.2 Criterios mapeables de exploración

El desarrollo de los objetivos de exploración y criterios mapeables, que sirvieron como insumo para la generación de las capas de entrada de los algoritmos computacionales, se establecieron con base en la escala de trabajo y los procesos de fuente, transporte, trampa y precipitación de los depósitos minerales de interés.

3.2.2.1 Procesos fuente

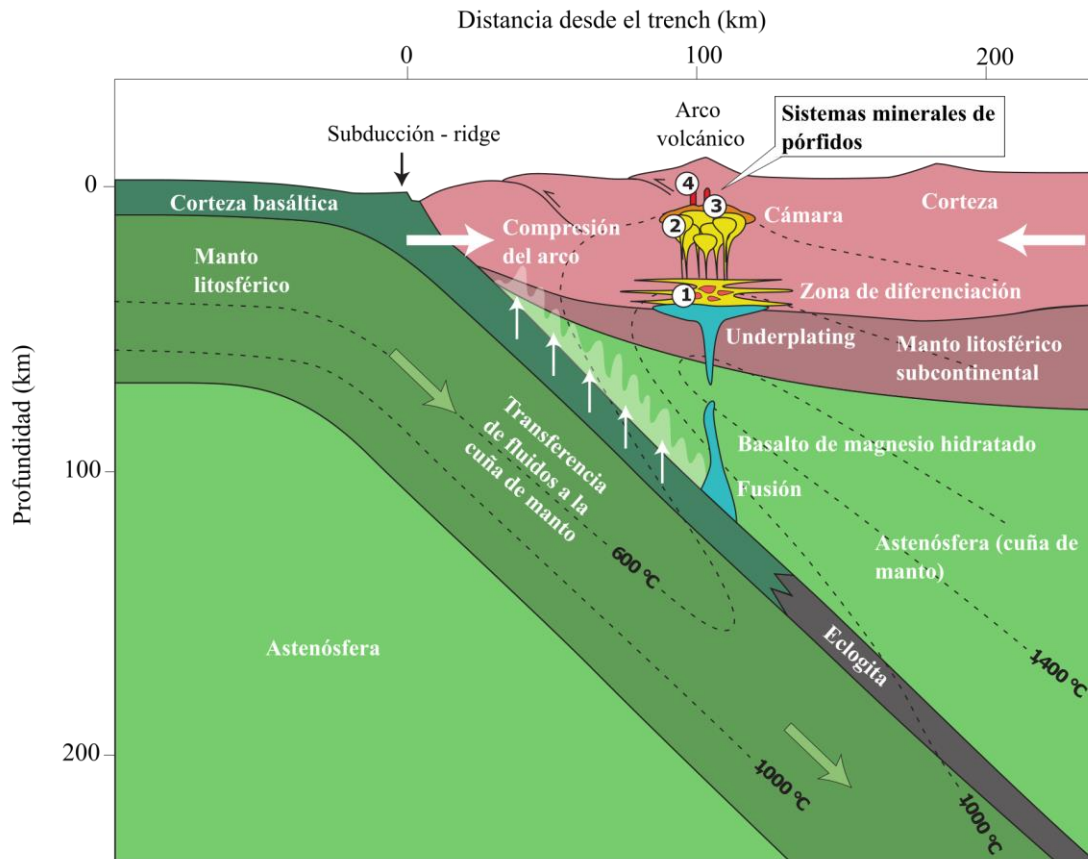
Los pórfidos se forman a partir de fluidos hidrotermales de magmas alcalinos y/o calcoalcalinos ricos en agua y emplazados en arcos volcanoplutónicos a lo largo de zonas de subducción en márgenes convergentes (Richards, 2009) durante períodos discretos de levantamiento regional, acortamiento tectónico y engrosamiento de la corteza (Camus, 2005). Para más información sobre la fuente de los fluidos hidrotermales puede dirigirse a Wilkinson (2013).

Los fluidos ascienden hasta el manto litosférico subcontinental donde se produce la cristalización fraccionada (proceso 1 en la Figura 2), el ascenso y fraccionamiento de fundidos magmáticos de composición andesítica a dacítica (proceso 2 en la Figura 2), emplazamiento de cuerpos intrusivos y vulcanismo en la corteza superior (procesos 3 y 4 en la Figura 2). Por esto, se

consideraron las rocas volcánicas y plutónicas de composición félsica e intermedia como capas de entrada.

Figura 2

Configuración de la zona de subducción para la formación de depósitos de pórfido



Nota. Los cuatro procesos desencadenantes de estos sistemas minerales están diferenciados por (1), (2), (3) y (4), respectivamente. Adaptado de Wilkinson (2013).

3.2.2.2 Procesos de transporte

Cuando se separa una fase magmática volátil y se produce una sobresaturación que excede las presiones litostáticas de las rocas suprayacentes, se generan fracturas que permiten el escape

de fluidos hidrotermales ricos en Au, Cu, Ag, Mo, entre otros elementos (Sinclair, 2007). Estos se ven favorecidos por la actividad de fallas pre-existentes que forman conductos verticales de alta permeabilidad y las estructuras translitosféricas paralelas al orógeno (Richards, 2003). Por esto, se tuvieron en cuenta la proximidad a las fallas geológicas y a los contactos de terrenos litotectónicos. La selección de estas dos capas de entrada se soporta en la sección 3.1.1.

3.2.2.3 Trampas físicas

Las rocas anfitrionas (o roca caja) pueden impedir el escape de fluidos y la dispersión de metales en otros sistemas (Wilkinson, 2013). Son barreras físicas que generan el estancamiento del magma y, por lo tanto, contrastes de competencia entre las unidades litológicas. Los métodos geofísicos de magnetometría permiten detectar bordes o discontinuidades asociadas a estas estructuras. Por eso, el mapa de campo total magnético residual y su primera derivada vertical se consideraron como capas de entrada.

3.2.2.4 Precipitación química

La precipitación de los minerales en sistemas pórfidos se presenta de manera detallada en Sinclair (2007), Kreuzer et al., (2015) y Wilkinson (2013). Los mecanismos de precipitación generan grandes halos de dispersión química cuyos principales indicadores en sistemas tipo pórfido son Pb y Zn como elementos secundarios (Leybourne, 2007, Kreuzer et al., 2015); Au, Ag, Cu y Mo (Sillitoe, 2008). Con base en la disponibilidad y calidad de datos se seleccionaron los siguientes indicadores como mapas predictores: Au, Cu, Pb, Mo y Zn. En la Tabla 2 se detallan los procesos que dan lugar a los criterios mapeables.

Tabla 2

Criterios mapeables de los sistemas minerales tipo pórfido en Yukón (Canadá)

Procesos críticos	Fuente (magma, H ₂ O)	Vía activa/conducto	Trampa física	Precipitación química
Procesos constituyentes	(1) Deshidratación de la placa oceánica por el proceso de subducción. Liberación de agua y otros elementos (CO ₂ , Fe, Pb, Th).	(1) Separación de una fase magmática volátil enriquecida en H ₂ O, CO ₂ , HCl, H ₂ S, SO ₂ que transporten los metales.	(1) Generación de zonas de dilatación durante la deformación tectónica.	(1) Expansión adiabática, separación de fases, interacción de los fluidos hidrotermales con la roca caja, mezcla con agua meteórica y/o neutralización de acidez.
	(2) Mezcla de magma basáltico con la corteza continental en la cuña de manto astenosférico y la litósfera del manto subcontinental.	(2) Sobresaturación de volátiles que exceden la presión litostática de rocas suprayacentes y generan fracturas.		
	(3) Oxidación y enriquecimiento en volátiles y sílice de los fluidos hidrotermales.	(3) Cambios en el entorno la subducción, lo que facilitan los conductos permeables.	(3) Contactos litológicos con diferentes contrastes de competencia constituyen las principales trampas para los fluidos metalíferos.	
	(4) Cristalización fraccionada: magmas de composición intermedia-félsica (andesíticos a dacíticos).	(4) Actividad de fallas pre-existentes por los eventos de acreción de terrenos litotectónicos.		
	(5) Disposición de metales (Au y Cu provienen del manto, Mo de la corteza).			
Elementos objetivo	Emplazamiento de cuerpos intrusivos y vulcanismo en la corteza superior.	Fallas geológicas. Terrenos litotectónicos.	Contrastes litológicos y respuestas magnéticas.	Halos de dispersión química.
Mapas predictores	Proximidad a los cuerpos intrusivos y rocas volcánicas.	- Proximidad a fallas geológicas. - Proximidad a contactos de terrenos litotectónicos.	- Proximidad a contactos litológicos. - Campo total magnético residual y primera derivada vertical.	Concentración geoquímica de Au, Cu, Pb, Mo, Zn.

3.3 Aprendizaje Automático

El aprendizaje automático (*ML*, por sus siglas en inglés), es una rama de las ciencias computacionales que se enfoca en estimar estadísticamente funciones complejas y probar intervalos de confianza alrededor de estas funciones (Goodfellow et al., 2016). Esto lo hace con el uso de algoritmos que son capaces de aprender de un conjunto de datos de entrada para resolver una tarea específica mediante modelos matemáticos. En este caso, los datos son una colección de características de un objeto o evento que generalmente se representa como un vector $x \in \mathbb{R}$ que contiene las características o atributos de entrada.

Las tareas más comunes que pueden resolver los algoritmos de *ML* son: clasificación, clasificación con valores faltantes, regresión, transcripción, traductor, salida estructurada, detección de anomalías, síntesis y muestreo, eliminación de ruido, estimación de la función de masa de probabilidad, entre otras que se detallan en (Goodfellow et al., 2016). En este caso, la predicción de depósitos minerales consiste en un problema de clasificación binaria, en el cual los datos tienen una etiqueta de 1 (depósito) y 0 (no-depósito).

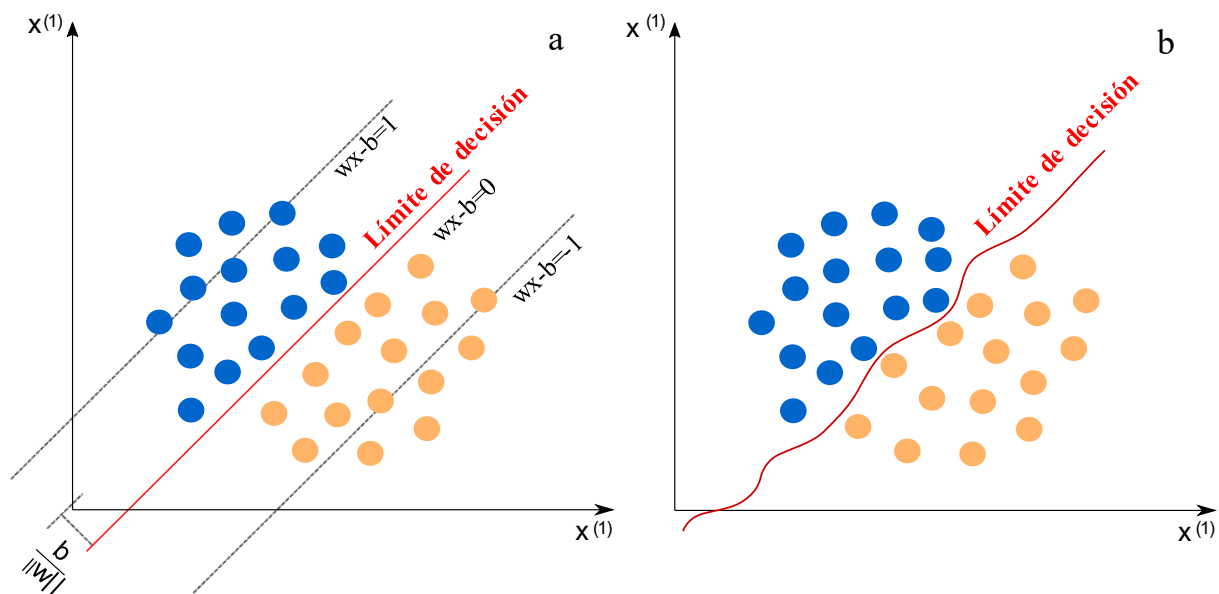
El problema de clasificación puede abordarse teniendo en cuenta que cada uno de los datos tiene unas características espaciales, por lo tanto, si un dato está bajo o sobre una región en particular, se considerará de la clase A o B (depósito y no-depósito). Es decir, para separar las clases se genera un límite de decisión que, según el número de dimensiones, puede ser un plano o hiperplano.

Por excelencia los algoritmos usados para resolver tareas de clasificación son: redes neuronales artificiales, bosques aleatorios y máquinas de soporte vectorial. El método para generar la superficie de decisión de cada uno es distinto y, por lo tanto, también su superficie de decisión.

Por ejemplo, para las máquinas de soporte vectorial está representado por una línea recta, pero para las redes neuronales artificiales y bosques aleatorios puede ser curva dependiendo de la distribución de los datos. Con base en la superficie de decisión se determina la precisión del modelo. En la Figura 3 se ilustran dos tipos de líneas divisorias.

Figura 3

Superficie de decisión para tareas de clasificación usando técnicas de aprendizaje automático



Nota. Formas que puede tener el límite de decisión en un algoritmo computacional. a) La línea que divide los datos es una línea recta que está dada por $wx-b=0$. Este es un ejemplo típico del algoritmo de máquinas de soporte vectorial. b) La línea que divide los datos es una línea curva. Adaptado de Burkov (2019).

Para calcular la probabilidad de ocurrencia mineral, se debe establecer una función f cuyo dominio x sean las variables geológicas (geología, geofísica, geoquímica) y la salida y sea la probabilidad. Sin embargo, f es una función no trivial que no puede expresarse en términos de funciones conocidas.

Algunos esfuerzos en el estado de arte para aproximar esta función incluyen métodos bivariados, heurísticos, pesos de evidencia y regresión logística. Sin embargo, estos no expresan un sentido matemático que permita extraer las características de las variables. Por ejemplo, en regresión logística se conoce f . No obstante, se considera que el conjunto de datos está separado linealmente, lo cual representa una desventaja en el uso de variables geológicas que, generalmente, no están correlacionadas entre sí. Por lo tanto, una forma de encontrar la aproximación para f es mediante el uso de algoritmos de aprendizaje automático. Estos modelos pueden ser representados como una función $M(x)$. A continuación, se explica cómo los algoritmos de redes neuronales artificiales, bosques aleatorios y máquinas de soporte vectorial pueden encontrar la aproximación de f .

3.3.1 Redes neuronales artificiales

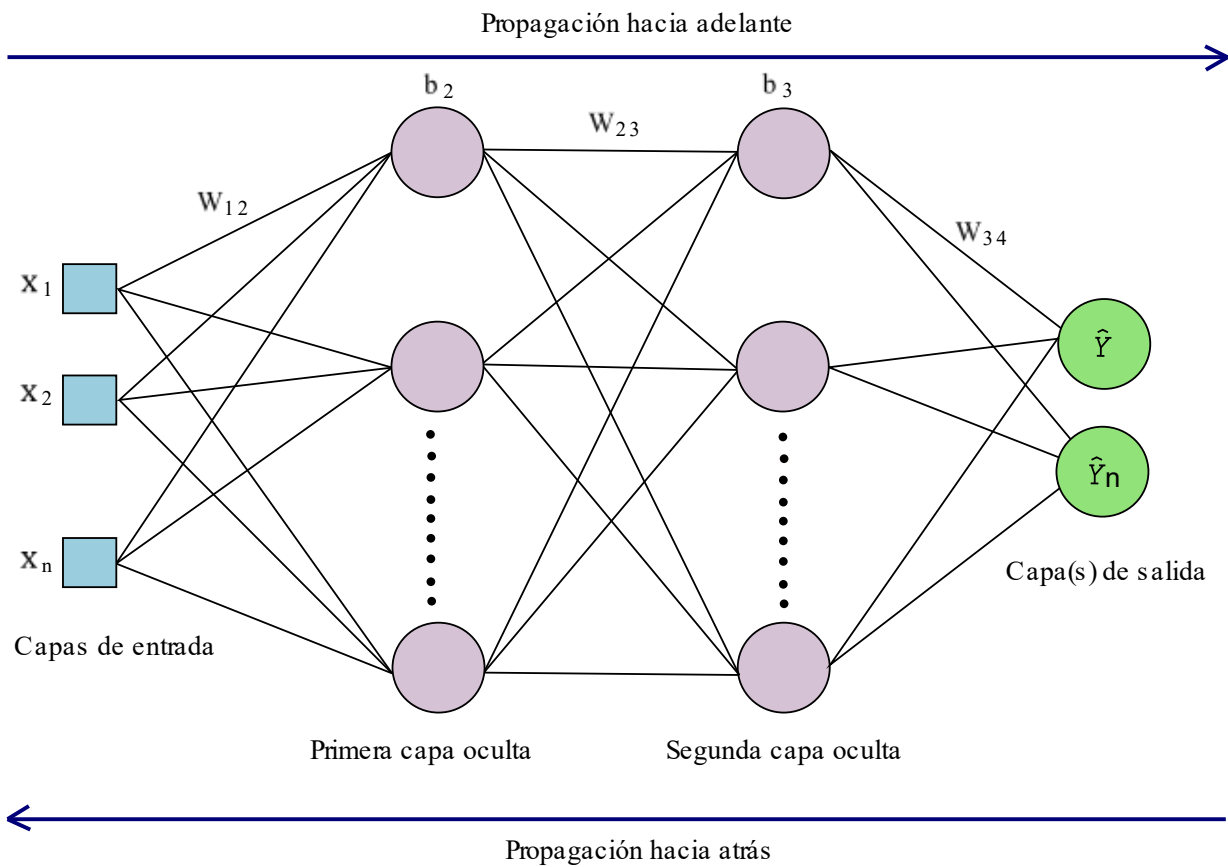
Las redes neuronales artificiales (ANN, por sus siglas en inglés) son una herramienta de propósito general flexible y no lineal que se componen de los siguientes elementos: capas de entrada, neuronas, capas ocultas, función de activación, función de pérdida, optimizador y tasa de aprendizaje. La forma matemática en que la red calcula las salidas de las capas se describe mediante la ecuación 1a.

$$y_j = f_a \left(b_j + \sum_i W_{ji} X_i \right) \tag{1a}$$

Donde y_j es el valor de salida de las neuronas en la capa j , f_a representa una función de activación, b_j un parámetro independiente conocido como bias, y W_{ji} los pesos de la red, que están representados mediante las conexiones entre las capas i y j . Es decir, que para calcular los valores en la capa j se necesitan los valores de la capa anterior. Por ejemplo, en la Figura 4 hay cuatro capas que corresponden a: una capa de entrada, dos capas ocultas y una capa de salida.

Figura 4

Estructura y elementos que conforman un modelo de las redes neuronales artificiales



Nota. Azul: capas de entrada, morado: neuronas, verde: capa(s) de salida. Los parámetros entrenables de la red W_{12} corresponden a los pesos entre la primera y segunda capa, b_2 al bias de la segunda capa, y \hat{Y} a la capa de salida que en este caso hace referencia a la probabilidad de ocurrencia mineral.

Teniendo en cuenta que una de las características de las ANN es emular la complejidad de las funciones que se quieren aproximar, f_a debe representar también una función no lineal. Algunas de las más usadas son:

Tabla 3

Funciones de activación para modelos ANN

Nombre	Función
Unidad Lineal Rectificada (ReLU)	$f(x) = \max(0, x)$
Sigmoide	$f(x) = \frac{1}{1 + e^{-x}}$
Límite estricto	$f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$
Tangente hiperbólica	$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
Gaussiana	$f(x) = e^{-\frac{(x-c)^2}{2\sigma^2}}$
Lineal	$f(x) = x$

Nota. Tomado de Keras: The Python Deep Learning API (<https://keras.io/>).

En este caso, la aproximación de f está soportada por el Teorema de Aproximación Universal que otorga las condiciones necesarias para que se pueda estimar M a partir de los parámetros entrenables óptimos θ , es decir W y b , tal que el error de la función de pérdida \mathcal{L} sea mínimo. Este problema de optimización se resuelve con el uso de algoritmos como el gradiente descendiente, Adagrad, RMSprop, Adam, Nadam, SGD y Adadelta (Gerón, 2019) que retropropagan la actualización de los parámetros entrenables en las capas ocultas de la red hasta

generar un nuevo valor \hat{Y} . En este caso el valor de \mathcal{L} en la ecuación 1b debe ser cercano a cero, lo que quiere decir que el modelo converge y, por lo tanto, la función f fue aproximada.

$$\min_{\theta} \mathcal{L}(M(x) - y) \quad (1b)$$

Entre las funciones de pérdida más usadas en problemas de probabilidad se encuentran las presentadas en la Tabla 4. En este trabajo se usó la función de entropía cruzada binaria que es usada en problemas de clasificación que abordan la naturaleza del problema de predicción de ocurrencias minerales, donde la etiqueta en los datos de entrenamiento contiene valores de 1 (depósito) y 0 (no-depósito). Adicionalmente, se escogió la función de activación sigmoide para la capa de salida debido a que representa la probabilidad de ocurrencia del evento estudiado cuyo rango es de [0,1]. Finalmente, el algoritmo de retropropagación (*backpropagation*) se realizó con el optimizador Adam, que se basa en el gradiente, por lo que resulta eficiente para resolver el problema de optimización en conjuntos de datos grandes.

Tabla 4

Funciones de pérdida usadas en problemas de probabilidad

Nombre	Función
Entropía cruzada binaria	$L(Y, \hat{Y}) = -\frac{1}{N} \sum_{i=0}^N [Y_i \log(\hat{Y}_i) + (1 - Y_i) \log(1 - \hat{Y}_i)],$
Entropía cruzada categórica	$L(Y, \hat{Y}) = -\frac{1}{N} \sum_{i=1}^N \sum_{j=0}^C Y_{i,j} \log(\hat{Y}_{i,j})$
Poisson	$L(Y, \hat{Y}) = \frac{1}{N} \sum_{i=0}^n \hat{Y}_i - Y_i \log(\hat{Y}_i)$

Nota. Tomado de Keras: The Python Deep Learning API (<https://keras.io/>)

El proceso se detiene hasta que se completan un número determinado de épocas o iteraciones. Una época significa entrenar la red neuronal con todos los datos en un ciclo. Es decir, cuando se usan todos los datos una sola vez en ambas direcciones: propagación hacia adelante y hacia atrás. El parámetro que determina el tamaño del paso en cada iteración mientras el algoritmo de optimización se mueve hacia un mínimo de la función de pérdida es denominado “tasa de aprendizaje”. Si este valor es muy pequeño se requieren mayor número de épocas porque los cambios realizados en la actualización serán muy pequeños y, si por el contrario el valor es muy grande, el modelo converge más rápido, pero puede caer en oscilaciones.

3.3.2 *Bosques aleatorios*

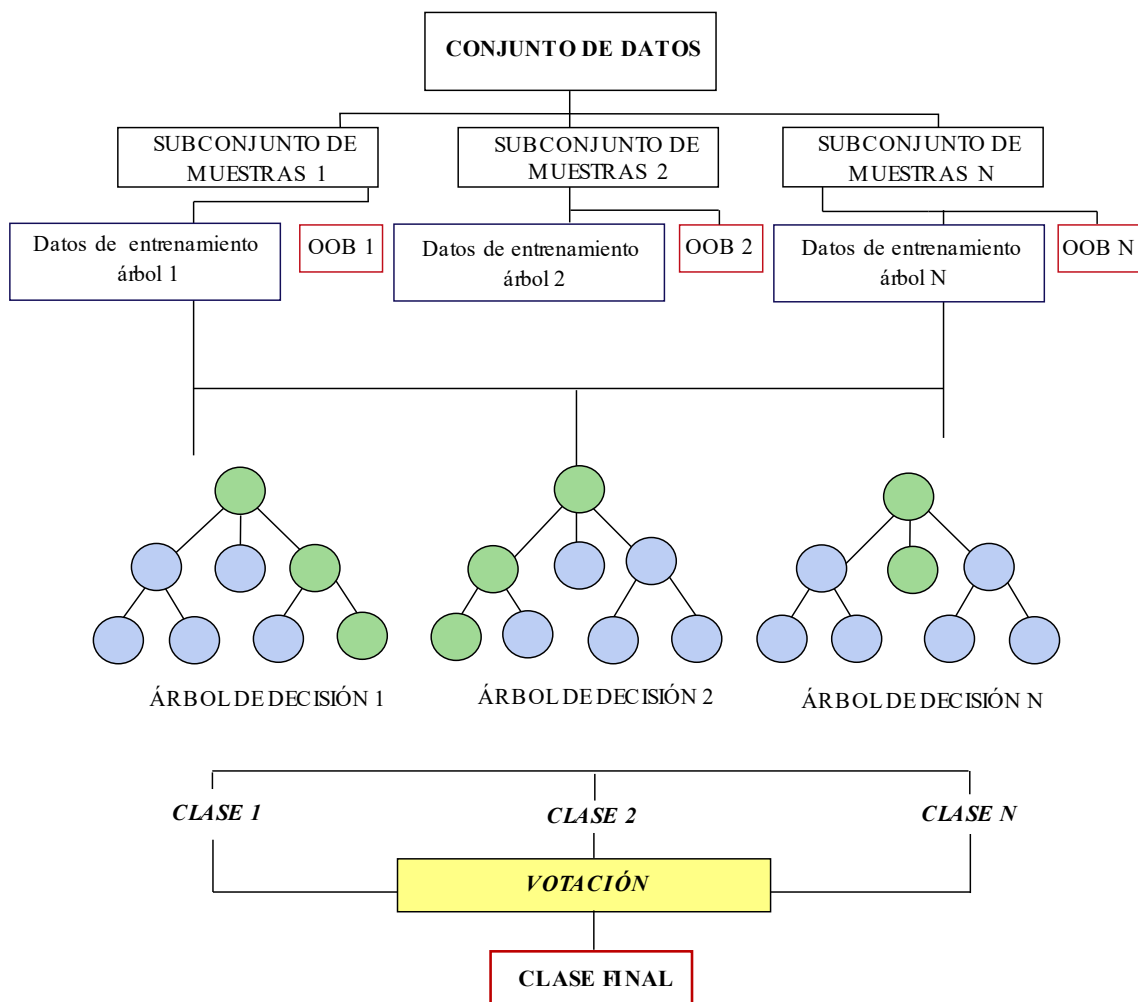
Los bosques aleatorios (*RF*, por sus siglas en inglés) fueron originalmente desarrollados por Breiman (2001) como un ensamblaje de árboles de decisión cuya predicción depende de los valores de un vector aleatorio muestreado de forma independiente. Primero, el algoritmo genera un conjunto de datos “*bootstrapped*” tomando muestras aleatoriamente del conjunto de datos original. A este proceso se le conoce como “*bagging*”. Luego, crea un número determinado de árboles de decisión escogiendo de manera aleatoria atributos y evaluándolos en cada nodo mediante el cálculo de su impureza. Existen varias aproximaciones para cuantificarla, por ejemplo, el índice Gini o criterio de entropía. El primero mide la desigualdad de cada nodo, por lo cual a mayor índice de Gini menor pureza. Está dado por:

$$I_G(f) = \sum_{i=1}^m f_i(1 - f_i) \quad (2)$$

Donde f_i es la probabilidad de la clase i en el nodo m y puede calcularse con $f_i = \frac{n_j}{n}$, en la cual n_j es el número de muestras que pertenecen a la clase j y n el número total de muestras de un nodo específico.

Figura 5

Estructura de un modelo de bosques aleatorios



Por otro lado, el criterio de entropía tiende a crear nodos balanceados en el número de observaciones y se relaciona con la ganancia de información, la cual busca la división con menor entropía ponderada. Su ecuación se presenta a continuación:

$$E = -p(A)\log(p(A)) - p(B)\log(p(B)) \quad (3)$$

Donde p es la proporción de cada clase A o B .

La elección de la medida de impurezas tiene poco efecto sobre el rendimiento de los algoritmos de aprendizaje del árbol de decisión, debido a que estas son consistentes entre sí. Raileanu & Stoffel (2004) reportan que solo en el 2% de los casos esto es importante.

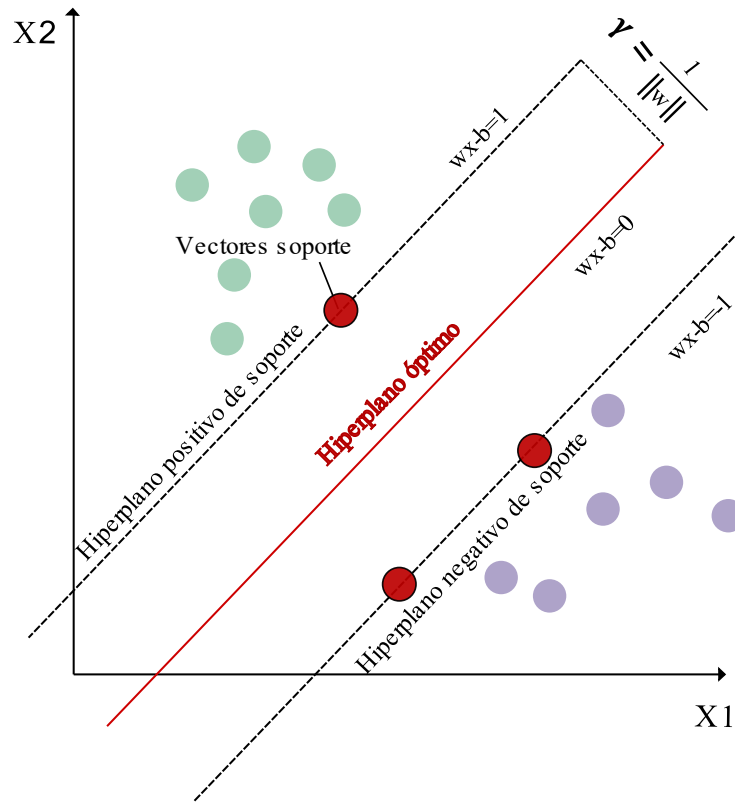
Por lo general, un tercio de los datos originales no son usados para el entrenamiento, a estos se les llama “*out of bag dataset (OOB)*”, y se usan para testear el modelo, evaluando los subconjuntos OOB en cada árbol de decisión. De esta manera se calcula la precisión de la clasificación de los bosques aleatorios. La clase final es definida mediante una votación entre todos los árboles de decisión. En la Figura 5 se muestra el proceso para generar un resultado final con base en N árboles de decisión.

3.3.3 Máquinas de soporte vectorial

Las máquinas de soporte vectorial (*SVM*, por sus siglas en inglés) tienen como objetivo encontrar un hiperplano óptimo que separe en un espacio de N -dimensiones dos clases de datos etiquetados con valores de $+1$ y -1 tal que la distancia entre ambas clases sea máxima. Por eso, además del hiperplano óptimo, existen dos hiperplanos delimitadores de clases (o hiperplanos de soporte) paralelos entre sí. Los puntos de datos clasificados cerca o sobre el margen de los hiperplanos de soporte se llaman vectores de soporte (Figura 6).

Figura 6

Estructura de un modelo de máquinas de soporte vectorial



Nota. En este caso el conjunto de datos se puede separar y clasificar mediante una línea recta. Las líneas punteadas corresponden a los hiperplanos de soporte, sobre los cuales se encuentran los vectores de soporte. La distancia entre el hiperplano óptimo y los hiperplanos de soporte se denota como γ .

Si las clases son linealmente separables entonces se satisface que:

$$wx_i + b \geq +1, \quad \text{si } y_i = +1 \tag{4}$$

$$wx_i + b \leq -1, \quad \text{si } y_i = -1 \tag{5}$$

Donde w y b representan los coeficientes de la ecuación del hiperplano y y_i las etiquetas de los datos de entrada. Cuando se requiere clasificar el conjunto de datos se resuelve un problema de optimización mediante el cual se busca que se cumplan dos condiciones. La primera hace referencia a que la distancia γ entre los hiperplanos de soporte sea la máxima, es decir el valor de w sea mínimo (6). La segunda, que los datos sean clasificados correctamente mediante la solución de la ecuación (7).

$$\gamma = \frac{1}{\|w\|} \quad (6)$$

$$y^i(w \cdot x^i + b) \geq 1, \quad i = 1, 2, \dots, n \quad (7)$$

Debido a que generalmente los datos no son completamente separables mediante una línea recta, existen técnicas que permiten dividir las dos clases objetivo. Por ejemplo, los métodos “el margen suave (*soft margin*)” o “el truco de kernel”. En el primero se permiten muestras clasificadas en el lado opuesto del límite de decisión al cual realmente pertenecen, es decir, en este caso existe un límite de violación de los hiperplanos delimitadores que establece una “penalización” denotada como parámetro C , el cual controla el número y la importancia de las violaciones que se pueden tolerar durante el entrenamiento.

Si $C = \infty$ ninguna violación es permitida y por lo tanto el resultado es el equivalente a usar un hiperplano óptimo lineal. En cambio, si $C = 0$ las observaciones pueden estar del lado incorrecto del límite de decisión o hiperplano. Cuanto mayor es el valor de C más estrecho es el margen que separa los hiperplanos, y por lo tanto menor número de observaciones serán vectores soporte. En efecto, el hiperparámetro C permitirá controlar el error que genera el modelo, por lo tanto, ajustarlo correctamente es de suma importancia. Para esto se usan técnicas de validación

cruzada que permiten estimar el error para cada valor de C y escoger aquel cuyo resultado sea menor.

El segundo método mencionado anteriormente como “truco de kernel” se basa en mapear los datos en una dimensión superior conocida como “espacio de características” y agregar valores adicionales llamados “*slack variables*” para acomodar valores atípicos. El procedimiento se realiza bajo la premisa de que puede existir un límite lineal en ese nuevo espacio de mayor dimensión (Cardoso-Fernandes et al., 2020). De acuerdo con Rodríguez-Galiano et al., (2015) algunas de las funciones más conocidas en aplicaciones del *SVM* son:

Tabla 5

Funciones de Kernel

Nombre	Función
K_{lineal}	$y = x, x'$
$K_{polinomial}$	$y = (\gamma xx' + r)^p$
K_{RBF}	$y = exp(-\gamma \ x - x'\ ^2)$
$K_{sigmoide}$	$y = \tanh(\gamma xx' + r)$

Nota. Funciones de Kernel usadas en el método del “truco de kernel” para asignar clases a un conjunto de datos durante el entrenamiento de *SVM*.

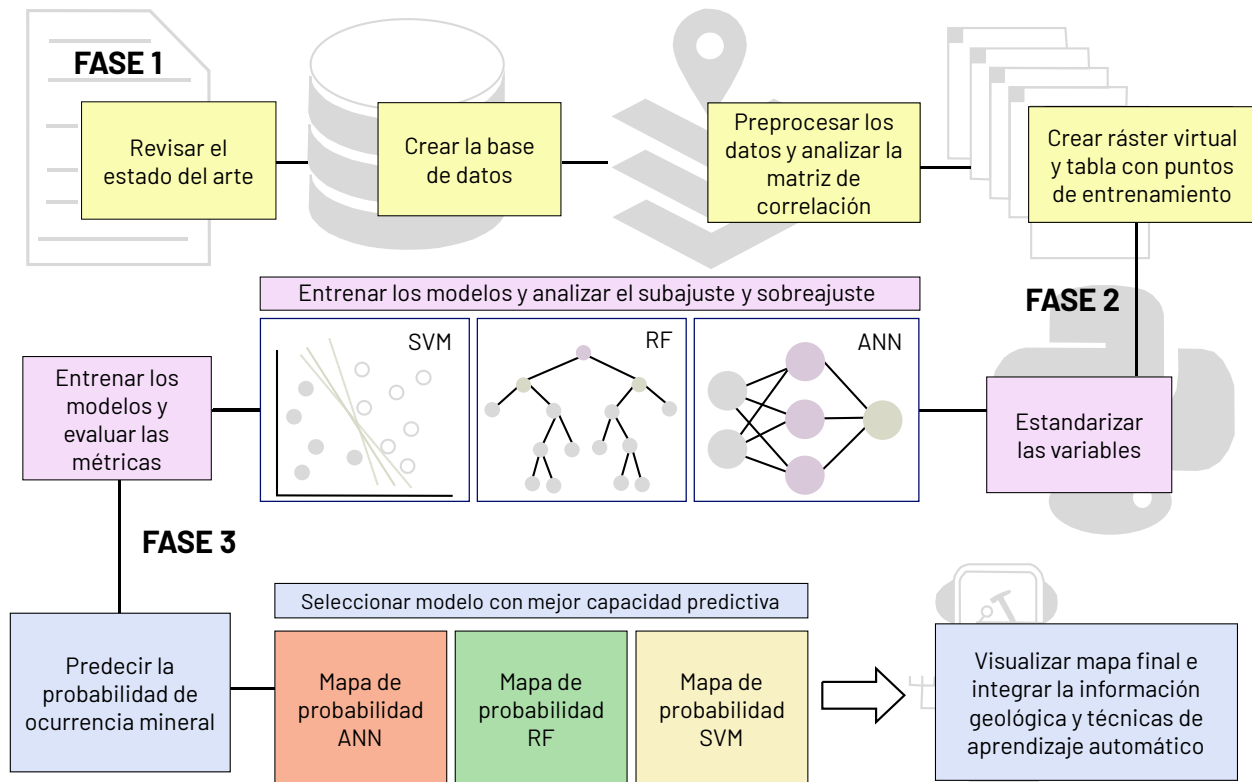
La función puede afectar el error total del modelo. Para ejemplificar, en Zuo & Carranza (2011) se realizan varios experimentos abordando la temática, el porcentaje de error varía considerablemente al escoger diferentes funciones Kernel.

4 Metodología

Para llevar a cabo el desarrollo de los objetivos específicos planteados se aplicó la metodología descrita en la Figura 7.

Figura 7

Metodología establecida para el proyecto actual



4.1 FASE 1: Análisis de los insumos geológicos, geofísicos y geoquímicos relacionados con la ocurrencia de depósitos minerales tipo pórfido.

ACTIVIDAD 1.1. Revisar el estado de arte sobre la formación de depósitos minerales tipo pórfido, enfatizando en los trabajos llevados a cabo en la zona de estudio (Yukón, Canadá). Con base en esto se escogieron los insumos para la zona de interés.

ACTIVIDAD 1.2. Crear una base de datos a partir de la información disponible en la página web del Servicio Geológico de Canadá (<https://data.geology.gov.yk.ca/>) y el Gobierno de Canadá (<https://www.canada.ca/en.html>) que contiene información tipo vector y ráster para las capas analizadas en la ACTIVIDAD 1.1.

ACTIVIDAD 1.3. Preprocesar los datos obtenidos en la ACTIVIDAD 1.2 en un sistema de información geográfica (SIG) haciendo uso del software libre y de código abierto QGIS. Los insumos tuvieron un preprocesamiento estándar que garantizó que todas las capas tuvieran una distribución normal e igual sistema de coordenadas geográficas, posición de valores nulos, número de filas y columnas y resolución espacial.

ACTIVIDAD 1.4. Analizar la relación entre las variables de entrada a partir de métodos estadísticos como correlación, dispersión, distribución, varianza, entre otros.

ACTIVIDAD 1.5. Crear Ráster Virtual y tabla con puntos de entrenamiento. El primero contiene las capas de entrada del modelo y, el segundo, los valores de píxel de cada atributo donde hay depósitos minerales tipo pórfido.

4.2 FASE 2: Diseño modelos de aprendizaje automático usando redes neuronales artificiales, bosques aleatorios y máquinas de soporte vectorial en el lenguaje de programación Python.

ACTIVIDAD 2.1. Cargar y estandarizar los datos obtenidos en la FASE I con Colab PRO, un entorno de Jupyter Notebook que permite ejecutar códigos en el lenguaje de programación Python. El proceso anterior es requerido porque el rango de valores de los datos es muy variado debido a que provienen de diferentes fuentes de información, lo cual puede generar problemas durante la etapa de entrenamiento de aprendizaje automático.

ACTIVIDAD 2.2. Crear los modelos de redes neuronales artificiales, máquinas de soporte vectorial y bosques aleatorios usando la librería TensorFlow. Durante esta actividad se ajustaron los hiperparámetros, la función de pérdida y las métricas para cada uno.

ACTIVIDAD 2.3. Entrenar los modelos usando Colab PRO en un entorno de ejecución de 12 GB RAM y GPU T4 para cada uno de los métodos de aprendizaje automático usando la información generada en la ACTIVIDAD 2.1. El acceso a este servicio se garantiza mediante el pago por parte de los autores. En esta etapa se generaron los análisis de sobreajuste y/o subajuste de cada modelo, variando los parámetros mediante validación cruzada. Al finalizar esta actividad se escogió el modelo con mejor capacidad de predicción.

4.3 FASE 3: Implementación de una metodología utilizando los modelos de aprendizaje automático y la información geológica y geofísica para predecir la probabilidad de ocurrencia de depósitos minerales tipo pórfido.

ACTIVIDAD 3.1. Predecir la probabilidad de ocurrencia de depósitos minerales tipo pórfido con el modelo seleccionado en la ACTIVIDAD 2.3 y los valores del ráster generado en la ACTIVIDAD 1.5. Durante esta etapa se generó el mapa de predicción de probabilidad de ocurrencia de depósitos minerales tipo pórfido.

ACTIVIDAD 3.2. Visualizar en QGIS el mapa generado en la ACTIVIDAD 3.1 haciendo uso de modelos digitales de elevación y sensores remotos para mejorar el análisis e interpretación de la ocurrencia de depósitos minerales tipo pórfido con factores como la topografía, fallas, unidades geológicas, depósitos minerales conocidos, y otros que apliquen.

ACTIVIDAD 3.3. Integrar una metodología que incluya desde el acondicionamiento de datos hasta la construcción de modelos de aprendizaje automático. De tal manera, que se establezca un flujo de trabajo coherente y sistemático en la unificación de información y algoritmos computacionales. Durante esta actividad se hizo la comparación y análisis entre la información

geológica del área de estudio con el resultado de la ACTIVIDAD 3.1 donde se resaltan las zonas de mayor potencial mineral.

5 Resultados

5.1 Análisis de los insumos geológicos, geofísicos y geoquímicos relacionados con la ocurrencia de depósitos minerales tipo pórfido

En este capítulo se realizó una revisión del estado de arte en la predicción de depósitos tipo pórfido usando técnicas de aprendizaje automático. De esta forma se creó una base de datos con insumos geológicos, geofísicos y geoquímicos que fueron preprocesados y analizados.

5.1.1 Revisión del estado de arte en la predicción de depósitos minerales usando aprendizaje automático

Los mapas de prospectividad mineral (MPM) tienen como objetivo predecir ubicaciones para las cuales la probabilidad de ocurrencia de un evento objetivo, por ejemplo, un pórfido, es máxima dadas algunas variables predictoras favorables (Kost et al., 2021). Por lo cual, es necesario combinar criterios relevantes que contribuyan a la formación del depósito mineral estudiado. Por ello, el aprendizaje automático es una herramienta fundamental para identificar zonas de exploración mineral debido a sus ventajas en comparación con los métodos tradicionales que, por lo general, conllevan un costo financiero y consumo de tiempo altos.

Entre las principales ventajas que caracterizan los algoritmos computacionales se encuentra su capacidad para no requerir suposiciones con respecto a la distribución de datos y establecer de forma adecuada las relaciones no lineales entre los depósitos minerales y las características de

entrada (Zuo, 2017). Por tal razón, actualmente varios métodos de MPM se basan en la aplicación de modelos de aprendizaje automático para predecir la ocurrencia de depósitos minerales.

Los algoritmos destacados son redes neuronales artificiales, bosques aleatorios y máquinas de soporte vectorial (Carranza & Laborte, 2015a; Chen & Wu, 2017; Ghezelbash et al., 2019; Harris et al., 2015; Kost et al., 2021; Li et al., 2020a; Maepa & Smith, 2017, 2019; M. McMillan et al., 2021; V. Rodriguez-Galiano et al., 2015a; Sun et al., 2019a, 2020; Zhang et al., 2018; Zuo, 2017). Además, recientemente se han desarrollado modelados geológicos tridimensionales con base en el análisis espacial de los datos, lo cual ha dado lugar a una mejor comprensión de los sistemas minerales (Mao et al., 2020; Xiang et al., 2020).

Algunos casos de estudio exitosos donde se usaron técnicas de aprendizaje automático para generar MPM son los depósitos de oro en España (Rodriguez-Galiano et al., 2014) Australia (Hariharan et al., 2017), China (Zhang et al., 2018), pórfidos en Irán (Keykhay-Hosseinpour et al., 2020), skarn de Cu (Sun et al., 2019), entre otros. Con base en esta revisión se establece que redes neuronales artificiales, bosques aleatorios y máquinas de soporte vectorial, por excelencia pueden compararse en tareas de clasificación.

Entre las ventajas que poseen los algoritmos de *ML* se encuentran su capacidad para abordar datos de diferente naturaleza. Por ejemplo, en Carranza & Laborte, (2015b) se demuestra que los bosques aleatorios generan buenos resultados cuando procesan conjuntos de datos con valores nulos y número de prospectos minerales reducido (<20). En Chen & Wu (2017) se modela una regresión logística extrema (ELM) para minimizar el tiempo de entrenamiento de los algoritmos, lo que en muchas ocasiones puede limitar su capacidad. En Daviran et al., (2021) y Lin et al., (2021) se desarrolla un modelo híbrido que, junto con *SVM*, *ANN* y *RF*, ajusta

automáticamente el valor de los hiperparámetros e incluso, aumenta la precisión predictiva del modelo. Es decir, en la literatura se han presentado varios casos de estudio relacionados con soluciones a problemas que puedan generarse con estos modelos lo cual representa una ventaja adicional para aplicar estas técnicas en los flujos de trabajo de la industria minera. Por otro lado, algunos aspectos que deben considerarse cuando se hace modelado de prospectividad mineral basado en GIS usando *ML* según Carranza (2009) son la selección adecuada de un tamaño de celda unitaria (o píxel), ubicaciones del depósito de interés y validación cruzada para evaluar los modelos.

A) Selección adecuada de un tamaño de celda unitaria (o píxel): Teniendo en cuenta que los mapas predictores están en formato ráster es importante resaltar la resolución de estos con base en la escala de trabajo y la disponibilidad de los datos geocientíficos. Por eso, se hace alusión a la metodología propuesta por Carranza (2009) para seleccionar un tamaño de píxel adecuado, el cual representa la precisión de los mapas generados como capas de entrada para los algoritmos computacionales, pues un tamaño de píxel fino representará una densidad de muestreo alto y por, lo tanto, una precisión superior (Carranza, 2009). A su vez, la selección del tamaño de celda debe estar fundamentado en que cada depósito mineral es único y cada celda unitaria debe contener solo una de las ocurrencias minerales conocidas (Carranza, 2009), pues las mineralizaciones son un fenómeno geológico raro.

B) Selección de depósitos de interés: la importancia de este factor radica en la selección adecuada de los puntos no-depósito y el desbalance de datos. En Prado et al., (2020) se plantea que el número de ocurrencias de depósitos minerales conocidos es naturalmente mucho menor que el número de ubicaciones no mineralizadas. No obstante, la respuesta de los algoritmos a estas situaciones ha

sido ampliamente estudiado obteniendo muy buenos resultados (Chawla et al.2004; Hariharan et al.2017, como se cita en Li et al., 2020b). Por otro lado, con respecto a la selección adecuada de los puntos no-depósito se han planteado soluciones alternativas. Por ejemplo, Nykänen et al., (2015) proponen usar puntos de otro tipo de ocurrencias minerales o muestras aleatorias como no-depósitos. Carranza (2009) apoya esta idea afirmando que, de esta manera, las firmas de los datos espaciales multivariantes probablemente no sean coherentes. Para garantizar esta distribución Li et al., (2020b) sugiere el análisis de patrón de puntos.

Los algoritmos de aprendizaje automático pueden procesar conjuntos de datos no lineales, con valores nulos, con desbalance de datos, con número limitado de datos de entrenamiento, entre otras situaciones que generalmente se presentan en la aplicación de los métodos. Para complementar, a continuación, se presentan las principales capas de entrada usadas en la literatura para generar MPM de sistemas minerales tipo pórfido alrededor del mundo (Tabla 6). Teniendo en cuenta estos criterios y la información geológica de la sección 3.2.2, se seleccionaron las capas de entrada para la predicción (Tabla 7).

Tabla 6

Capas de entrada usadas en el estado de arte para generar MPM de depósitos tipo pórfido usando técnicas de aprendizaje automático

Algoritmos	País	Depósito	Capas de entrada	Referencia
RF y RBM	Distrito Dehsalm (Irán)	Pórfidos de Cu-Au	<ul style="list-style-type: none"> - Proximidad a rocas intrusivas y a fallas geológicas - Mapas geofísicos (magnetometría y aeromagnetometría) - Mapa de alteraciones hidrotermales de argílica y óxidos de Fe - Geoquímica de sedimentos de corriente 	(Keykhay-Hosseinpoor et al., 2020b)
FL	Columbia Británica (Canadá)	Pórfidos de Cu	<ul style="list-style-type: none"> - Mapas geofísicos (primera derivada del campo magnético, primera derivada de Bouguer, polarización inducida) - Intersecciones con fallas de alto ángulo - Anomalías geoquímicas asociadas a sedimentos de corriente 	(Montsion et al., 2019)
FL y AHP	China	Pórfidos	<ul style="list-style-type: none"> - Buffer a rocas intrusivas (500 m), zonas alteración (3000 m), diques (2500 m) - Distancia, densidad y dirección de fallas - Mapas geofísicos (magnetometría y gravimetría) - Geoquímica de elementos: Au, Ag, Cu, Mo, Pb, Zn 	(Zhang et al., 2017)
EBF, superposición de índices y FL	Seridune (Iran)	Pórfidos de Cu	<ul style="list-style-type: none"> - Buffer a fallas (<20 m, 20–35 m, 35–45 m, 45–50 m, y >50 m) - Mapa de alteración hidrotermal, tipos de roca, factor sulfuro - Geoquímica de concentración de metales (Cu, Pb y Mo) - Mapas geofísicos (magnetometría, resistividad aparente, IP “cargabilidad”, señal analítica, pseudo gravedad) 	(Abedi et al., 2017)
RF	Columbia Británica (Canadá)	Pórfido epitermal de Au-Cu	<ul style="list-style-type: none"> - Primera derivada vertical magnética - Alteración argílica - Alteración por óxidos de Fe - Alteración fílica, propilítica y sílica - Distancia a intrusivos, fallas y roca caja en intervalos de 500 m - Geoquímica de Au 	(Lachaud et al., 2021a)
RF y WoE	Abra (Filipinas)	Pórfidos de Cu	<ul style="list-style-type: none"> - Distancia a márgenes batolíticos de granodiorita, diorita - Distancia a centroides de los plutones pórfido - Distancia a fallas/fracturas de tendencia NS, SE, NW, EW - Intersecciones con estructuras regionales - Geoquímica de <u>Cu</u> y Zn 	(E. J. M. Carranza & Laborte, 2015c)

Nota. RF: Bosques aleatorios, RBM: Máquina de Boltzmann Restringida, FZ: Lógica difusa, AHP: Proceso de jerarquía analítica difuso, EBF: Funciones de creencias evidenciales, WoF: Pesos de evidencia

Tabla 7

Capas de entrada para los algoritmos ANN, RF y SVM

Proxi	Factor condicionante
Geología regional	- Proximidad a cuerpos intrusivos
Fuente	- Proximidad a rocas volcánicas
Geología estructural	-Proximidad a las fallas geológicas
Transporte	-Proximidad a contactos de terrenos litotectónicos
Geofísica	-Campo total magnético residual
Trampa física	-Primera derivada vertical del campo total magnético residual
Geoquímica	-Mapa de concentración geoquímica para Au, Cu, Mo,
Precipitación geoquímica	Pb y Zn

5.1.2 Creación de la base de datos

La creación de la base de datos consistió en descargar la información disponible en la página web del Servicio Geológico de Canadá (<https://data.geology.gov.yk.ca/>) y el Gobierno de Canadá (<https://www.canada.ca/en.html>). A continuación, se describe la información sobre cada base de datos.

5.1.2.1 Geología

El conjunto de datos incluye capas acerca de la geología del lecho rocoso, fallas locales, contactos y geocronología [actualizada en febrero de 2022]. También se encuentra disponible información sobre fallas regionales [actualizada el 8 de septiembre del 202] y pliegues [actualizada en septiembre de 2021]. La geología del lecho rocoso incluye polígonos de las unidades geológicas

mapeadas a escala 1:250.000 con breves descripciones del tipo y subclase de roca. Las fallas geológicas contienen polilíneas con el nombre, escala y grado de confianza de cada una de las fallas regionales y locales presentes en Yukón.

5.1.2.2 Geoquímica

La base de datos de geoquímica es una compilación de más de 30000 puntos de análisis geoquímicos en sedimentos de corriente tomados en el territorio de Yukón. Se basa principalmente en datos publicados y no publicados adquiridos por el Servicio Geológico de Yukón y el Servicio Geológico de Canadá como parte de campañas de mapeo regionales. La versión inicial fue publicada el 19 de noviembre de 2020 y desde entonces no ha sido actualizada. Los análisis se realizaron en varios laboratorios utilizando una gama de métodos de alta precisión resumidos para los elementos de interés en la Tabla 8.

Tabla 8

Métodos usados y límites de detección de los elementos geoquímicos de interés

Método	Límite de detección				
	Zn	Au	Cu	Mo	Pb
AAS	2 ppm	1-2 ppb	2 ppm	-	2 ppm
INA	-	2 ppb	-	-	-
ICP-MS	0.1 ppm	0.2 ppb	0.01 ppm	0.01 ppm	0.01 ppm

Nota. Límites de detección con base en la base de datos geoquímicos de Yukon Geological Survey (2022a, b).

5.1.2.3 Geofísica

Los mapas geofísicos 1:250.000 son compilaciones de varios levantamientos geofísicos en el territorio de Yukón. El área de estudio se compone de 9 planchas de vuelo. En específico se tuvieron en cuenta los mapas de campo total magnético residual y primera derivada del campo total magnético residual, que tienen tres bandas y tamaño de píxel de 147.467 m.

5.1.2.4 Ocurrencias minerales

La base de datos de ocurrencias minerales contiene puntos con información geológica, económica y geográfica sobre más de 2500 minas metálicas, minerales industriales y de carbón, depósitos y ocurrencias en Yukón. Este conjunto de datos GIS muestra la ubicación del punto, el nombre, el tipo de depósito, el estado del depósito y las materias primas. La última fecha de actualización fue del 15 julio de 2019.

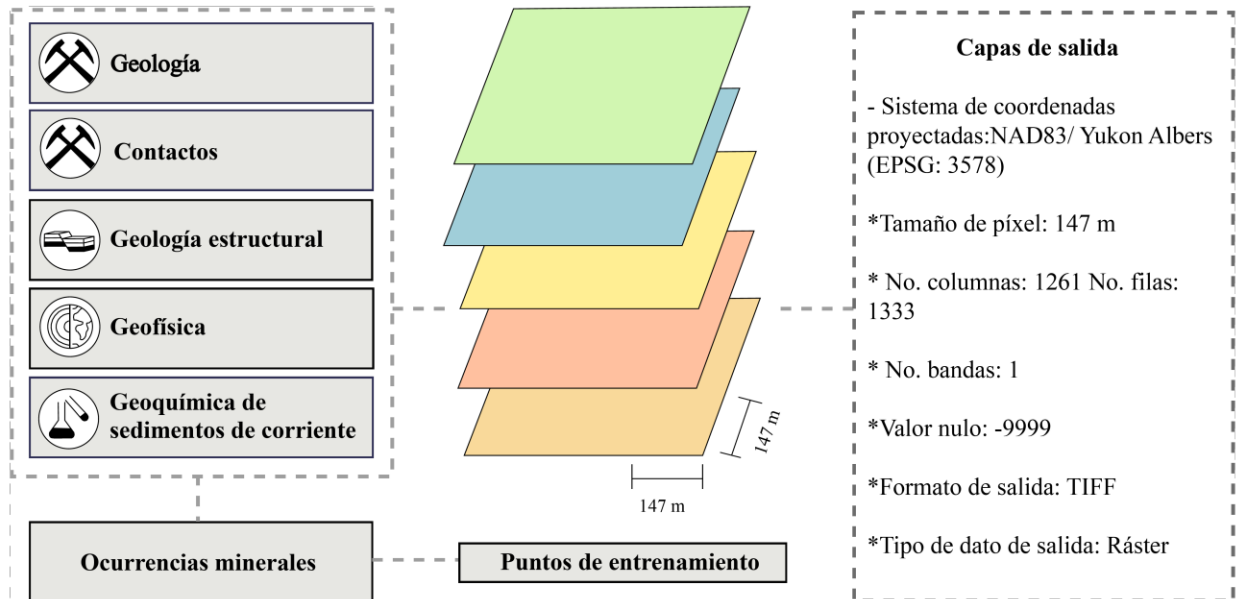
En la siguiente sección se mostrará el procesamiento realizado a cada una de las capas de entrada incluyendo las ocurrencias minerales que servirán para generar la tabla con puntos de entrenamiento.

5.1.3 *Preprocesamiento de la base de datos*

El procesamiento de las capas de entrada consistió en ajustar los datos con el mismo sistema de coordenadas, tamaño celda, número de filas y columnas, extensión, número de bandas, valores nulos y formato de salida (Figura 8). El primer paso consistió en realizar un recorte en QGIS a todas las capas con base en el polígono del área de estudio.

Figura 8

Características de las capas de salida con el procesamiento de los datos



5.1.3.1 Geología

Rocas intrusivas y volcánicas: Se tuvieron en cuenta las rocas volcánicas y plutónicas definidas a escala 1: 250.000 con composición félsica e intermedia. Se calculó la proximidad a estos polígonos con la herramienta de distancia euclidiana.

Fallas: Las fallas que se tuvieron en cuenta para generar el mapa de proximidad fueron aquellas trazadas en escala 1:250.000 con un grado de confianza definido, aproximado e inferido. Se usaron las herramientas de distancia y buffer para generar el mapa final.

5.1.3.2 Geofísica

Las capas de campo total magnético residual y primera derivada vertical tuvieron el mismo preprocesamiento estándar que consistió en realizar un promedio aritmético de bandas con la calculadora ráster para obtener una sola banda en cada capa.

5.1.3.3 Geoquímica

Los valores geoquímicos de Au, Cu, Pb, Mo y Zn fueron generados con base en los datos medidos con el método ICP-MS (Tabla 8). El procedimiento para el procesamiento de estos datos consistió en: primero, eliminar los datos que estaban por debajo del límite de detección de la técnica ICP-MS, pues representaban anomalías producto de error instrumental o humano que podrían sesgar los modelos de aprendizaje automático. Segundo, realizar un análisis estadístico mediante la revisión de los histogramas de cada elemento (Figura 9). A través de estas gráficas se deduce que los datos no tienen una distribución normal porque su distribución no se asemeja a una campana de Gauss. Por consiguiente, se requirió aplicar una transformación logarítmica para que la varianza de los datos tuviera una tendencia a cero. Luego se aplicó la técnica de interpolación Kriging Ordinario con un modelo de semivariograma gaussiano para generar los mapas geoquímicos predictores.

5.1.3.4 Ocurrencias minerales

Las 402 ocurrencias minerales que contiene la capa en la zona de estudio pertenecen a depósitos de tipo skarn, SEDEX, VMS, pórfidos, carbón, paleoplacer, oro orogénico, venas, epitermales, ultramáficos, entre otros. El primer paso del preprocesamiento consistió en hacer una selección por atributos para tener en cuenta únicamente el tipo de depósito de interés que corresponde a los pórfidos de Cu-Mo-Au, Cu-Au y Mo.

También, se filtraron aquellos cuyo estado correspondía a depósito, prospecto perforado, productor pasado, productor y prospecto. No se tuvieron en cuenta las anomalías, pues su grado de confianza para hospedar depósitos minerales no es alto y, por lo tanto, pueden generar sesgo en el modelo computacional. De este modo el total de depósitos minerales tipo pórfido fue de 74.

Por otro lado, para seleccionar el conjunto de datos no-depósitos se tuvo en cuenta la zona de influencia propuesta por Mihalasky et al., (2011) y expuesta en la sección Depósitos pórfidos de Yukón donde se exponen 10 km alrededor de cada cuerpo plutónico donde hay probabilidad de encontrar depósitos tipo pórfido. De este modo, se seleccionaron puntos que, en primer lugar, no se encontraban dentro de la zona de influencia y, en segundo lugar, estaban en zonas que no cumplían con los parámetros geológicos para hospedar depósitos de este tipo.

Si bien, algunos autores recomiendan seleccionar los puntos no-depósito de forma aleatoria (Carranza, 2009), también sugieren que estas localidades deben estar ubicadas en puntos distales a los puntos de depósitos. Esto garantiza que las características espaciales y geológicas de ambos conjuntos de datos sean distintas y, por lo tanto, el algoritmo tenga mayor facilidad para clasificar y dividir las dos clases. De esta manera la precisión será mayor y, por lo tanto, el mapa de probabilidad será también más confiable.

Sin embargo, otros autores han revisado de forma exhaustiva el efecto de escoger los puntos no-depósito con diferentes técnicas. Por ejemplo, en Lachaud et al., (2021a) se determinó que la selección de estos puntos con base en los depósitos no-prospectivos tiene mejor rendimiento que si se escogen aleatoriamente. Por eso, en este trabajo de investigación se seleccionaron los puntos no-depósito con base en criterios geológicos relacionados con el modelo conceptual de exploración a escala regional y los atributos específicos de la zona de estudio como, por ejemplo, la roca caja, las estructuras geológicas, y la lejanía a los puntos ya conocidos.

En este orden de ideas, se seleccionaron igual número de depósitos y no-depósitos, lo cual en total suma 148 datos de entrenamiento. A los depósitos clasificados como pórfidos se les asignó una etiqueta de 1 y a los no-depósitos de 0. Los mapas de entrada que representan el insumo para

los algoritmos de *ANN*, *RF* y *SVM* se muestran en la Figura 10. Los mapas en alta resolución se encuentran en el Apéndice A. Las variables fueron nombradas como se muestra en la Tabla 9 para facilitar su reconocimiento en los modelos de aprendizaje automático.

Tabla 9

Nombre de las variables de entrada en los algoritmos computacionales

Variable	Tipo de variable	Descripción
X1	Continua	Proximidad a las rocas plutónicas
X2	Continua	Proximidad a las rocas volcánicas
X3	Continua	Proximidad a las fallas geológicas
X4	Continua	Proximidad a contactos de terrenos litotectónicos
X5	Continua	Valores del campo total magnético residual
X6	Continua	Valores de la primera derivada del campo magnético residual
X7	Continua	Concentración geoquímica de Zn
X8	Continua	Concentración geoquímica de Au
X9	Continua	Concentración geoquímica de Cu
X10	Continua	Concentración geoquímica de Mo
X11	Continua	Concentración geoquímica de Pb
Ocurrencias minerales	Binaria	Ocurrencias minerales tipo pórfido y puntos no-depósitos.

Figura 9

Histogramas

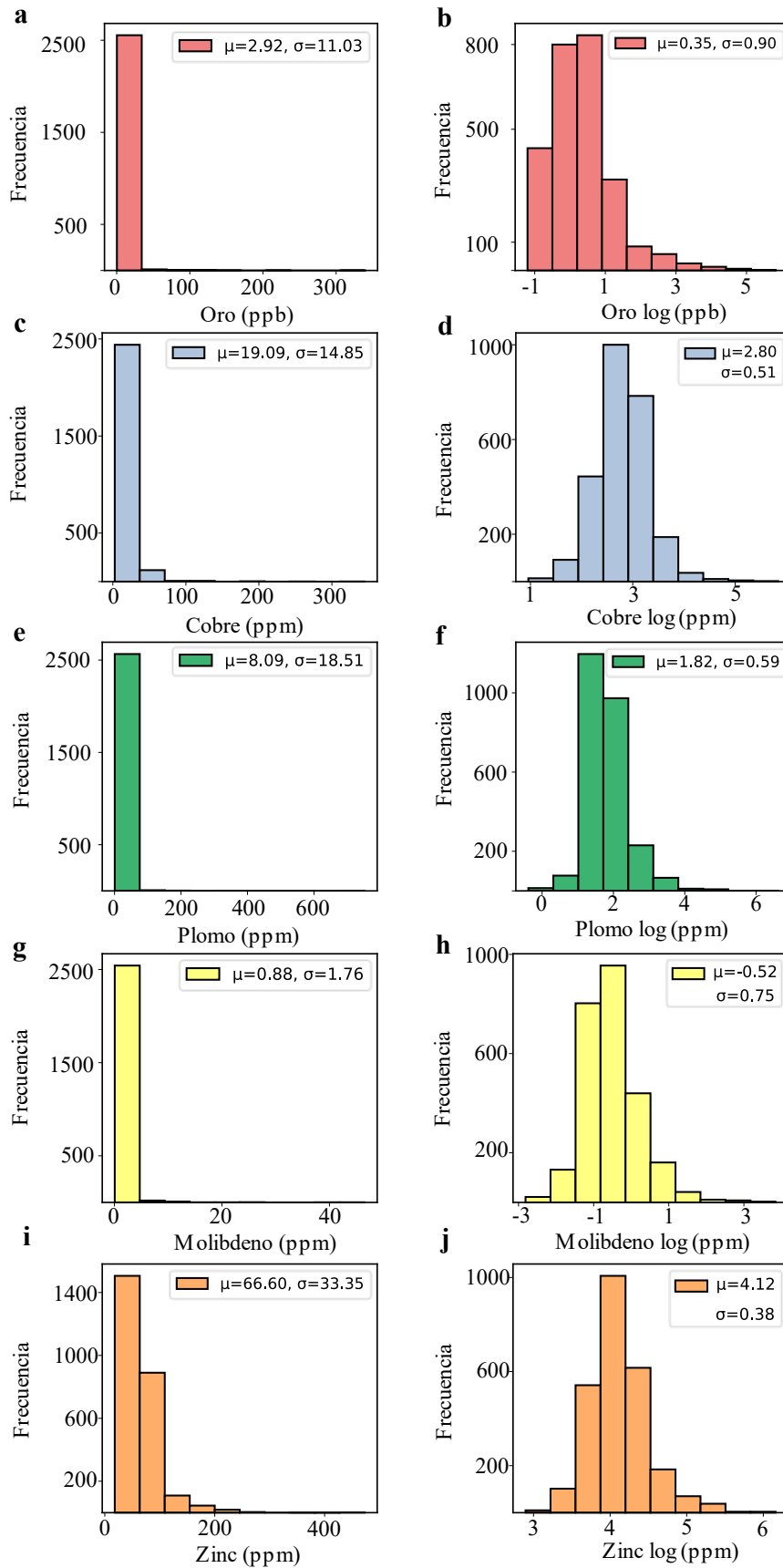
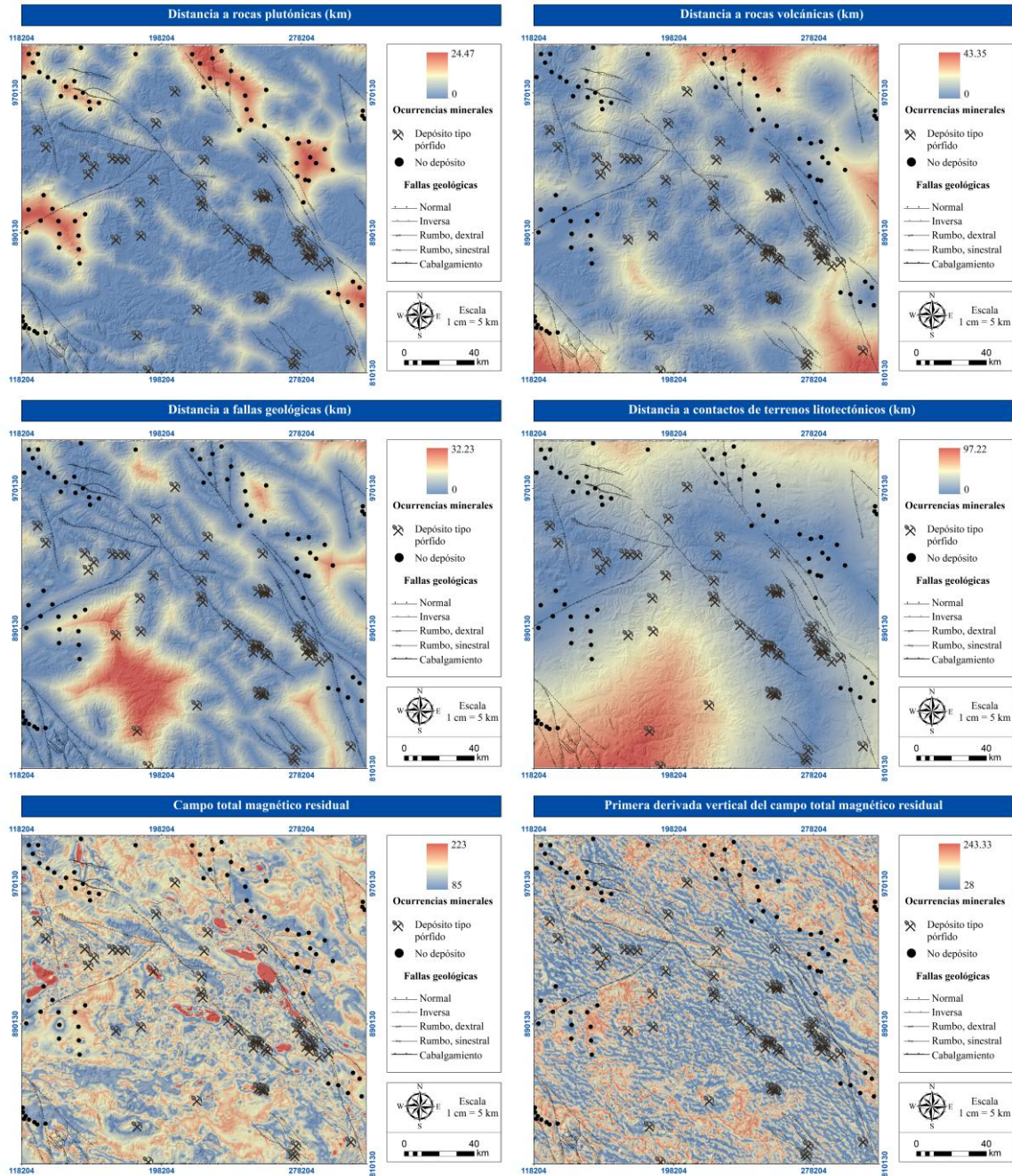
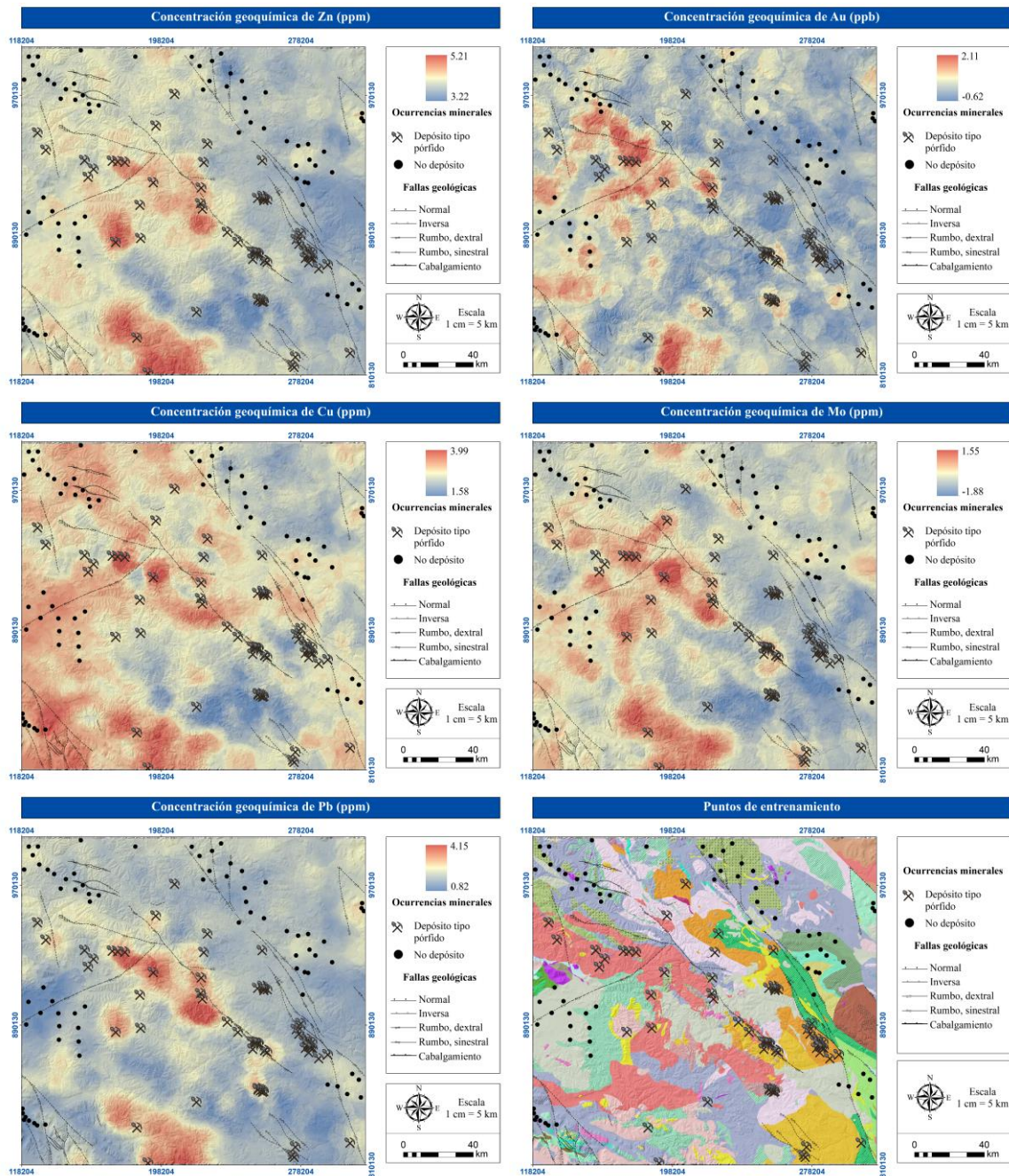


Figura 10

Mapas predictores de depósitos tipo pórfido en Yukón





5.1.4 Análisis entre las variables de entrada

Esta sección consistió en estandarizar las variables, analizar la dispersión y correlación, extraer los puntos de entrenamiento y crear ráster virtual. El análisis entre las variables de entrada garantiza que los factores que pueden afectar el rendimiento de los algoritmos computacionales

sean considerados y corregidos. En este capítulo se analizaron algunas características estadísticas del conjunto de datos. Primero, se extrajeron los valores de los puntos de depósito y no-depósito en los mapas predictores.

Con base en esto se estandarizaron las variables con el uso de la librería Scikit-learn (<https://scikit-learn.org/stable/>) en Python. El procedimiento consistió en escalar los valores para que tengan las propiedades de una distribución normal estándar con $\mu = 0$ y $\sigma = 1$ donde μ es el promedio y σ la desviación estándar (Burkov, 2019). De esta manera se analizó la distribución mediante gráficos de dispersión. En la Figura 11 se evidencia que la variable X1, que corresponde a la distancia a las rocas plutónicas, contiene muchos valores iguales a cero. Es decir, los datos no tienen dispersión y tampoco siguen una distribución normal porque están concentrados en un rango de valores específicos. Además, se destaca la presencia de valores atípicos como el punto cercano a 30000 cerca al eje y, así como algunos otros mayores a cero.

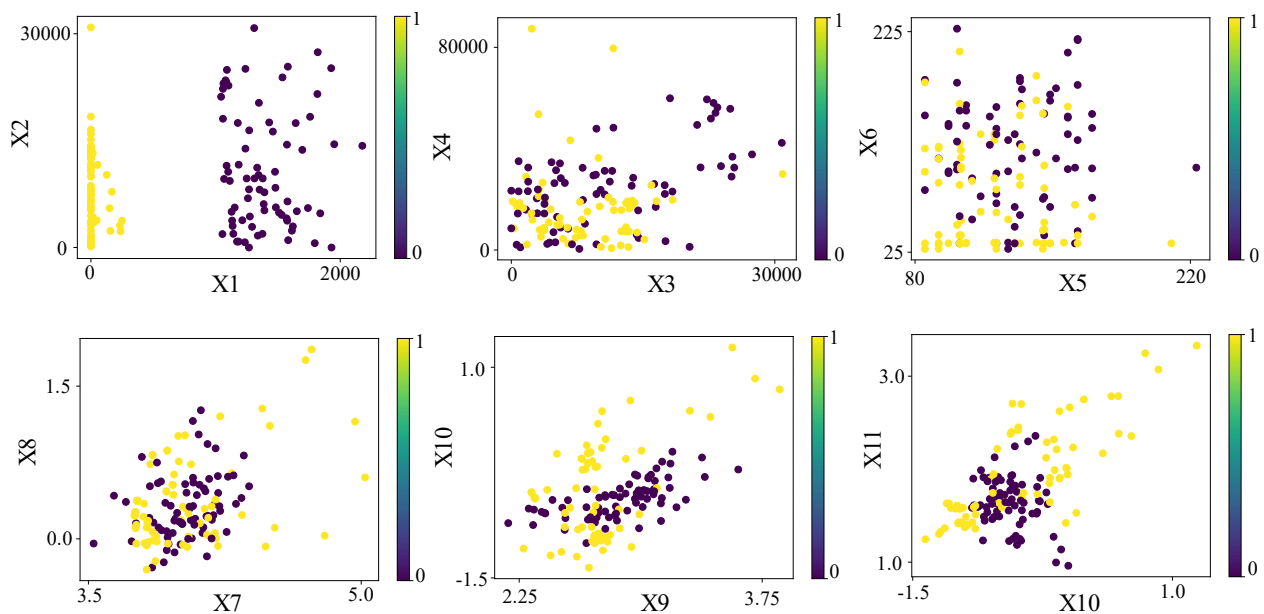
La estadística descriptiva de este conjunto de datos tiene un promedio de 7048.06, desviación estándar de 7107.50, mediana de 6471.19 y varianza de 50516675.09. Este resultado indica que la variable no posee una distribución cercana a la normal estándar donde la media es cero y la varianza es uno. Por lo tanto, se pone a consideración su uso en el entrenamiento de los modelos de aprendizaje automático, porque se demostró que su naturaleza estadística no favorece la predicción de depósitos y no-depósitos. Dado que los datos no son simétricos respecto a su media hay una alta probabilidad de que la mayoría de los valores que sean predichos pertenezcan a la clase 1, lo cual generará sesgo en el modelo.

Por otro lado, en la Figura 11 algunas de las variables sugieren una dependencia lineal que dificultará separar los datos mediante una superficie de decisión. De esta manera, el segundo factor

que se analizó fue la correlación o multicolinealidad entre las variables de entrada. La multicolinealidad existe cuando se encuentran correlaciones significativas entre las variables predictoras (Goyes & Hernandez, 2021), es decir cuando los datos están relacionados linealmente entre sí y por lo tanto tienen coeficientes de correlación altos (Tabla 10).

Figura 11

Análisis de dispersión para las variables de entrada



La estimación de la colinealidad se realizó mediante el cálculo del coeficiente de correlación de Pearson con el fin de probar la interdependencia entre las 11 variables predictoras (Figura 12). Se deduce que hay 5 variables cuyos coeficientes absolutos de correlación r son moderados a altos según la Tabla 10. Los valores están entre 0.77, 0.62, 0.63 y 0.59 y corresponden a las variables X7-X11, X8-X11, X10-X11, X7-X10 y X7-X9 (Figura 13).

Tabla 10

Interpretación de los valores del coeficiente de correlación de Pearson

Valor de r	Fuerza de la correlación
$ r < 0.3$	Ninguna o débil
$0.3 < r < 0.5$	Débil
$0.5 < r < 0.7$	Moderada
$ r > 0.7$	Fuerte

Nota. Tomado de Mindrila & Balentyne (2017).

Figura 12

Matriz de correlación para las variables predictoras

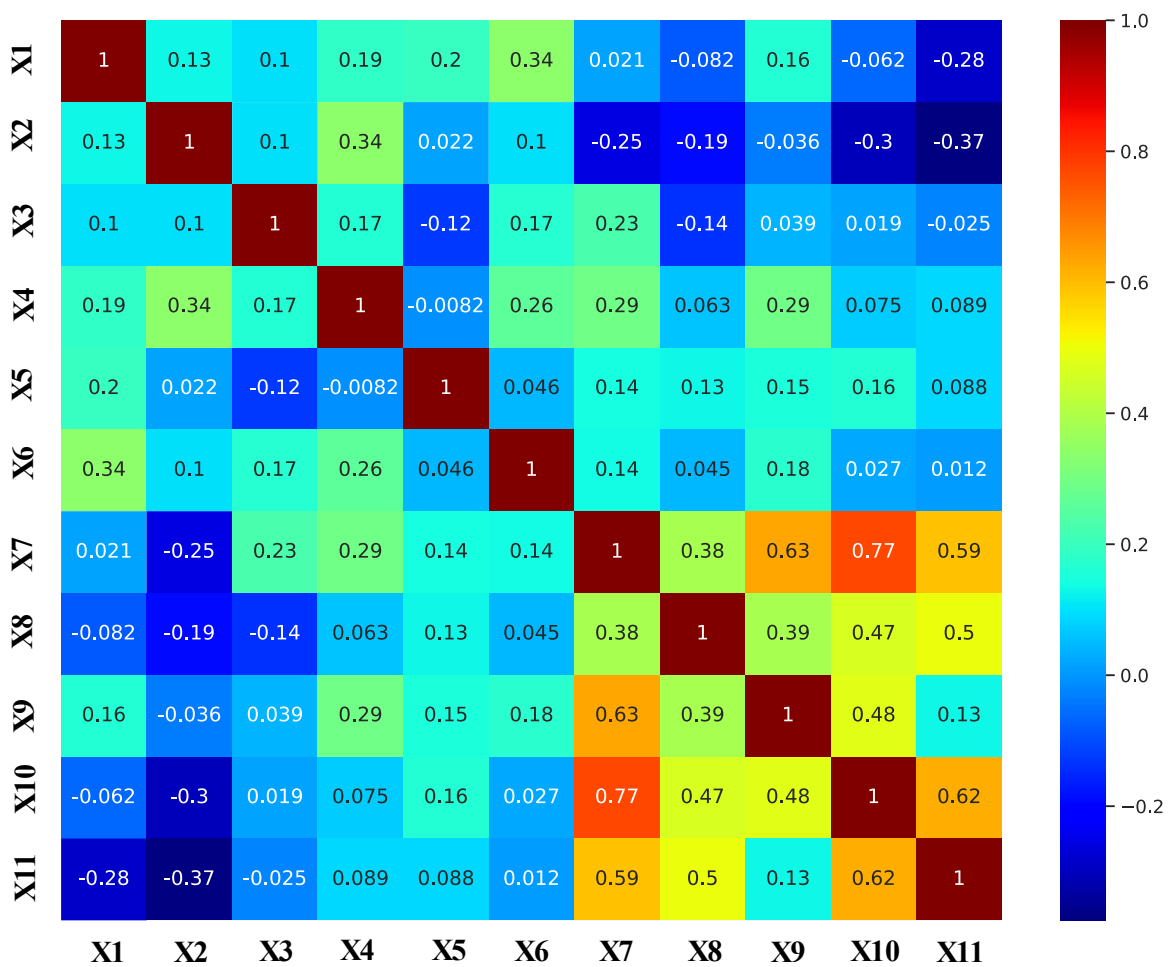
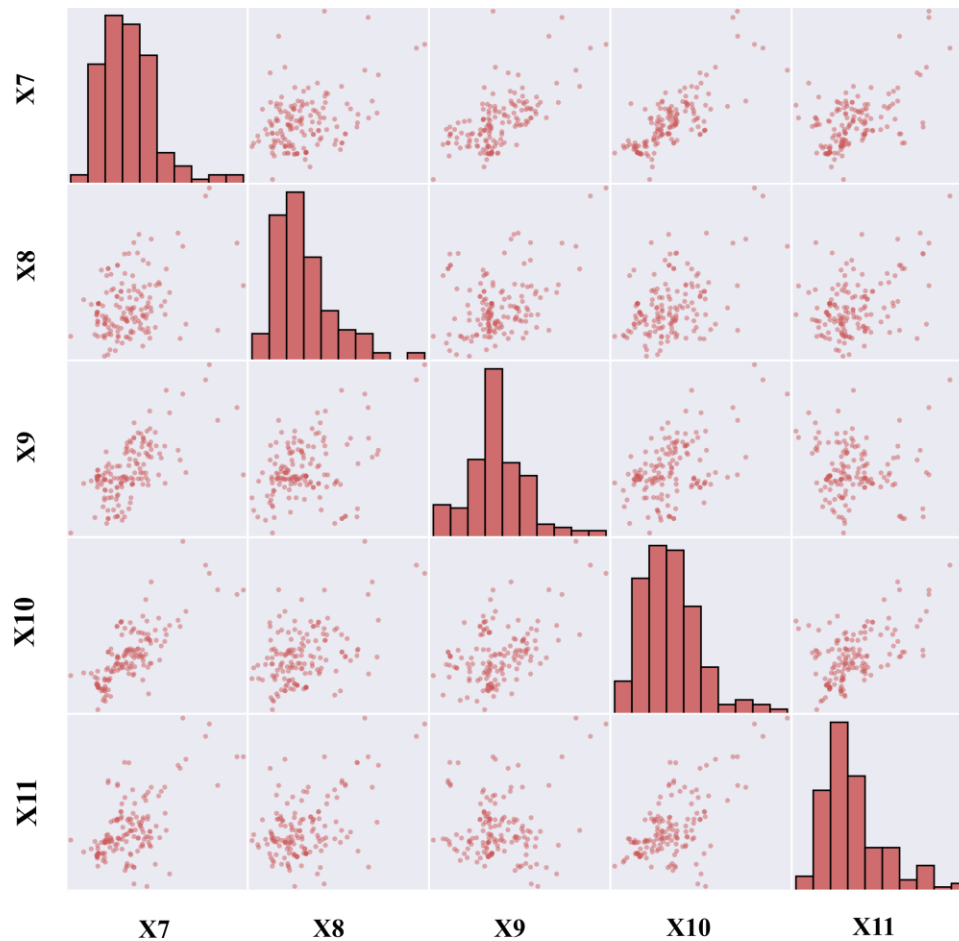


Figura 13

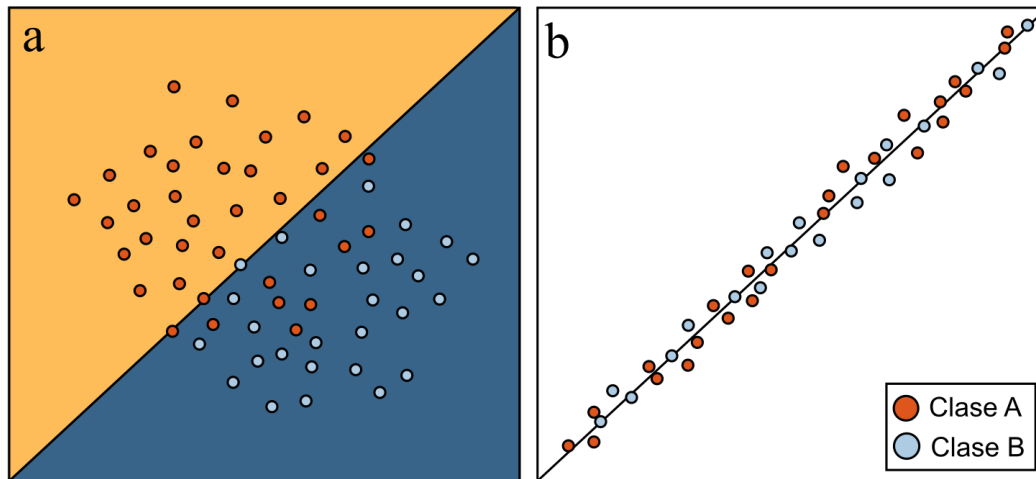
Matriz de dispersión para las variables fuertemente correlacionadas



Es importante resaltar las desventajas de la multicolinealidad en la predicción de depósitos minerales usando técnicas de aprendizaje automático. El problema radica en que la superficie de decisión entre la clase A y B, es decir pórfido y no-pórfido respectivamente, puede representarse como una línea recta o curva dependiendo del algoritmo que resuelva la clasificación. Sin embargo, las variables con coeficientes de correlación altos indican una tendencia lineal que impedirá que los algoritmos puedan trazar la superficie de decisión, como se muestra en la Figura 14. Por tal razón, se requirió transformar los datos con una técnica de remoción de la multicolinealidad.

Figura 14

Problema de la colinealidad en tareas de clasificación



Según Gregorich et al., (2021) algunas de las metodologías que se aplican para resolver este problema son: la omisión de una de las variables afectadas, la combinación de las variables que están fuertemente correlacionadas en una única puntuación compuesta o el cambio a modelados capaces de manejar este tipo de variables. Algunas de las técnicas más comunes son aquellas que reducen la dimensionalidad, por ejemplo, el análisis de componentes principales (*PCA*, por sus siglas en inglés) (Jolliffe, 1987), análisis de componentes principales probabilístico (ACPP) (Tipping & Bishop, 1999), análisis de componentes principales con Kernel (APK) (Scholkopf et al., 1999), entre otros.

En este caso se optó por usar el *PCA* que es una técnica estadística que describe la estructura de covarianza de los datos mediante componentes que representan combinaciones lineales de las variables originales y que clasifican la variabilidad de los datos a través de las direcciones de varianza (Awange et al., 2018). Es decir, el proceso consiste en descomponer la matriz de covarianza en las direcciones de máxima varianza de N variables de entrada con D

número de muestras distribuidas en P dimensiones, con base en las cuales se aplica una transformación lineal que proyecta los datos en un número Z de dimensiones donde se cumple que $Z < P$. Para esto se calculan los *eigenvectors* que representan la dirección en la cual hay más varianza en los datos y cuya magnitud está dada por los *eigenvalues*. Este proceso implica que se cumpla la ecuación (8)

$$A\vec{v} = \lambda\vec{v} \quad (8)$$

Donde A es la matriz de transformación, \vec{v} el *eigenvector* y λ el *eigenvalue*. A partir de estos se calculan los componentes principales PC_1, PC_2, \dots, PC_N , los cuales por teoría son ortogonales entre sí, es decir que no están correlacionados (Goyes-Peñañiel & Hernandez-Rojas, 2021). Cabe resaltar que PC_1 es el componente asociado con el *eigenvalue* más alto, seguido por PC_2 que corresponde al segundo *eigenvalue* más alto, es decir la segunda dirección de mayor varianza pero que a su vez es ortogonal al PC_1 , garantizando que ninguna de las dos esté correlacionada. El mismo proceso es realizado para calcular los otros componentes principales.

En resumen, el *PCA* actúa sobre una matriz de datos A que tiene P variables de entrada con D número de muestras, realizando combinaciones lineales tal que se obtenga una nueva matriz de datos B con D número de muestras y Z variables de entrada, tal que $P < Z$.

$$A \in \mathbb{R}^{D \times P} \xrightarrow{PCA} B \in \mathbb{R}^{D \times Z} \quad (9)$$

Si bien, es necesario garantizar que el número de componentes principales almacene la cantidad de información estadística suficiente que represente las nuevas variables de entrada. Para esto se recurre a la varianza explicada acumulada, cuyo valor explica la variación de los datos y

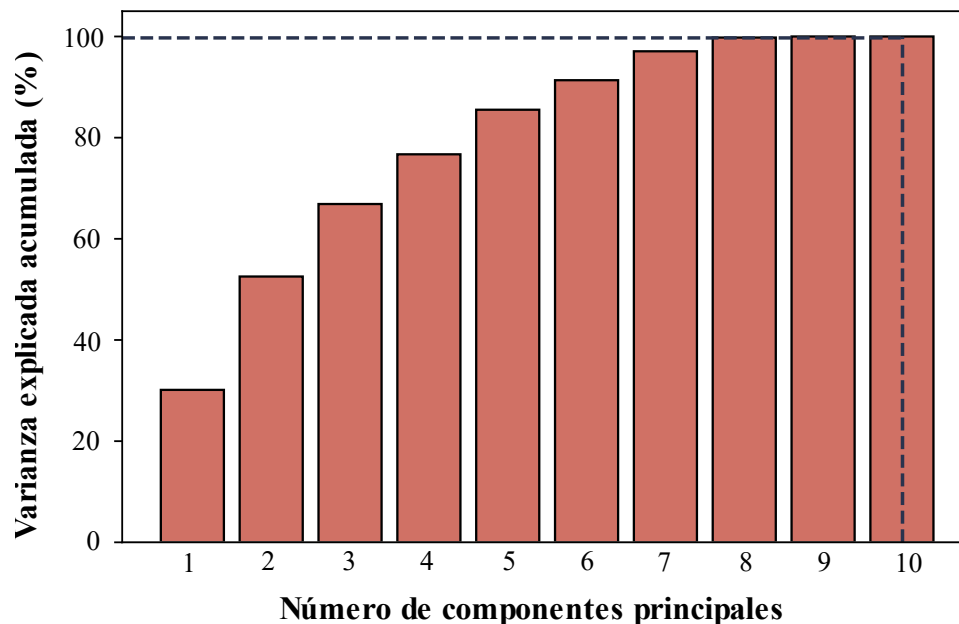
garantiza que no exista pérdida de información estadística. Con base en la información extraída de la Figura 11, Figura 12 y Figura 13 se deduce que:

- a) La naturaleza de la variable X1 afecta la estimación de probabilidad de depósitos tipo pórfido en los algoritmos de aprendizaje automático.
- b) Las variables están fuertemente correlacionadas por lo cual se requiere aplicar una técnica de reducción de dimensionalidad.

Por lo tanto, la variable X1 no se usará en el entrenamiento de los modelos y, por lo tanto, tampoco será analizada ni transformada junto con X2-X11. En este orden de ideas, se aplicó la técnica *PCA* a los diez mapas predictores originales y se volvieron a extraer los valores en los puntos de depósito y no-depósito que representan los puntos de entrenamiento.

Figura 15

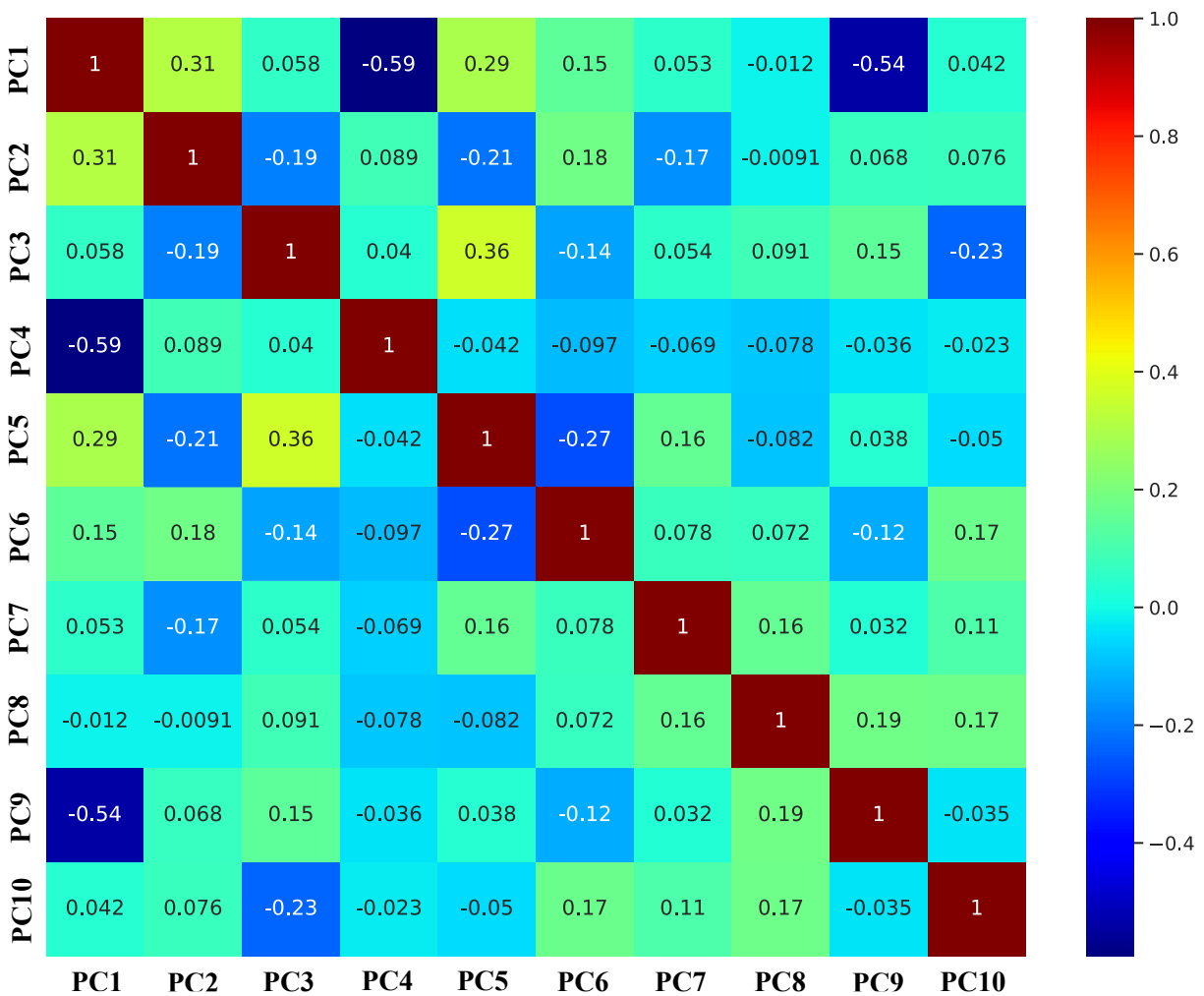
Porcentaje de varianza explicada acumulada con diez componentes principales



Posteriormente, se analizó el porcentaje de varianza acumulada para diez componentes principales. En este caso, se obtuvo un valor de 99.99 % (Figura 15). De esta forma se garantizó la descorrelación entre las variables de entrada. Por lo tanto, los diez componentes principales son los nuevos mapas predictores que representan las capas de entrada para el cálculo de probabilidades.

Figura 16

Matriz de correlación para las diez variables predictoras PC1-PC11



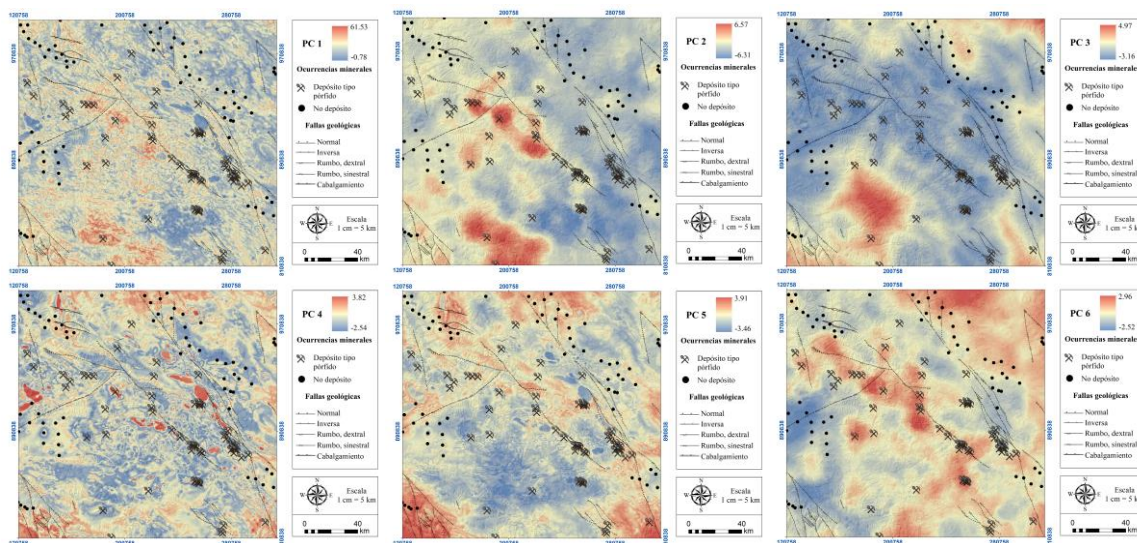
Para verificar que la técnica de *PCA* fue efectiva y cumplió con su objetivo, se analizó la matriz de correlación con los diez componentes principales. Se obtuvieron coeficientes de correlación de Pearson máximos de -0.59, 0.36, 0.29 para el caso de PC1-PC4, PC3-PC5 y PC1-PC5, respectivamente, lo cual se considera correlación moderada a débil según la Tabla 10. El resto de los valores son del orden $\times 10^{-2}$, es decir, se removió la multicolinealidad y de esta manera los datos se ajustaron para ser usados en el entrenamiento computacional. El resultado se presenta en la Figura 16.

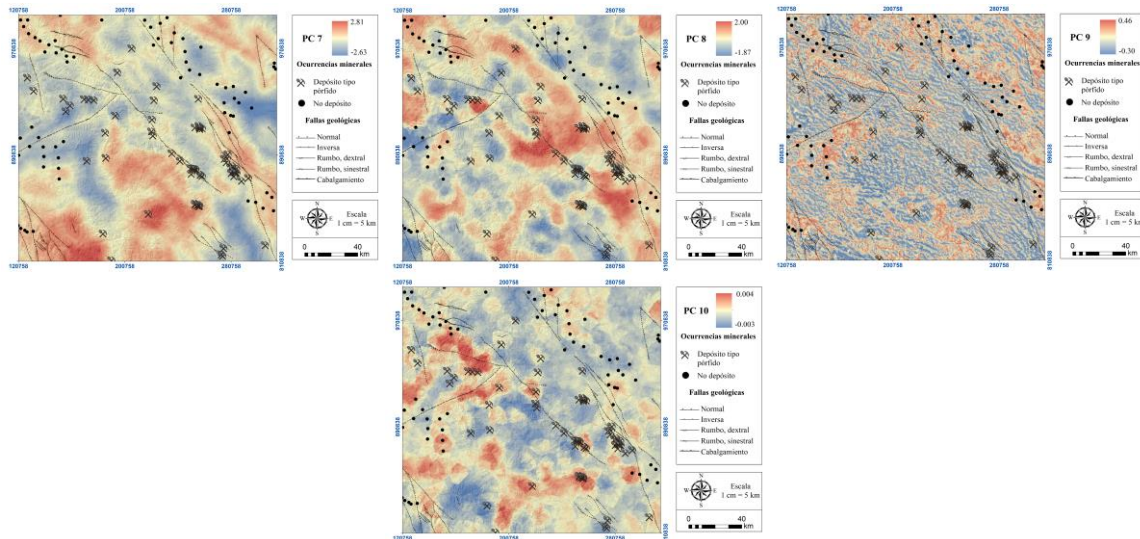
5.1.5 Creación de Ráster Virtual

Una vez se verificó que la colinealidad fue removida, se procedió a crear el ráster virtual con los diez componentes principales. Esta capa contiene los valores de píxeles que serán usados en la predicción de ocurrencias minerales tipo pórfido en el área de estudio. Los nuevos mapas predictores se presentan en la Figura 17 y en el Apéndice B en alta resolución.

Figura 17

Mapas predictores con PCA





5.2 Diseño modelos de aprendizaje automático usando redes neuronales artificiales, bosques aleatorios y máquinas de soporte vectorial en el lenguaje de programación *Python*.

El diseño y entrenamiento de los modelos de aprendizaje automático conduce a la clasificación de depósitos minerales tipo pórfido y no-pórfido y por lo tanto a la creación del mapa de prospectividad mineral de la zona de estudio. Se implementaron tres algoritmos computacionales que corresponden a redes neuronales artificiales (*ANN*), bosques aleatorios (*RF*) y máquinas de soporte vectorial (*SVM*). La creación y el entrenamiento de los modelos se realizó usando el lenguaje de programación Python en el entorno de desarrollo integrado (IDE) Google Colab Pro con una GPU. Los datos de entrenamiento fueron cargados y visualizados en Colab PRO con la librería Pandas (<https://pandas.pydata.org/>).

Dado que el conjunto de datos de entrenamiento no es lo suficientemente grande (148) como para dividirlo en datos de entrenamiento y datos de validación, se seleccionaron aleatoriamente 20 datos del conjunto usado en el entrenamiento para medir métricas como *ROC*, *AUC* y matriz de confusión.

5.2.1 Redes Neuronales Artificiales (ANN)

El modelo de redes neuronales artificiales se implementó en la librería Keras de TensorFlow, con los hiperparámetros definidos en la Tabla 11. Se usaron diez variables de entrada correspondientes a los diez componentes principales. La arquitectura consistió en diez capas ocultas, cada una con diez neuronas y una capa de *dropout*. Esta técnica evita el *overfitting* (sobreentrenamiento) y consiste en desconectar aleatoriamente ciertas unidades de la red neuronal, es decir, eliminarlas temporalmente del modelo junto con todas sus conexiones entrantes y salientes (Srivastava et al., 2014). Para más detalles y comparaciones sobre el uso de este método se recomienda leer Srivastava et al., (2014).

Tabla 11

Hiperparámetros definidos en el modelo de ANN

Hiperparámetros	Descripción
Variables de entrada	10
Capas ocultas	10
Dropout	0.2 (entre cada capa oculta)
Neuronas requeridas en cada capa oculta	10
Función de activación capas ocultas	ReLU
Función de activación capa de salida	Sigmoide
Optimizador	Adam
Tasa de aprendizaje	1×10^{-3}
Función de pérdida	Entropía cruzada binaria
Métricas	Precisión, ROC, AUC, matriz de confusión
Épocas	5000

El entrenamiento y la predicción de los depósitos minerales con los diez componentes principales presentó un sesgo que no permitió definir una superficie de decisión. Por lo tanto, se realizaron pruebas variando los factores condicionantes iniciales para revisar cuál de las variables

desfavorecía la predicción. De esta forma, X2 se eliminó debido a que no resultaba ventajosa para el entrenamiento del modelo. Por consiguiente, se realizó nuevamente el análisis de componentes principales teniendo únicamente en cuenta X3-X11, lo cual resultó en nueve componentes principales (PC), que representan el nuevo insumo para generar los mapas de probabilidad. Los mapas predictores con estas variables se encuentran en el Apéndice C.

Para garantizar la confiabilidad en el modelo se realizó 10 veces el entrenamiento de la red neuronal y con base en este, se hizo un promedio de las métricas de precisión y función de pérdida en cada etapa. Lo anterior teniendo en cuenta que se presentan variaciones debido a que los pesos de la red se inicializan aleatoriamente (ver 3.3.1). En la Tabla 12 se presenta un resumen estadístico de la variación en la precisión y el valor de la función de pérdida en los 10 entrenamientos. El total de parámetros entrenables fue de 1.101. Se alcanzó una precisión de 94.59 % y el valor de la función de entropía cruzada binaria convergió hasta 0.1688. En la Figura 18 se muestra la variación de estas métricas con el número de épocas.

Tabla 12

Resumen del entrenamiento con ANN

Atributo	Valor
Parámetros entrenables	1.101
Precisión	0.9459 (94.59 %)
Función de pérdida	0.1688
Tiempo de entrenamiento	258.6 (s), 4.31 (m)
Matriz de confusión	Verdaderos positivos (VP): 10 Falsos positivos (FP): 0 Falsos negativos (FN): 2 Verdaderos negativos (VN): 8
AUC	0.945

Figura 18

Resumen estadístico de la evolución de la precisión y valor de función de pérdida en cada entrenamiento ANN

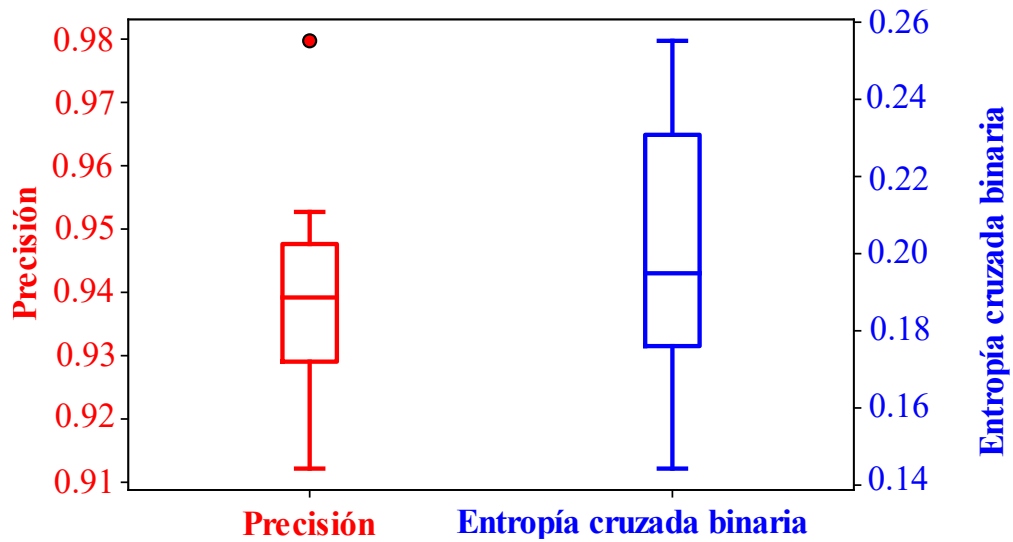
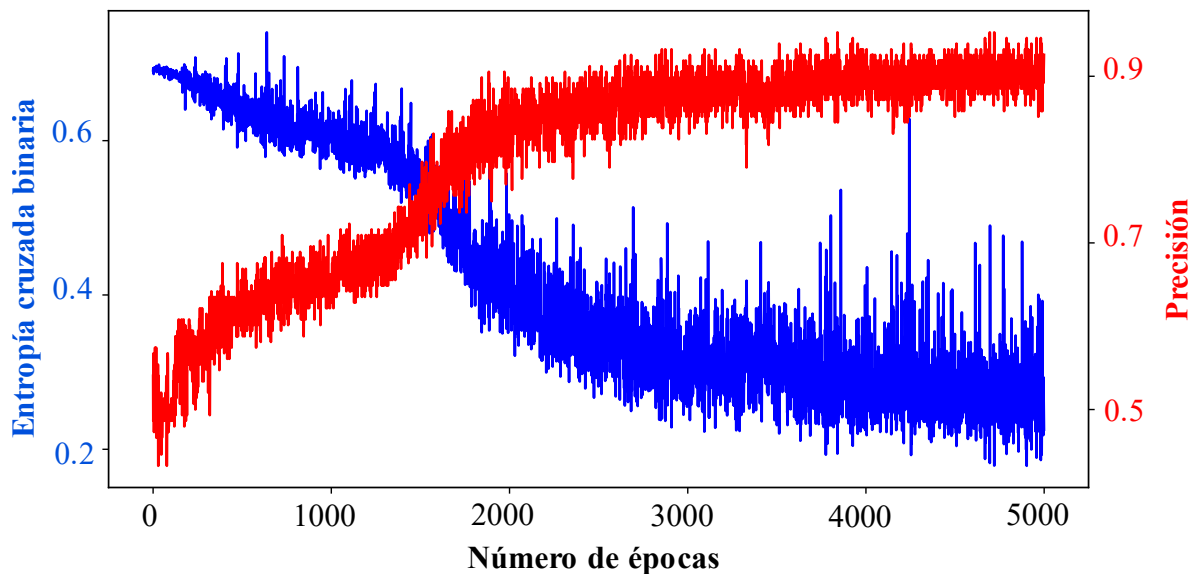


Figura 19

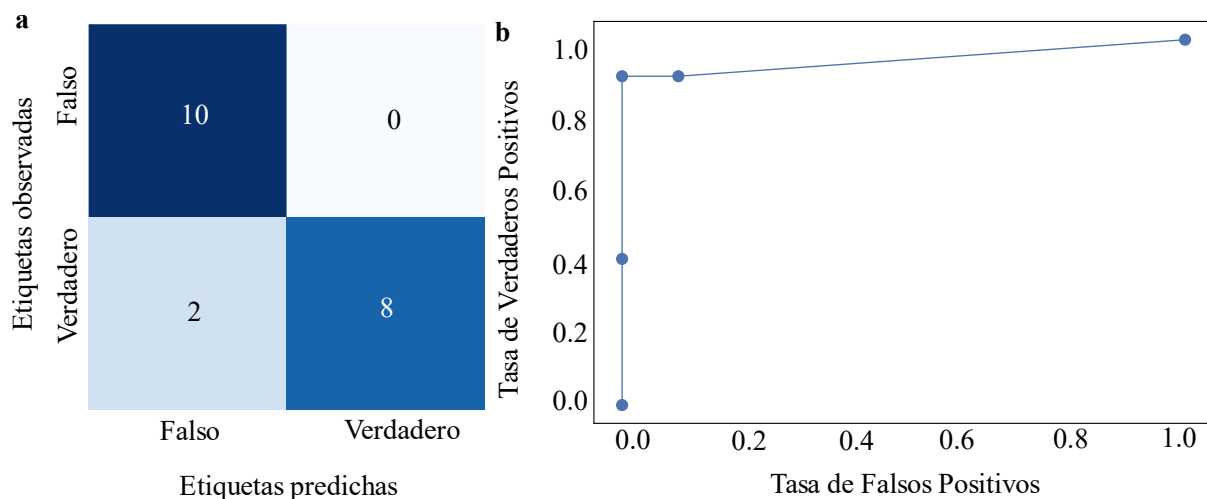
Métricas del modelo ANN



Las métricas evaluadas con los datos seleccionados para validación mostraron un buen rendimiento para el modelo. En la Figura 20 se muestra la matriz de confusión, la cual muestra la relación entre la tasa de verdaderos positivos y falsos positivos. De esta forma, se extraen los errores de predicción del modelo teniendo en cuenta que, si en el eje x hay valores pequeños, el número de falsos positivos es bajo y el de verdaderos negativos alto. Por otro lado, si en el eje y hay valores grandes el número de verdaderos positivos es alto y el de falsos negativos es bajo.

Figura 20

Métricas de validación para ANN



La matriz de confusión muestra que el algoritmo clasificó 10 puntos como verdaderos positivos (VP), es decir, datos cuya etiqueta real era 1 y se predijo como 1. Ningún punto fue clasificado como falso positivo (FP), es decir, un dato cuya etiqueta real era 0 pero el algoritmo predijo como 1. También, se presentaron 2 falsos negativos (FN), es decir, etiquetas cuyo valor real era 1 pero fueron predichos como 0. Finalmente, hubo 8 verdaderos negativos (VN), es decir, etiquetas con valor real de 0 y cuya predicción fue 0. En general, el modelo clasificó de manera

correcta 18 puntos de entrenamiento y tan solo 2 de manera errónea, lo que indica que el rendimiento de *ANN* para clasificar los depósitos (1) y no-depósitos fue muy bueno.

Por otro lado, se evaluó el comportamiento de la curva Característica Operativa del Receptor (*ROC*, por sus siglas en inglés) que define la relación entre la tasa de verdaderos positivos y falsos positivos. De esta forma, se extraen los errores de predicción del modelo teniendo en cuenta que, si en el eje x hay valores pequeños, el número de falsos positivos es bajo y el de verdaderos negativos alto. Por otro lado, si en el eje y hay valores grandes el número de verdaderos positivos es alto y el de falsos negativos es bajo.

Por eso, en un caso ideal la curva *ROC* tendría un comportamiento donde que se desplaza hacia la esquina superior izquierda, lo cual quiere decir que hay mayor número de verdaderos positivos. Es decir, el área bajo la curva (*AUC*, por sus siglas en inglés), es mayor y se acerca a 1. En este caso, la calidad del modelo está dada por la Figura 20 y el *AUC* obtenido fue de 0.945 que representa resultado muy cercano al ideal donde este valor es igual a uno.

5.2.2 Bosques aleatorios (*RF*)

El modelo de bosques aleatorios se generó usando la librería TensorFlow. Se definieron los hiperparámetros de la Tabla 13. En este caso se implementó el algoritmo categórico *CART* que usa el índice Gini definido en la Ecuación (2) para analizar los valores numéricos y calcular la impureza en los datos.

Tabla 13

Hiperparámetros definidos en el modelo de RF

Hiperparámetros	Descripción
Número de árboles de decisión	2000

Algoritmo categórico	CART
Función de pérdida	Entropía cruzada binaria
Métricas	Precisión, ROC, AUC, matriz de confusión

Figura 21

Modelo grafo del modelo RF con máxima profundidad de 2

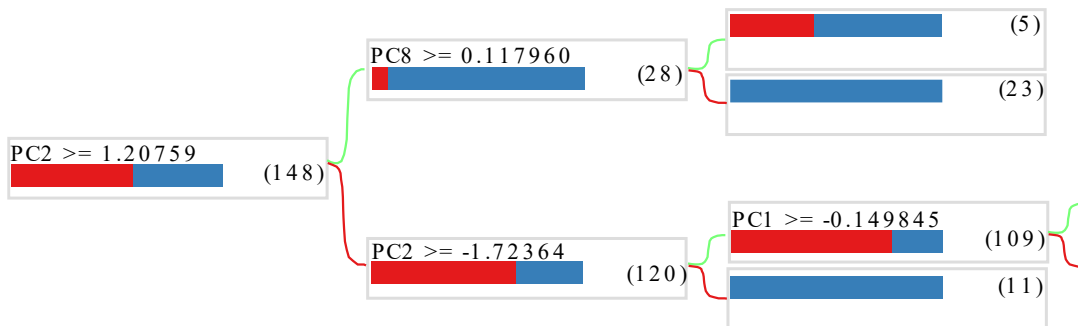
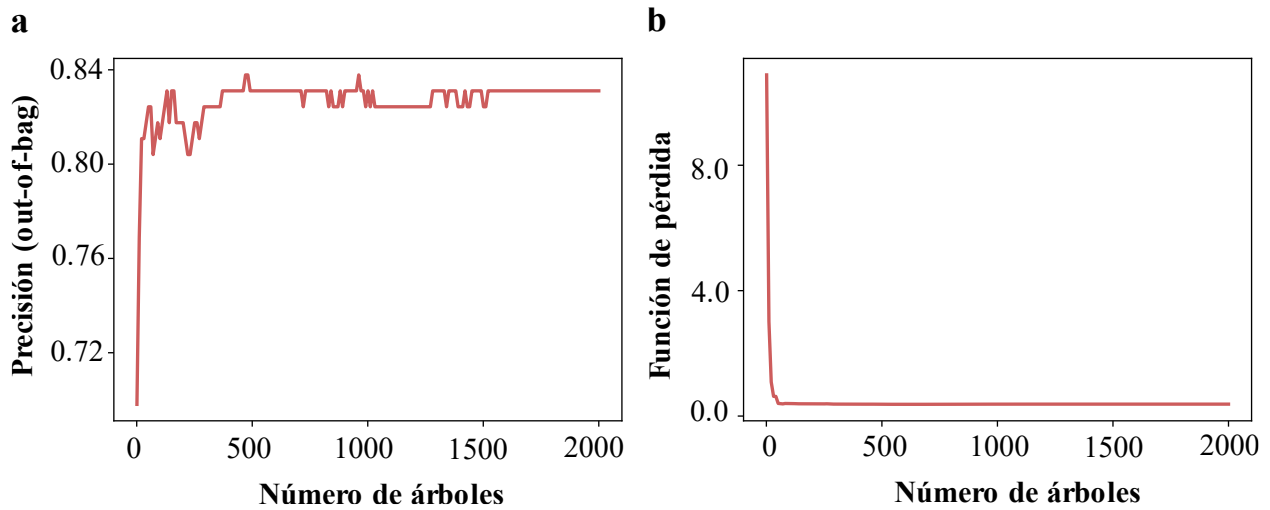


Figura 22

Métricas del entrenamiento del modelo RF. evolución de a) la precisión y b) la función de pérdida respecto al número de árboles de decisión



La precisión alcanzada por el modelo fue de 83.10 % y se estabiliza a partir de aproximadamente 1500 árboles de decisión. La función de pérdida alcanzó un mínimo de 0.380721 y el tiempo total de entrenamiento fue de 7.1851 (s). En la Figura 22 y Tabla 14 se presenta un resumen de los resultados obtenidos.

En la Figura 23 se muestra la matriz de confusión donde se evidencia que el algoritmo clasificó 10 puntos como verdaderos positivos (VP), 0 como falsos positivos (FP), 0 falsos negativos (FN) y 10 verdaderos negativos (VN). En decir, el modelo clasificó de manera correcta los 20 puntos de validación. De esta manera el *AUC* tuvo un valor de 0.9996.

5.2.3 Máquinas de soporte vectorial (SVM)

El modelo de máquinas de soporte vectorial se generó usando la librería scikit-learn. Se definieron los hiperparámetros de la Tabla 15. Se escogieron con la técnica de validación cruzada implementada en Python, la cual evalúa todas las combinaciones posibles de modelos de entrenamiento, calcula la precisión con cada una de ellas y, finalmente, escoge la de mejor rendimiento.

Tabla 15

Hiperparámetros definidos en el modelo de SVM

Hiperparámetros	Descripción
C	10
Gamma	0.1
Kernel	rbf (<i>radial basis function</i>)
Métricas	Precisión, ROC, AUC, matriz de confusión

El entrenamiento obtuvo una precisión del 94.59 % y el tiempo de entrenamiento de 0.011771 (s). En la Tabla 16 se resumen las métricas del modelo. La matriz de confusión (Figura

24) muestra que el algoritmo clasificó 10 puntos como verdaderos positivos (VP), 0 como falsos positivos (FP), 2 falsos negativos (FN) y 8 verdaderos negativos (VN). En decir, el modelo clasificó de manera correcta los 20 puntos de validación. El *AUC* tuvo un valor de 0.9.

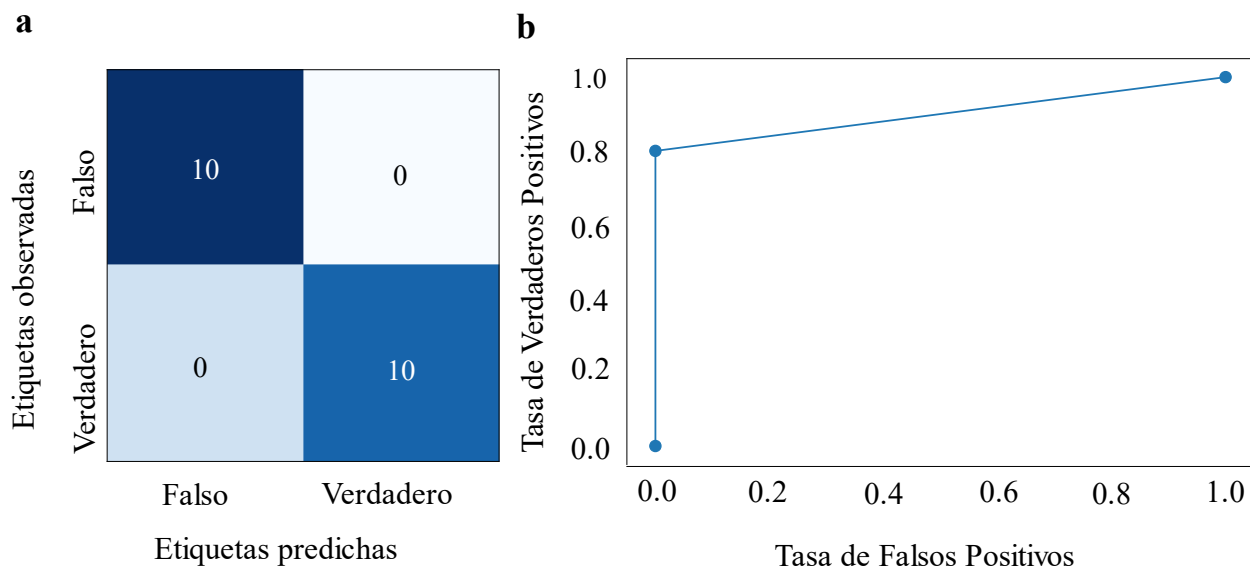
Tabla 16

Resumen del modelo SVM

Atributo	Nueve PC
Precisión	0.9459 (94.59 %)
Tiempo de entrenamiento	0.011771 (s)
Matriz de confusión	Verdaderos positivos (VP): 10 Falsos positivos (FP): 0 Falsos negativos (FN): 2 Verdaderos negativos (VN): 8
AUC	0.9

Figura 24

Métricas de validación SVM



De manera general, el entrenamiento de los tres modelos de aprendizaje automático *ANN*, *RF* y *SVM* tuvo buenos resultados. Se destaca que en los tres casos la precisión fue alta, con el

valor más bajo obtenido por *RF* con 83%. Sin embargo, este modelo fue el que clasificó menor número de depósitos de forma incorrecta, lo cual se evidenció en la matriz de confusión con los datos seleccionados para validación.

Los algoritmos evaluados convergieron a un error mínimo, por lo tanto, tanto la estructura del modelo como de los datos es suficientemente buena para realizar la predicción de probabilidad en la zona de estudio y, por consiguiente, definir objetivos de exploración. En la siguiente sección se muestran los mapas obtenidos.

5.3 Implementación de una metodología utilizando los modelos de aprendizaje automático y la información geológica y geofísica para predecir la probabilidad de ocurrencia de depósitos minerales tipo pórfido.

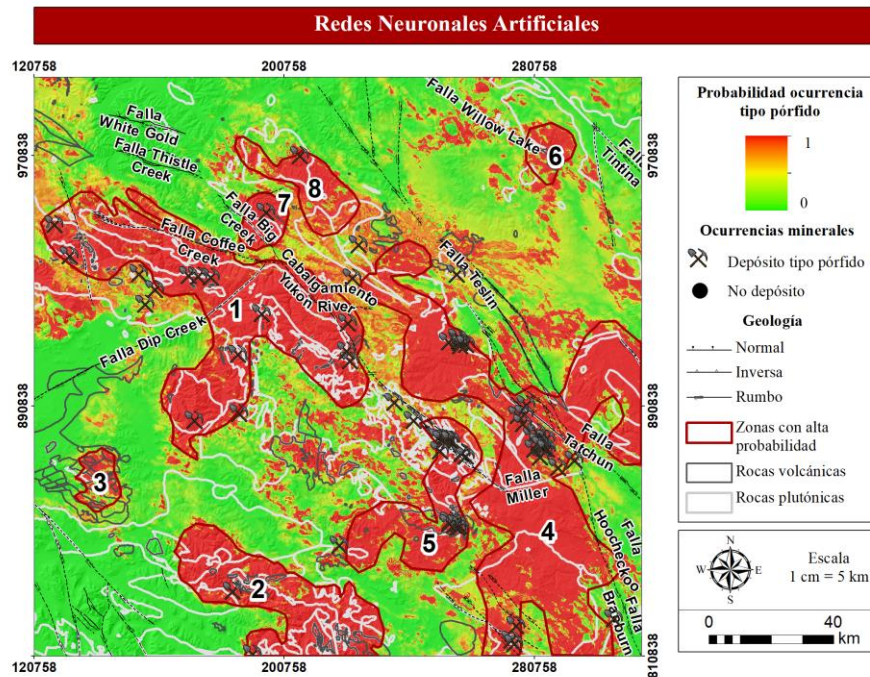
Los mapas de probabilidad derivados de las variables geológicas explicatorias fueron generados de forma efectiva y satisfactoria usando tres modelos que cumplen con la tarea de clasificación, pero presentan diferencias en su superficie de decisión. A continuación, se presentan los mapas de probabilidad obtenidos por cada modelo y su respectiva descripción.

Los mapas en alta resolución se presentan en el Apéndice D.

5.3.1 Mapa de probabilidad con redes neuronales artificiales

Figura 25

Mapa de probabilidad con ANN



En la Figura 25 se delimitaron siete zonas con probabilidades mayores al 80% ubicadas al sur, sureste y noroeste de la zona de estudio que coinciden con la mayoría de los puntos de depósitos tipo pórfido extraídos de la base de datos del Servicio Geológico de Yukón. Se destaca un área significativa con alta probabilidad de ocurrencia (60-80% y >80%) y que corresponde a la sección (1). Es una franja en dirección NW-SE hacia el occidente de la zona de estudio, la cual es subparalela a la Falla Big Creek y el Cabalgamiento Yukon River en el SE y a la Falla Coffee Creek en el NW. Está conformada por el batolito Dawson Range del Cretácico Temprano y localmente por la unidad Sulphur Creek del Pérmico cuya litología es principalmente de monzogranitos y gneiss augen. El límite norte de este polígono rojo está en contacto con rocas

metamórficas de la unidad Snowcap del Devónico Superior que corresponde a cuarcitas, esquistos y psammitas.

Los polígonos en color rojo se encuentran mayormente sobre cuerpos plutónicos, lo cual es de suma importancia tener en cuenta debido a que es donde se hospedan la mayoría de los depósitos tipo pórfido en Yukón (Mortensen & Friend, 2021), por ejemplo, en las suites Whitehorse, Stikine y Casino (Figura 1). Así mismo, con base en el mapa se deduce que no hay necesariamente una relación directa con las rocas volcánicas y la ocurrencia mineral tipo pórfido, excepto en el sur donde se concentran depósitos en zonas de probabilidad entre 60-80% hacia las secciones (2), (3) y (5).

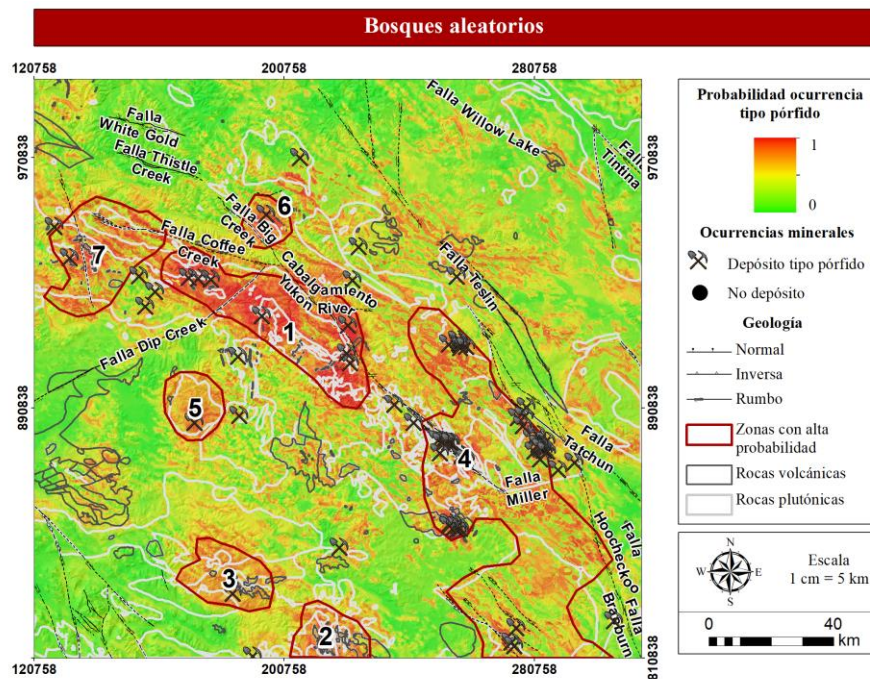
Algunas de las unidades volcánicas relacionadas con estas zonas son la suite Rhyolite Creek del Paleoceno compuesta por andesitas, dacitas, riolitas, brechas y menormente basaltos. Lo anterior está sustentado ampliamente en los modelos geológicos de Yukón, donde se afirma que solo algunas de las intrusiones más antiguas del Triásico tardío se superponen en edad con las rocas volcánicas en el Terreno de Stikinia (Sack et al., 2020, como se cita en Mortensen & Friend, 2021), por lo tanto, las rocas volcánicas no están exclusivamente relacionadas con la ocurrencia de depósitos tipo pórfido en la zona de estudio. Además, en la sección (3) están en contacto con la unidad Finlayson del Devónico Superior cuya litología son rocas metasedimentarias carbonáceas, metachert y esquistos pelíticos.

Es decir, el algoritmo marcó como una zona de alta probabilidad a un conjunto de unidades ígneas y metamórficas en el área de estudio, por lo cual es preciso tener esta variable en cuenta como capa de entrada en próximos trabajos de investigación relacionados con la temática. En las secciones (4), (6), (7) y (8) las zonas rojas corresponden principalmente a rocas plutónicas.

5.3.2 Mapa de probabilidad con bosques aleatorios

Figura 26

Mapa de probabilidad con RF

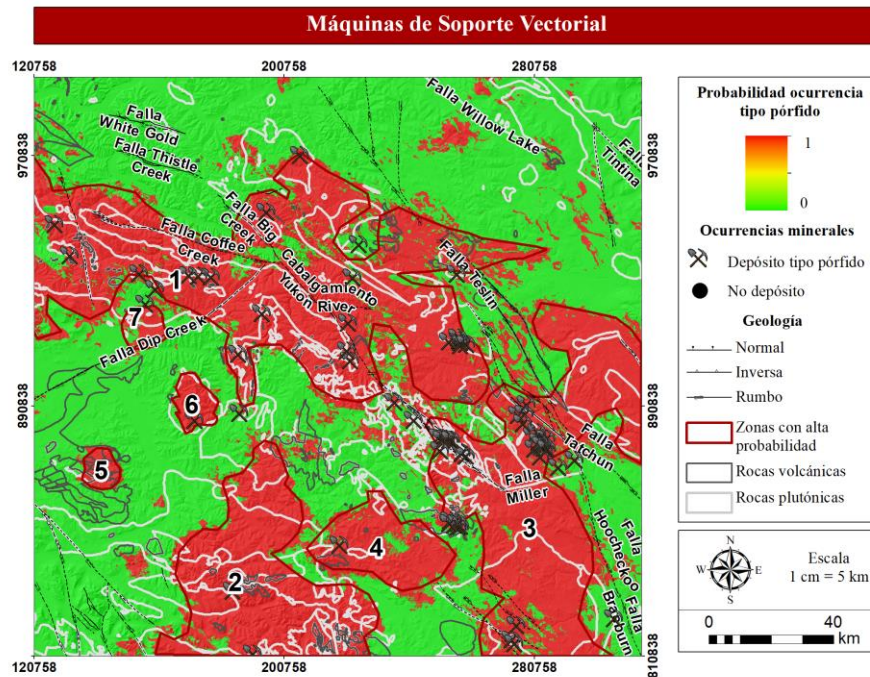


En la Figura 26 se delimitaron siete zonas prospectivas. Se destaca solamente un polígono con probabilidades mayores al 80% correspondiente a la zona (1) y que coincide con la zona (1) del modelo *ANN*. Están asociados con el terreno Yukon Tanana (Figura 1). Así mismo, *RF* delimitó con probabilidades del 60-80% los polígonos de la zona sur (2) y (3) que fueron marcados con valores >80% en el modelo *ANN*. Además, hacia el sector suroccidental sobresalen cuerpos volcánicos que en *RF* son marcados con bajas probabilidades (40-60%) pero que en *ANN* son delimitados con valores >80% en la sección (3).

5.3.3 Mapa de probabilidad con máquinas de soporte vectorial

Figura 27

Mapa de probabilidad con SVM



En la Figura 27 se delimitaron seis zonas prospectivas: la sección (1) clasificada con alta probabilidad al igual que en ANN y RF. Los límites de los polígonos en este mapa son menos transicionales, por lo cual se tienen valores muy cercanos a 1 o muy cercanos a 0. De la misma forma con las zonas (5) y (4) en ANN y SVM, respectivamente, ubicadas al suroriente donde se encuentra la unidad Snowcap y el depósito Maloney. En esta sección el modelo RF identificó valores muy bajos de probabilidad (<40%).

5.3.4 *Análisis descriptivo y comparativo de los mapas de probabilidad*

En este apartado se compararon las zonas prospectivas delimitadas con los mapas de probabilidad de los tres modelos de aprendizaje automático. Los resultados se relacionaron con la información de reportes geológicos de exploración mineral y los puntos de depósitos conocidos.

En las inmediaciones de la falla Teslin hacia el NE, cerca de la sección (4), el modelo *RF* clasifica la zona con valores de baja probabilidad entre el rango 40-60% e incluso <40%, mientras que *ANN* predice valores >80%. En este caso es más confiable el modelo *ANN* porque entre el sistema estructural se encuentra el depósito Stu - Zone D, por lo cual se esperan altos valores de probabilidad. También se distingue un polígono subredondeado en la sección (5) que, si bien en *ANN* se delimita con color rojo, no posee una forma definida a diferencia del modelo *RF*. De manera similar, en la zona (6) se marcan valores altos de probabilidad hacia el centro del cuerpo y medios hacia sus bordes, lo cual es concordante con la presencia del depósito Scroggie. Este cuerpo está relacionado con el Plutón Onasick de la suite Whitehorse que está delimitado al sur por la Falla Big Creek y conforma la parte con probabilidades >80% en la sección (6).

En términos generales, *RF* clasifica con probabilidades medias a bajas la mayor parte de los polígonos que *ANN* predijo con valores altos, a excepción de la franja (1) donde aflora el batolito Dawson Range, en el cual ambos modelos coinciden con valores >80%.

En resumen, al menos tres zonas fueron identificadas con probabilidades altas y medias en los tres mapas predictores y corresponden según la Figura 25 a las secciones (1), (2) y (4). De esta manera, en este trabajo de investigación se consideran que los tres modelos tienen muy buena capacidad predictiva, pues las zonas clasificadas con alta probabilidad son concordantes con la presencia de depósitos minerales tipo pórfido y atributos geológicos favorables para su formación.

Por tal razón, se plantea una metodología para implementar la creación de un mapa de probabilidad más confiable usando los tres algoritmos computacionales.

5.3.5 Propuesta de metodología

La propuesta adicional para implementar la metodología consiste en realizar un aumento de datos de entrenamiento y dos operaciones entre bandas: la primera corresponde a un promedio aritmético que extraiga las características distintivas en común de los tres modelos. La segunda corresponde a un promedio geométrico que extraiga las características distintivas, es decir, las altas probabilidades, que se cumplan en todos los modelos.

5.3.5.1 Aumento de datos

El aumento de datos consistió en asignar nuevos puntos de depósito tipo pórfido teniendo en cuenta los siguientes factores: en primer lugar, la cercanía a los depósitos donde existe una alta probabilidad de que exista otro depósito mineral. En segundo lugar, la premisa de que las mineralizaciones son un fenómeno geológico raro, por lo cual se garantizó que existiera máximo un depósito en cada celda unitaria, es decir, en cada píxel (Carranza, 2009). En tercer lugar, la distancia mínima y máxima entre los depósitos tipo pórfido en la zona de estudio. En Mihalasky et al., (2011) se plantea una zona de influencia de 2.5 km de radio para cada punto clasificado como depósito.

En cuarto lugar, se consideró que los píxeles donde se aumentaron los datos estuvieran en zonas donde la probabilidad de ocurrencia fuera alta en los tres mapas de probabilidad generados en la sección anterior. Por consiguiente, se tomaron en cuenta puntos en los píxeles cercanos a los depósitos ya conocidos que estuvieron dentro de un búfer de 5 km (2.5 km de radio) y en zonas de alta probabilidad de ocurrencia mineral. Teniendo en cuenta estas consideraciones se garantizó

que no se indujera sesgo en el modelo computacional. De este modo, se aumentaron los datos de entrenamiento a 289 y se volvieron a ejecutar los tres modelos de aprendizaje automático. Las métricas obtenidas con este nuevo entrenamiento se presentan en la Tabla 17. Los nuevos mapas de probabilidad se encuentran en la Figura 28 y el Apéndice E.

Tabla 17

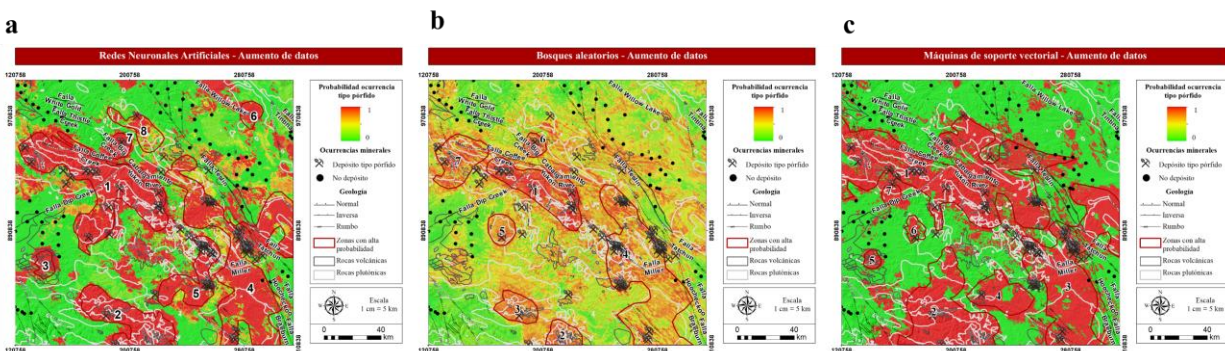
Precisión alcanzada por cada modelo con el aumento de datos a 289 puntos de entrenamiento

Modelo	Precisión con aumento de datos (%)	Precisión inicial (%)
ANN	98.96	94.59
RF	89.61	83.10
SVM	97.92	94.59

Figura 28

Mapas de probabilidad con el aumento de datos de entrenamiento para los modelos de: a)

ANN, b) RF y c) SVM.



Con base en la Tabla 17 el modelo con mejor resultado fue el de ANN, seguido por SVM y luego por RF. En comparación con la precisión inicial mostrada en la sección 5.2 se mejoró significativamente el rendimiento de cada uno de los modelos. Los mapas de probabilidad

obtenidos en esta etapa con el aumento de datos de entrenamiento se presentan en la Figura 28. El resumen de las características que se mantuvieron o eliminaron en cada zona delimitada como prospecto según el modelo se presenta en la Tabla 18.

Tabla 18

Comparación de probabilidades con el aumento de datos de entrenamiento

<i>Comparación de probabilidades con el aumento de datos</i>		
<i>ANN</i>	<i>RF</i>	<i>SVM</i>
En las zonas (1), (2), (4), (5), (6) y (7) se mantuvieron en altas probabilidades	Se mantuvo la zona (1) con alta probabilidad, concentrándose a lo largo de la Falla Coffee Creek y el cabalgamiento Yukon River	Las zonas (1) y (2) se mantuvieron similares al modelo anterior
En la zona (3) el polígono redujo su tamaño y adoptó una forma más redondeada	Se destacan las zonas (2) y (3) que tienen valores más altos de probabilidad en comparación con el modelo sin el aumento de datos	Sobresale una disminución de los valores de probabilidad de altos a bajos en la zona norte del polígono (2), así como el NE y SW de la sección (4) que son clasificados como no depósito según el modelo con el aumento de datos
En la zona (8) el algoritmo disminuyó el área delimitada anteriormente con alta probabilidad. Por eso, ya no se considera objetivo de exploración debido a que su tamaño es muy reducido y, no representa un punto homogéneo de valores de probabilidad	En las secciones (4), (5), (6) y (7) no se distinguen cambios significativos.	El polígono (6) redujo su área original y en la sección (3) aparecieron zonas con baja probabilidad, específicamente hacia el NE de la Falla Tatchun
En general, los cuerpos son más homogéneos y solamente tienen probabilidades del rango 80-100%	En general, se mantuvo la zona (1) con alta probabilidad y aumentaron los valores para las zonas (2) y (3)	En general, se mantuvo la zona (1) y disminuyó parcialmente la probabilidad de las zonas (2), (4), (6) y (3).

5.3.5.2 Operaciones entre bandas

Una vez se garantiza que el aumento de datos fue favorable para la delimitación de áreas con alta probabilidad de ocurrencia mineral, se propone realizar el promedio aritmético y promedio geométrico con los tres mapas generados en la sección anterior. Los resultados se presentan en la Figura 29 y Figura 30. Con base en la Figura 29 se establecieron 11 zonas como prospectos para exploración de depósitos tipo pórfido, la probabilidad de ocurrencia mineral es mayor al 80%, con valores máximos de hasta 99.98 %. En cambio, en la Figura 30 se destaca, principalmente la sección (1) con muy alta probabilidad.

5.3.5.3 Estadística zonal

Se determinó la estadística para cada uno de los polígonos o zonas. Se calculó la media, desviación estándar, la varianza, el valor máximo y el área. Los resultados se presentan en la Tabla 19 y Tabla 20.

Figura 29

Mapa del promedio aritmético

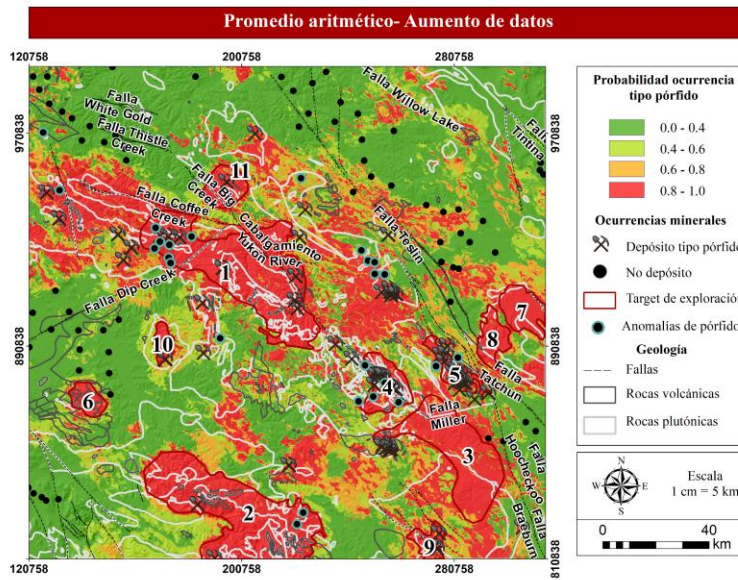


Figura 30

Mapa del promedio geométrico

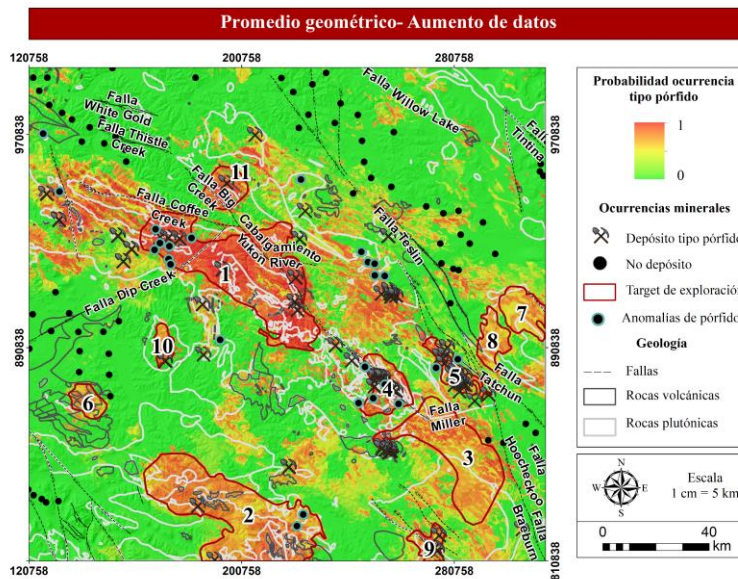


Tabla 19*Estadística zonal con el mapa de promedio aritmético*

Zona	Promedio	Desviación estándar	Varianza	Máximo	Área (ha)	% área
1	0.9701	0.0398	0.0016	0.9998	153943.97	4.21
2	0.9191	0.0424	0.0018	0.9947	141523.48	3.87
3	0.9017	0.0763	0.0058	0.9995	85982.33	2.35
4	0.9381	0.0865	0.0075	0.9998	28275.01	0.77
5	0.9177	0.1006	0.0101	0.9998	17093.40	0.47
6	0.8841	0.0792	0.0063	0.9793	12070.33	0.33
7	0.8830	0.0502	0.0025	0.9838	23188.54	0.63
8	0.8869	0.0687	0.0047	0.9895	17243.26	0.47
9	0.8920	0.1181	0.0139	0.9973	11297.00	0.31
10	0.9262	0.0284	0.0008	0.9815	7812.20	0.21
11	0.9405	0.0762	0.0058	0.9990	19538.85	0.53
Área total					517968.37	14.18
Área total zona de estudio					3652407.1	100

Tabla 20*Estadística zonal con el mapa de promedio geométrico*

Zona	Promedio	Desviación estándar	Varianza	Máximo	Área (ha)	% área
1	0.3036	0.0382	0.0015	0.3332	153943.97	4.21
2	0.2526	0.0409	0.0017	0.3280	141523.48	3.87
3	0.2383	0.0609	0.0037	0.3328	85982.33	2.35
4	0.2745	0.0742	0.0055	0.3332	28275.01	0.77
5	0.2589	0.0677	0.0046	0.3332	17093.40	0.47
6	0.2214	0.0623	0.0039	0.3127	12070.33	0.33
7	0.2179	0.0465	0.0022	0.3172	23188.54	0.63
8	0.2232	0.0520	0.0027	0.3228	17243.26	0.47
9	0.2358	0.0914	0.0084	0.3307	11297.00	0.31
10	0.2588	0.0284	0.0008	0.3148	7812.20	0.21
11	0.2763	0.0653	0.0043	0.3323	19538.85	0.53
Área total					517968.37	14.18
Área total zona de estudio					3652407.1	100

Caso con promedio aritmético: la zona con mejores resultados corresponde a la (10), con la menor desviación estándar (0.0284) y varianza (0.0008). Su promedio es cercano al de una distribución normal estándar (0.9262) y tiene valores máximos de 98.15%. La segunda zona con mejores resultados de Figura 29 corresponde a la sección (1), con el promedio más alto (0.9701), el segundo valor más bajo de varianza (0.0016) y desviación estándar (0.0398), además tiene los máximos valores de probabilidad de ocurrencia tipo pórfido (99.98%). En orden de mayor a menor probabilidad máxima las zonas se organizan de la siguiente manera: 1, 4, 5, 3, 11, 9, 2, 8, 7, 10 y 6. En orden de menor a mayor desviación estándar las zonas se organizan así: 10, 1, 2, 7, 8, 11, 3, 6, 4, 5 y 9.

Caso con promedio geométrico: las zonas con mejores resultados corresponden a la (10) y (1). En el primer caso la desviación estándar es de 0.0284 y, en el segundo de 0.0382. Dado que en este tipo de promedio la interpretación tiene que ver, sobre todo, con la geometría, se discriminó con base en las zonas donde los tres modelos convergieron con valores de alta o baja probabilidad. De esta manera, la zona (3), (6), (7), (8) y (9) ya no se tienen en cuenta como secciones prospectivas. En cambio, sobresalen los polígonos (1), (11) y parcialmente el (2). Si se comparan con la estadística zonal de la Tabla 20 teniendo en cuenta únicamente estas tres zonas, en orden de menor a mayor desviación estándar se organizan así: 1, 2 y 11. En orden de mayor a menor varianza se organizan así: 1, 2 y 11. Es decir, con el promedio geométrico se redujeron de 11 a 3 zonas prospectivas que son las de mayor probabilidad en todos los modelos de aprendizaje automático que se aplicaron.

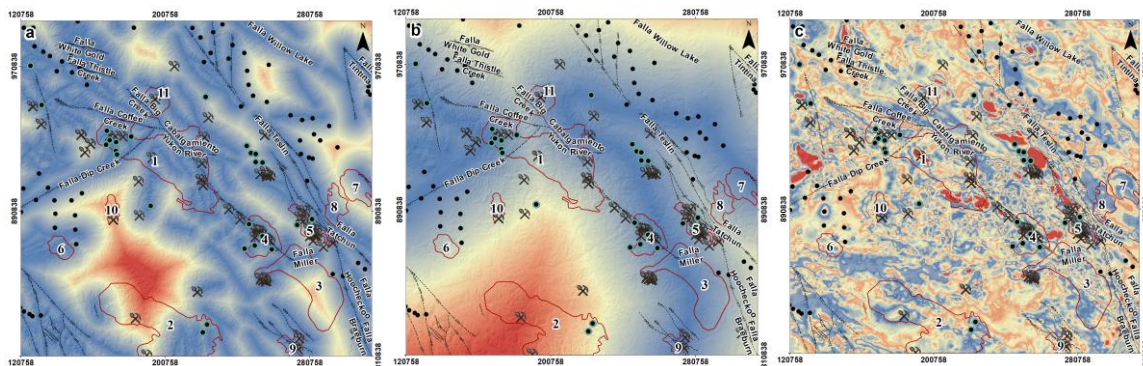
De esta forma, las secciones (1), (2), (4), (10) Y (11) constituyen las más confiables en términos estadísticos. Por consiguiente, para complementar el objetivo de predecir ocurrencias tipo pórfido usando técnicas de aprendizaje automático, se realizó un análisis de los polígonos obtenidos en la Figura 29 y Figura 30 con las variables de entrada definidas X3-X11.

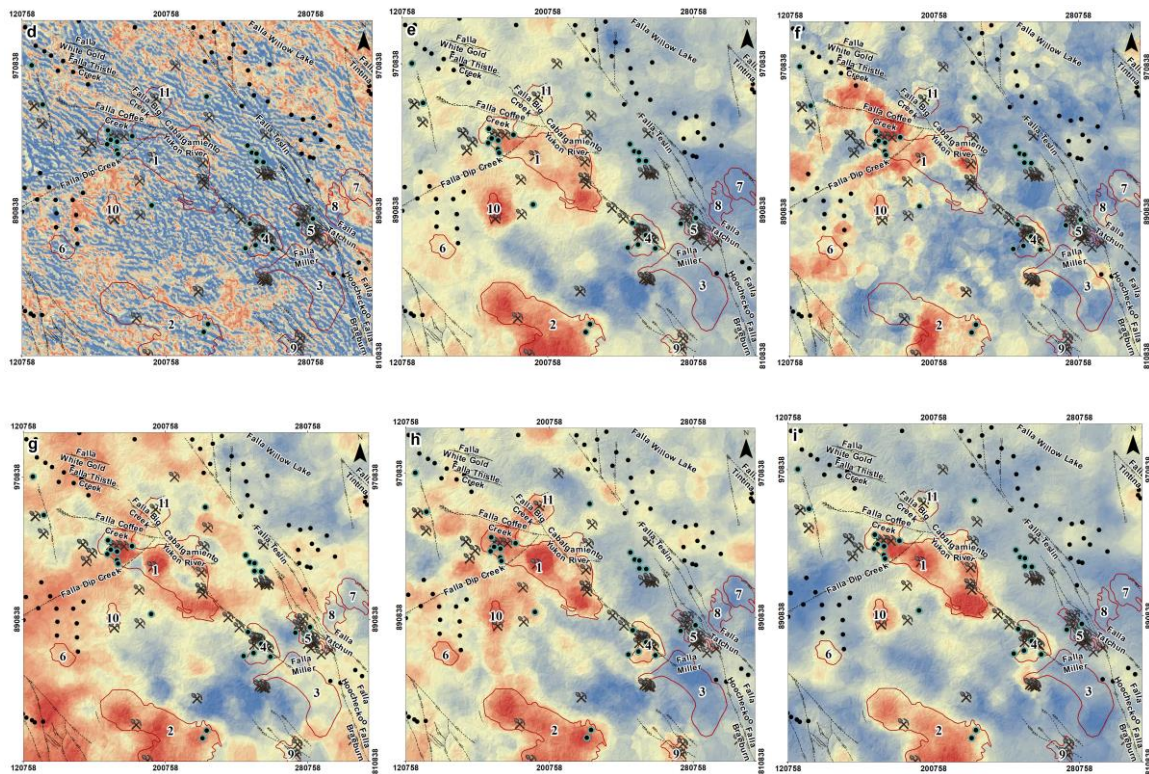
Dado que se requiere unificar la información geológica, geoquímica y geofísica para discriminar los objetivos de exploración y verificar que la información obtenida con los algoritmos computacionales tenga coherencia con la presencia de las ocurrencias minerales en la zona de estudio, se hará un análisis de las estructuras, tipos de roca, información de reportes de exploración mineral, entre otros, con las zonas delimitadas prospectivas.

5.3.5.4 Interpretación de las zonas prospectivas con la información geológica, geoquímica y geofísica

Figura 31

Relación de las variables de entrada con los polígonos de probabilidad >80%. a) X3, b) X4, c) X5, d) X6, e) X7, f) X8, g) X9, h) X10, i) X11.





En la Figura 31 (Apéndice G) se destacan tres factores importantes: en primer lugar, las secciones delineadas con color rojo están, mayormente, cerca de las fallas geológicas y contactos de terrenos litotectónicos exceptuando el polígono (2) en la parte sur. Además, poseen una tendencia NW-SE que tiene sentido lógico con las fallas en la zona de estudio de dirección NW. Las estructuras secundarias como venas y diques se hospedan en estas fallas geológicas. Por ejemplo, la Falla Teslin y Big Creek.

En segundo lugar, los polígonos (1), parcialmente el (2) y el (6) tienen valores altos-muy altos del campo total magnético residual, lo cual es concordante con las zonas de alteración de las ocurrencias minerales. Estas anomalías geofísicas están bien documentadas en la literatura, como el caso del depósito Moly Zone, cuyos valores altos están asociados con pórfidos de cuarzomonzonita, altos contenidos de sulfuros y silicificación; estas permanecen abiertas tanto al

W como al N del depósito, y según las interpretaciones de la base de datos del Servicio Geológico de Yukón, pueden correlacionarse con los contactos mapeados del Pórfido Tad del Cretácico Superior.

Otras secciones como la (8) y la (7) no muestran necesariamente una relación con estos valores de campo total magnético residual. Por otro lado, la capa de primera derivada del campo total magnético residual parece tener valores bajos en la mayoría de los polígonos delineados, exceptuando los (7), (8) y (10).

En tercer lugar, las anomalías geoquímicas Zn, Au, Cu, Mo y Pb tienen una relación directa con los resultados obtenidos con aprendizaje automático. Los polígonos (1), (2) y (6) están prácticamente delimitados por las concentraciones de Mo, Pb y Zn. Sobre todo, se destacan los altos valores de Pb que discriminan las demás áreas de interés. Es decir, esta variable influye de manera importante en las decisiones del algoritmo computacional, pues delimita zonas de alta probabilidad donde, incluso, no hay depósitos (zona 6). El polígono (4) también fue marcado con base en estas características geoquímicas exceptuando las concentraciones de Cu, donde no hay fuerte correlación.

Se resalta que algunos depósitos las áreas de muy alta probabilidad están mayormente sobre altos topográficos, lo cual se ha documentado en varios reportes de exploración, por ejemplo, para el depósito Casino, ubicado en la zona (1) en el cual se considera que las mineralizaciones se desarrollan mejor donde la oxidación de sulfuros de cobre ocurre por encima del nivel freático, es decir, en pendientes bien drenadas.

Finalmente, con el fin de integrar el procedimiento para generar mapas de prospectividad mineral teniendo en cuenta la creación y procesamiento de una base de datos geológica,

geoquímica y geofísica, así como el análisis de las variables de entrada para la construcción de modelos de aprendizaje automático y, posteriormente, la creación de mapas de probabilidad, se plantea y/o unifica una metodología generalizable para cualquier tipo de depósito mineral y algoritmo computacional enfocado en predecir zonas con ocurrencia mineral.

5.3.5.5 Metodología propuesta para la creación de mapas de prospectividad mineral usando técnicas de aprendizaje automático

La metodología propuesta en este trabajo de investigación consta de tres etapas que se ilustran en la Figura 32:

Primera etapa: consiste en la creación y análisis de la base de datos, incluyendo el debido preprocesamiento y la asignación de los puntos no-depósito, teniendo en cuenta las técnicas propuestas en la literatura tales como la selección aleatoria, manual mediante puntos de zonas desfavorables y otros depósitos que no representen características geológicas similares al tipo que se está estudiando. Posteriormente, con base en el análisis de las variables se discriminan aquellas que no tengan una distribución estadística favorable. Lo anterior teniendo en cuenta que, estos datos pueden usarse en el análisis final de zonas de probabilidad, pero, no necesariamente, en el entrenamiento de los algoritmos computacionales. Seguido de esto, es necesario garantizar que las variables de entrada no estén correlacionadas para que la superficie de decisión entre las clases depósito y no-depósito pueda ser calculada. En caso de que estén correlacionadas se debe aplicar una técnica de reducción de dimensionalidad, como por ejemplo el *PCA*. En caso de que no, se procede directamente a extraer los puntos de entrenamiento.

Segunda etapa: consiste en la creación, entrenamiento y validación de los modelos de aprendizaje automático en Python, usando las librerías TensorFlow, Keras y scikit-learn. Una vez

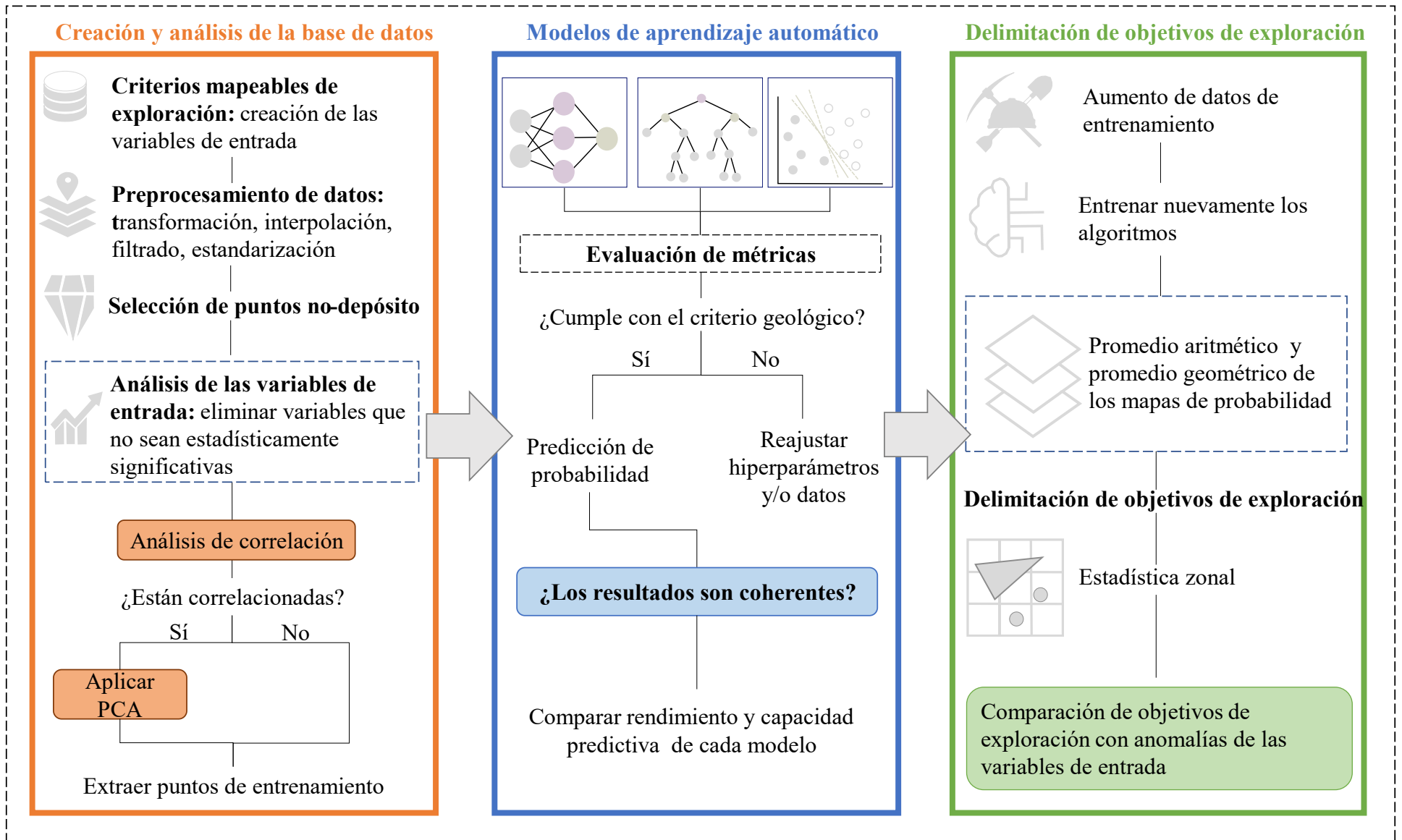
los resultados obtenidos sean acordes con el criterio geológico, se procede a realizar la predicción de probabilidad. En caso de que no, se deben reajustar los hiperparámetros del modelo (tuning, validación cruzada, ensayo y error). De otra forma, se debe re-evaluar la naturaleza de las variables explicatorias. Luego se compara el rendimiento y capacidad predictiva de cada uno de los modelos para verificar que las zonas de alta probabilidad sean consistentes con los puntos conocidos de ocurrencias minerales y, las zonas de baja probabilidad concuerden con los puntos marcados como no-depósitos y/o zonas con características geológicas desfavorables para la formación de estos.

Tercera etapa: consiste en delimitar los objetivos de exploración. Con el fin de refinar las zonas obtenidas en la predicción, se propone realizar un aumento del conjunto de datos de entrenamiento teniendo en cuenta una revisión detallada de la literatura de la zona de estudio para no introducir sesgo en el modelo. Finalmente se obtienen los nuevos mapas de probabilidad.

Por último, se propone realizar dos operaciones entre los mapas con el fin de delimitar las zonas de manera más precisa. Se calcula un promedio aritmético y un promedio geométrico para definir detalladamente los objetivos de exploración con mayor confiabilidad. Es importante, realizar una estadística zonal de los polígonos trazados con el fin de comparar su distribución. Los resultados se comparan con la información de la literatura, por ejemplo, reportes de exploración minera, descripción de perforaciones, anomalías geoquímicas, roca caja, extensión de alteraciones hidrotermales, entre otros.

Figura 32

Metodología propuesta para la creación de mapas de prospectividad mineral con técnicas de aprendizaje automático



6 Conclusiones

Se realizó una revisión sistemática de los criterios relevantes que contribuyen a la formación de depósitos minerales tipo pórfido y al estado de arte en la predicción mineral con técnicas de aprendizaje automático, y se establecieron once mapas predictores que corresponden a: la proximidad a las rocas plutónicas, rocas volcánicas, fallas geológicas, contactos de terrenos litotectónicos, valores del campo total magnético residual y primera derivada vertical del campo magnético residual, concentración geoquímica de Zn, Au, Cu, Mo y Pb. El análisis de estas variables consistió en el cálculo de parámetros estadísticos como la media, desviación estándar y varianza, lo que permitió eliminar la variable X1 del conjunto de datos de entrenamiento. También se calculó el coeficiente de correlación de Pearson que demostró que los mapas predictores estaban altamente correlacionados, por lo cual se aplicó la técnica de reducción de dimensionalidad de análisis de componentes principales (*PCA*) que removi6 la colinealidad. Los nuevos mapas predictores, llamados componentes principales, solamente se tuvieron en cuenta para la predicción de probabilidad, ya que estos no tienen una interpretación geológica. Con esto se dio por cumplido el objetivo 1.

Se usaron los algoritmos computacionales de redes neuronales artificiales, bosques aleatorios y máquinas de soporte vectorial que por excelencia pueden resolver problemas de clasificación mediante el trazo de una superficie de decisión. Las diferencias de esta división se evidenciaron en el mapa de probabilidad, pues los modelos de *ANN* y *RF* calcularon probabilidades porque su superficie de decisión puede ser curva, mientras que para *SVM* está representada por una línea recta, por eso este método solamente calculó valores de 0 y 1 que corresponden a cada clase objetivo. La capacidad predictiva de los tres modelos de aprendizaje automático se evaluó

con las métricas de precisión, función de pérdida, ROC, AUC y matriz de confusión. En términos de precisión se obtuvieron valores de 94.59 % para *ANN*, 83.10 % para *RF* y 94.59 % para *SVM*. Debido a que los algoritmos convergieron a un error mínimo se realizó la predicción de probabilidad en la zona de estudio y se definieron los objetivos de exploración. Con esto se dio por cumplido el objetivo 2.

Se propuso una metodología para detectar y refinar la relación entre las características geológicas y la selección precisa de prospectos de exploración con base en los mapas de probabilidad. La metodología consistió en tres etapas: primero, la creación y el análisis de la base de datos. Segundo, el entrenamiento de los modelos de aprendizaje automático que son usados para resolver tareas de clasificación y tercero, la delimitación de objetivos de exploración aplicando un aumento en el conjunto de datos de entrenamiento y dos operaciones entre bandas para detallar las zonas más prospectivas (promedio aritmético y promedio geométrico). Este paso incluyó el cálculo de la estadística zonal de ambos mapas resultantes. Con esto se redujeron de 11 a 4 zonas prospectivas que tienen probabilidades hasta del 98.96%. Los resultados obtenidos tienen coherencia con la información geológica reportada y el inventario de depósitos conocido. Con esto se concluye que la predicción de ocurrencias minerales tipo pórfido a partir de la integración de información geológica, geoquímica y geofísica usando técnicas de aprendizaje automático es una herramienta capaz de generar zonas con alta probabilidad de depósitos minerales que puede ser implementada por las organizaciones privadas y estatales de Colombia y del mundo, como un elemento clave en el diagnóstico y discriminación de objetivos de exploración con el fin de reducir los costos tanto económicos como operacionales con alta confiabilidad. Con esto se dio por cumplido el objetivo 3.

7 Recomendaciones y trabajo futuro

Se recomienda aplicar otras técnicas de reducción de dimensionalidad o remoción de multicolinealidad, como por ejemplo las redes autoencoders, con fines comparativos para evaluar la influencia de los nuevos mapas predictores en el cálculo de la probabilidad y, por lo tanto, del mapa de prospectividad mineral, lo cual puede contribuir a una mejora en el flujo de trabajo propuesto en este trabajo de investigación.

Se propone aplicar redes neuronales convolucionales para extraer las características más representativas de cada variable geológica, teniendo en cuenta que estas representan mapas predictores con ocurrencias minerales conocidas que no discriminan solo puntos de depósito y no-depósito, sino áreas que cumplen con ciertas condiciones geológicas que permiten la formación de depósitos minerales y, estas redes, a diferencia de las usadas en este trabajo, sí son capaces de realizar la tarea de clasificación teniendo en cuentas estas consideraciones.

8 Repositorio y acceso a los datos

Los datos usados en este trabajo están protegidos bajo una licencia de acceso libre y se encuentran publicados en la plataforma del Servicio Geológico de Yukón (<https://data.geology.gov.yk.ca/>) y el Gobierno de Canadá (<https://www.canada.ca/en.html>). Esta tesis fue elaborada en su totalidad usando código abierto en el lenguaje de programación Python y programas libres como QGIS. En el siguiente repositorio se pueden encontrar los códigos en Python 3x, así como los insumos geográficos usados <https://github.com/Anagabrielamantilla/MineralPrediction.git>

Referencias Bibliográficas

- Abedi, M., Mostafavi Kashani, S. B., Norouzi, G. H., & Yousefi, M. (2017). A deposit scale mineral prospectivity analysis: A comparison of various knowledge-driven approaches for porphyry copper targeting in Seridune, Iran. *Journal of African Earth Sciences*, 128, 127–146. <https://doi.org/10.1016/j.jafrearsci.2016.09.028>
- Awange, J. L., Paláncz, B., Lewis, R. H., & Völgyesi, L. (2018). *Mathematical Geosciences Hybrid Symbolic-Numeric Methods*.
- Brandmeier, M., Cabrera Zamora, I. G., Nykänen, V., & Middleton, M. (2020). Boosting for Mineral Prospectivity Modeling: A New GIS Toolbox. *Natural Resources Research*, 29(1), 71–88. <https://doi.org/10.1007/s11053-019-09483-8>
- Breiman, L. (2001). Random Forests. *Machine Learning 2001* 45:1, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Burkov, A. (2019). *The hundred-page machine learning book*.
- Camus, F. (2005). THE ANDEAN PORPHYRY SYSTEMS. In *Super Porphyry Copper & Gold Deposits: A Global Perspective* (Vol. 1, pp. 45–63).
- Cardoso-Fernandes, J., Teodoro, A. C., Lima, A., & Roda-Robles, E. (2020). Semi-automatization of support vector machines to map lithium (Li) bearing pegmatites. *Remote Sensing*, 12(14). <https://doi.org/10.3390/rs12142319>
- Carranza, E. J. (2009). *Chapter 8: Data-Driven Modeling of Mineral Prospectivity* (pp. 249–310). [https://doi.org/10.1016/s1874-2734\(09\)70012-9](https://doi.org/10.1016/s1874-2734(09)70012-9)

- Carranza, E. J. M. (2009). Objective selection of suitable unit cell size in data-driven modeling of mineral prospectivity. *Computers and Geosciences*, 35(10), 2032–2046. <https://doi.org/10.1016/j.cageo.2009.02.008>
- Carranza, E. J. M., & Laborte, A. G. (2015a). Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Computers and Geosciences*, 74, 60–70. <https://doi.org/10.1016/j.cageo.2014.10.004>
- Carranza, E. J. M., & Laborte, A. G. (2015b). Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Computers and Geosciences*, 74, 60–70. <https://doi.org/10.1016/j.cageo.2014.10.004>
- Carranza, E. J. M., & Laborte, A. G. (2015c). Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Computers and Geosciences*, 74, 60–70. <https://doi.org/10.1016/j.cageo.2014.10.004>
- Chen, Y., & Wu, W. (2017). Mapping mineral prospectivity using an extreme learning machine regression. *Ore Geology Reviews*, 80, 200–213. <https://doi.org/10.1016/j.oregeorev.2016.06.033>
- Colpron, M. (compiler). (2006). *Tectonic assemblage map of Yukon-Tanana and related terranes in Yukon and northern British Columbia (1:1 000 000 scale)*. Yukon Geological Survey, Open File 2006-1.

- Daviran, M., Maghsoudi, A., Ghezelbash, R., & Pradhan, B. (2021). A new strategy for spatial predictive mapping of mineral prospectivity: Automated hyperparameter tuning of random forest approach. *Computers and Geosciences*, 148. <https://doi.org/10.1016/j.cageo.2021.104688>
- Farnham, B., Tokyo, S., Boston, B., Sebastopol, F., & Beijing, T. (n.d.). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems SECOND EDITION*.
- Fu, G., Lü, Q., Yan, J., Farquharson, C. G., Qi, G., Zhang, K., Zhang, Y., Wang, H., & Luo, F. (2021). 3D mineral prospectivity modeling based on machine learning: A case study of the Zhuxi tungsten deposit in northeastern Jiangxi Province, South China. *Ore Geology Reviews*, 131. <https://doi.org/10.1016/j.oregeorev.2021.104010>
- Ghezelbash, R., Maghsoudi, A., & Carranza, E. J. M. (2019). Performance evaluation of RBF- and SVM-based machine learning algorithms for predictive mineral prospectivity modeling: integration of S-A multifractal model and mineralization controls. *Earth Science Informatics*, 12(3), 277–293. <https://doi.org/10.1007/s12145-018-00377-6>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. <https://www.deeplearningbook.org/>
- Goyes-Peñafiel, P., & Hernandez-Rojas, A. (2021). Landslide susceptibility index based on the integration of logistic regression and weights of evidence: A case study in

- Popayan, Colombia. *Engineering Geology*, 280.
<https://doi.org/10.1016/j.enggeo.2020.105958>
- Gregorich, M., Strohmaier, S., Dunkler, D., & Heinze, G. (2021). Regression with Highly Correlated Predictors: Variable Omission Is Not the Solution. *International Journal of Environment Research and Public Health*, 1–12.
- Hagemann, S. G., Lisitsin, V. A., & Huston, D. L. (2016). Mineral system analysis: Quo vadis. In *Ore Geology Reviews* (Vol. 76, pp. 504–522). Elsevier B.V.
<https://doi.org/10.1016/j.oregeorev.2015.12.012>
- 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(4), 489–507. <https://doi.org/10.1007/s11053-017-9335-6>
- Harris, J. R., Grunsky, E., Behnia, P., & Corrigan, D. (2015). Data- and knowledge-driven mineral prospectivity maps for Canada's North. *Ore Geology Reviews*, 71, 788–803. <https://doi.org/10.1016/j.oregeorev.2015.01.004>
- Hronsky, J. M. A., & Groves, D. I. (2008). Science of targeting: Definition, strategies, targeting and performance measurement. *Australian Journal of Earth Sciences*, 55(1), 3–12. <https://doi.org/10.1080/08120090701581356>
- Jolliffe, I. T. (1987). *Principal Component Analysis*. Springer.
- Keras: the Python deep learning API*. (n.d.). Retrieved April 8, 2022, from <https://keras.io/>

- Keykhay-Hosseinpour, M., Kohsary, A. H., Hossein-Morshedy, A., & Porwal, A. (2020a). A machine learning-based approach to exploration targeting of porphyry Cu-Au deposits in the Dehsalm district, eastern Iran. *Ore Geology Reviews*, 116. <https://doi.org/10.1016/j.oregeorev.2019.103234>
- Keykhay-Hosseinpour, M., Kohsary, A. H., Hossein-Morshedy, A., & Porwal, A. (2020b). A machine learning-based approach to exploration targeting of porphyry Cu-Au deposits in the Dehsalm district, eastern Iran. *Ore Geology Reviews*, 116. <https://doi.org/10.1016/j.oregeorev.2019.103234>
- Kost, S., Rheinbach, O., & Schaeben, H. (2021). Using logistic regression model selection towards interpretable machine learning in mineral prospectivity modeling. *Geochemistry*, 125826. <https://doi.org/10.1016/j.chemer.2021.125826>
- Kreuzer, O. P., Miller, A. V. M., Peters, K. J., Payne, C., Wildman, C., Partington, G. A., Puccioni, E., McMahon, M. E., & Etheridge, M. A. (2015). Comparing prospectivity modelling results and past exploration data: A case study of porphyry Cu-Au mineral systems in the Macquarie Arc, Lachlan Fold Belt, New South Wales. *Ore Geology Reviews*, 71, 516–544. <https://doi.org/10.1016/j.oregeorev.2014.09.001>
- Lachaud, A., Marcus, A., Vučetić, S., & Mišković, I. (2021a). Study of the influence of non-deposit locations in data-driven mineral prospectivity mapping: A case study on the Iskut Project in Northwestern British Columbia, Canada. *Minerals*, 11(6). <https://doi.org/10.3390/min11060597>

- Lachaud, A., Marcus, A., Vučetić, S., & Mišković, I. (2021b). Study of the influence of on-deposit locations in data-driven mineral prospectivity mapping: A case study on the iskut project in northwestern british columbia, canada. *Minerals*, 11(6). <https://doi.org/10.3390/min11060597>
- Leybourne, M. I. (2007). Aqueous geochemistry in mineral exploration. In *Mineral Deposits of Canada: A Synthesis of Major Deposit-Types, District Metallogeny, the Evolution of Geological Provinces, and Exploration Methods: Geological Association of Canada, Mineral Deposits Division, Special Publication* (Issue 5, pp. 1007–1033).
- Li, T., Xia, Q., Zhao, M., Gui, Z., & Leng, S. (2020a). Prospectivity Mapping for Tungsten Polymetallic Mineral Resources, Nanling Metallogenic Belt, South China: Use of Random Forest Algorithm from a Perspective of Data Imbalance. *Natural Resources Research*, 29(1), 203–227. <https://doi.org/10.1007/s11053-019-09564-8>
- Li, T., Xia, Q., Zhao, M., Gui, Z., & Leng, S. (2020b). Prospectivity Mapping for Tungsten Polymetallic Mineral Resources, Nanling Metallogenic Belt, South China: Use of Random Forest Algorithm from a Perspective of Data Imbalance. *Natural Resources Research*, 29(1), 203–227. <https://doi.org/10.1007/s11053-019-09564-8>
- Lin, N., Chen, Y., Liu, H., & Liu, H. (2021). A comparative study of machine learning models with hyperparameter optimization algorithm for mapping mineral prospectivity. *Minerals*, 11(2), 1–31. <https://doi.org/10.3390/min11020159>

- Maepa, F., & Smith, R. (2017). *Predictive mapping of the gold mineral potential in the Swayze Greentone Belt, ON, Canada.* 2456–2460. <https://doi.org/10.1190/segam2017-17660409.1>
- Maepa, F., & Smith, R. S. (2019). Radial Basis Function Link Nets method for predicting gold mineral potential from geological and geophysical data in the Swayze greenstone belt (SGB). *2018 SEG International Exposition and Annual Meeting, SEG 2018*, 1848–1852. <https://doi.org/10.1190/segam2018-2995704.1>
- Maepa, F., Smith, R. S., & Tessema, A. (2021). Support vector machine and artificial neural network modelling of orogenic gold prospectivity mapping in the Swayze greenstone belt, Ontario, Canada. *Ore Geology Reviews*, 130. <https://doi.org/10.1016/j.oregeorev.2020.103968>
- Mao, X., Zhang, W., Liu, Z., Ren, J., Bayless, R. C., & Deng, H. (2020). 3D mineral prospectivity modeling for the low-sulfidation epithermal gold deposit: A case study of the axi gold deposit, western Tianshan, NW China. *Minerals*, 10(3). <https://doi.org/10.3390/min10030233>
- Mccuaig, T. C., Beresford, S., & Hronsky, J. (2010). Translating the mineral systems approach into an effective exploration targeting system. *Ore Geology Reviews*, 38(3), 128–138. <https://doi.org/10.1016/j.oregeorev.2010.05.008>
- McMillan, M., Haber, E., Peters, B., & Fohring, J. (2021). Mineral prospectivity mapping using a VNet convolutional neural network. *Leading Edge*, 40(2), 99–105. <https://doi.org/10.1190/tle40020099.1>

- McMillan, W. J. (1991). Porphyry deposits in the Canadian Cordillera. In *Ore deposits, tectonics and metallogeny in the Canadian Cordillera: British Columbia* (pp. 253–276). Ministry of Energy, Mines and Petroleum Resources.
- Mihalasky, M. J., Bookstrom, A. A., Frost, T. P., Ludington, S., with contributions from James, L., Panteleyev, A., & Abbott, G. (2011). *Porphyry Copper Assessment of British Columbia and Yukon Territory, Canada: Scientific Investigations Report 2010-5090-C*.
- Mindrila, D., & Balentyne, P. (2017). *Scatterplots and Correlation*. https://www.westga.edu/academics/research/vrc/assets/docs/scatterplots_and_correlation_notes.pdf
- Montsion, R. M., Saumur, B. M., Acosta-Gongora, P., Gadd, M. G., Tschirhart, P., & Tschirhart, V. (2019). Knowledge-driven mineral prospectivity modelling in areas with glacial overburden: porphyry Cu exploration in Quesnellia, British Columbia, Canada. *Applied Earth Science: Transactions of the Institute of Mining and Metallurgy*, 128(4), 181–196. <https://doi.org/10.1080/25726838.2019.1675403>
- Mortensen, J. K. (1992). Pre-Mid-Mesozoic tectonic evolution of the Yukon-Tanana Terrane, Yukon and Alaska. *Tectonics*, 11(4), 836–853. <https://doi.org/10.1029/91TC01169>
- Mortensen, J. K., & Friend, M. (2021). An overview of porphyry style deposits in Yukon. In Sharman, Lang, & Chapman (Eds.), *Porphyry Deposits of the Northwestern Cordillera of North America: A 25-Year Update* (Vol. 57). Canadian

Institute of Mining and Metallurgy (CIM).

<https://www.researchgate.net/publication/347622588>

Nelson, J., & Colpron, M. (2007). TECTONICS AND METALLOGENY OF THE BRITISH COLUMBIA, YUKON AND ALASKAN CORDILLERA, 1.8 GA TO THE PRESENT. In *Mineral Deposits of Canada: A Synthesis of Major Deposit-Types, District Metallogeny, the Evolution of Geological Provinces, and Exploration Methods* (pp. 755–791). Geological Association of Canada, Mineral Deposits Division.

Nelson, J. L., Colpron, M., Piercey, S., Dusel-Bacon, C., & Murphy, D. (2006). Paleozoic tectonic and metallogenic evolution of pericratonic terranes in Yukon, northern British Columbia and eastern Alaska. In *Paleozoic Evolution and Metallogeny of Pericratonic Terranes at the Ancient Pacific Margin of North America, Canadian and Alaskan Cordillera* (Special Paper 45, pp. 323–360). Geological Association of Canada.
<https://www.researchgate.net/publication/257269093>

Nykänen, V., Lahti, I., Niiranen, T., & Korhonen, K. (2015). Receiver operating characteristics (ROC) as validation tool for prospectivity models - A magmatic Ni-Cu case study from the Central Lapland Greenstone Belt, Northern Finland. *Ore Geology Reviews*, 71, 853–860. <https://doi.org/10.1016/j.oregeorev.2014.09.007>

pandas - Python Data Analysis Library. (2022, April). <https://pandas.pydata.org/>

pandas documentation — pandas 1.4.2 documentation. (n.d.). Retrieved April 19, 2022, from <https://pandas.pydata.org/docs/>

- Prado, E. M. G., de Souza Filho, C. R., Carranza, E. J. M., & Motta, J. G. (2020). Modeling of Cu-Au prospectivity in the Carajás mineral province (Brazil) through machine learning: Dealing with imbalanced training data. *Ore Geology Reviews*, 124. <https://doi.org/10.1016/j.oregeorev.2020.103611>
- Raileanu, L. E., & Stoffel, K. (2004). Theoretical comparison between the Gini Index and Information Gain criteria. *Annals of Mathematics and Artificial Intelligence*, 41(1), 77–93. <https://doi.org/10.1023/B:AMAI.0000018580.96245.C6>
- Richards, J. P. (2003). Tectono-Magmatic Precursors for Porphyry Cu-(Mo-Au) Deposit Formation. *Economic Geology*, 98(8), 1515–1533. <https://doi.org/10.2113/98.8.1515>
- Richards, J. P. (2009). Postsubduction porphyry Cu-Au and epithermal Au deposits: Products of remelting of subduction-modified lithosphere. *Geology*, 37(3), 247–250. <https://doi.org/10.1130/G25451A.1>
- Rodriguez-Galiano, V. F., Chica-Olmo, M., & Chica-Rivas, M. (2014). Predictive modelling of gold potential with the integration of multisource information based on random forest: a case study on the Rodalquilar area, Southern Spain. *International Journal of Geographical Information Science*, 28(7), 1336–1354. <https://doi.org/10.1080/13658816.2014.885527>
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector

machines. *Ore Geology Reviews*, 71, 804–818.

<https://doi.org/10.1016/j.oregeorev.2015.01.001>

Scholkopf, B., Smola, A., & Muller, K.-R. (1999). Kernel Principal Component Analysis. In *Advances in Kernel Methods - Support Vector Learning* (pp. 327–352).

scikit-learn: machine learning in Python — scikit-learn 1.0.2 documentation. (n.d.).

Retrieved April 19, 2022, from <https://scikit-learn.org/stable/>

Silberling, N. J., Jones, D. L., Coney, P. J., & Plafker, G. (1984). *Lithotectonic terrane maps of the North American Cordillera*.

Sillitoe, R. (2008). Major Gold Deposits and Belts of the North and South American Cordillera. *Economic Geology*, 103, 663–687.

Sinclair, W. D. (2007). Porphyry deposits. In *Mineral Deposits of Canada: A Synthesis of Major Deposit-Types, District Metallogeny, the Evolution of Geological Provinces, and Exploration Methods* (5th ed., pp. 223–243). Geological Association of Canada, Mineral Deposits Division.

<https://www.researchgate.net/publication/228668686>

Singer, D. A., Berger, V. I., & Moring, B. C. (2008). *Porphyry Copper Deposits of the World: Database And Grade and Tonnage Models, 2008*.

<http://www.usgs.gov/pubprod>

Srivastava, N., Hinton, G., Krizhevsky, A., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. In *Journal of Machine Learning Research* (Vol. 15).

- Sun, T., Chen, F., Zhong, L., Liu, W., & Wang, Y. (2019). GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China. *Ore Geology Reviews*, *109*, 26–49. <https://doi.org/10.1016/j.oregeorev.2019.04.003>
- Sun, T., Li, H., Wu, K., Chen, F., Zhu, Z., & Hu, Z. (2020). Data-driven predictive modelling of mineral prospectivity using machine learning and deep learning methods: A case study from Southern Jiangxi Province, China. *Minerals*, *10*(2). <https://doi.org/10.3390/min10020102>
- Tipping, M., & Bishop, C. (1999). Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *61*, 611–622.
- Wheeler, J. O., Brookfield, A. J., Gabrielse, H., Monger, J. W. H., Tipper, H. W., & Woodsworth, G. J. (1991). *Terrane Map of the Canadian Cordillera*, Geological Survey of Canada, Map 1713A, scale 1:2 000 000.
- Wilkinson, J. J. (2013). Triggers for the formation of porphyry ore deposits in magmatic arcs. In *Nature Geoscience* (Vol. 6, Issue 11, pp. 917–925). <https://doi.org/10.1038/ngeo1940>
- Xiang, J., Xiao, K., Carranza, E. J. M., Chen, J., & Li, S. (2020). 3D Mineral Prospectivity Mapping with Random Forests: A Case Study of Tongling, Anhui, China. *Natural Resources Research*, *29*(1), 395–414. <https://doi.org/10.1007/s11053-019-09578-2>
- Zhang, N., Zhou, K., & Du, X. (2017). Application of fuzzy logic and fuzzy AHP to mineral prospectivity mapping of porphyry and hydrothermal vein copper deposits

in the Dananhu-Tousuquan island arc, Xinjiang, NW China. *Journal of African Earth Sciences*, 128, 84–96. <https://doi.org/10.1016/j.jafrearsci.2016.12.011>

Zhang, N., Zhou, K., & Li, D. (2018). Back-propagation neural network and support vector machines for gold mineral prospectivity mapping in the Hatu region, Xinjiang, China. *Earth Science Informatics*, 11(4), 553–566. <https://doi.org/10.1007/s12145-018-0346-6>

Zuo, R. (2017). Machine Learning of Mineralization-Related Geochemical Anomalies: A Review of Potential Methods. In *Natural Resources Research* (Vol. 26, Issue 4, pp. 457–464). Springer New York LLC. <https://doi.org/10.1007/s11053-017-9345-4>

Zuo, R., & Carranza, E. J. M. (2011). Support vector machine: A tool for mapping mineral prospectivity. *Computers and Geosciences*, 37(12), 1967–1975. <https://doi.org/10.1016/j.cageo.2010.09.014>