



UNIVERSIDAD DE TALCA
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA CIVIL DE MINAS

CLASIFICACIÓN DE LITOLOGÍA Y ESTIMACIÓN DE VETAS
EN TESTIGOS DE SONDAJES DIAMANTINO USANDO
MACHINE LEARNING Y VISIÓN COMPUTACIONAL

**MEMORIA PARA OPTAR AL TÍTULO
DE INGENIERO CIVIL DE MINAS**

SEBASTIÁN IGNACIO ARRIAGADA SILVA

PROFESOR GUÍA

Dr. Manuel Rolando Reyes Jara

MIEMBROS DE LA COMISIÓN

Mg. Marcelo Surjan Arellano

Mg. Manuel García Hurtado

CURICÓ - CHILE

2020

CONSTANCIA

La Dirección del Sistema de Bibliotecas a través de su encargado Biblioteca Campus Curicó certifica que el autor del siguiente trabajo de titulación ha firmado su autorización para la reproducción en forma total o parcial e ilimitada del mismo.



Curicó, 2022

RESUMEN

La exploración geológica y la descripción de muestras de corteza terrestre son de suma importancia para la evaluación de yacimientos minerales. Por esta razón, existen múltiples esfuerzos por mejorar la metodología de descripción o de *mapeo de rocas*, siendo los algoritmos de *machine learning* basados en redes neuronales una potencial herramienta para estandarizar, automatizar y agilizar estos procesos. Este trabajo aborda la aplicación de dos tipos de modelos de *machine learning*, con el objetivo de clasificar la litología y estimar la presencia de vetas utilizando fotografías digitales de sondajes tipo diamantino.

El primer tipo de modelo es generado a través del proceso *Automated Machine Learning* (AutoML), el cual adapta modelos de clasificación preexistentes con buenos rendimientos en bases de datos estandarizadas y los optimiza para ser entrenados en una nueva base de datos. La clasificación de imágenes fue utilizada para la inferencia de la litología, del estado geotécnico y de la presencia de vetas. El segundo tipo, correspondiente un modelo *cGAN* (Conditional Generative Adversarial Networks), está compuesto por dos redes neuronales enfrentadas en un juego de suma cero, en donde si uno gana el otro pierde, obteniendo destacados resultados en tareas como la transferencia de estilos en imágenes. Este modelo fue utilizado para la segmentación de mineralización en vetas. Para cuantificar el desempeño de los modelos se utilizaron 5 métricas derivadas de las correspondientes matrices de confusión.

Con respecto a los resultados de la clasificación litológica, el mejor rendimiento se consiguió utilizando la etiqueta binaria correspondiente a “litología de interés” y “otro tipo”, junto con entrenar un modelo diferente especializado en cada estado geotécnico presente (roca fracturada, molida, intacta y semi intacta). Se obtiene en conjunto un acierto del 79% de las veces y un 59% de posibilidades de que al reconocer la litología de interés, esta sea correctamente etiquetada. El mayor rendimiento se obtiene dentro de la roca intacta con 78% y 79% respectivamente. Al analizar los datos clasificados erróneamente según la información obtenida por observación directa (verdad fundamental), se detectan casos que corresponden a potenciales zonas de transición entre la litología de interés y otras. Por lo tanto, si bien aún no se obtiene un nivel de asertividad necesario para automatizar por completo el proceso de *mapeo*, esta metodología presenta una potencial utilidad para realizar un perfil coherente de cada litología y detectar errores dentro de *mapeos* anteriores.

Con respecto a la estimación de vetas, al segmentar las zonas de mineralización observable en las fotografías, es posible estimar su área con una exactitud del 97% y una precisión del 83% con respecto a los datos de prueba segmentados manualmente por un humano. La efectividad de estimación se ve reducida por factores como la variabilidad de tonalidades a lo largo de la roca y la presencia de zonas de alto brillo en la superficie.

Aplicando ambas metodologías, es posible realizar inferencias dentro de nuevos sondajes a una rapidez por sobre 1 metro por segundo. Se recomienda utilizar esta herramienta como apoyo para agilizar el proceso de *mapeo*, pre clasificando y segmentando las zonas con altos niveles de certeza y señalar las zonas con mayor incerteza para la validación manual por parte del equipo de geología.

PALABRAS CLAVES: Machine Learning, Clasificación, Litología, Segmentación, Vetas.

ABSTRACT

Geological exploration and description of earth's crust samples are of prime importance for the evaluation of mineral deposits. For this reason, there are multiple efforts to improve the methodology of description or rock *mapping*, being *machine learning* algorithms based on neural networks a potential tool to standardize, automate and streamline these processes. This work addresses the application of two types of machine learning models, with the objective of classifying lithology and estimating the presence of veins using digital photographs of diamond drill holes.

The first type of model is generated through the *Automated Machine Learning* (AutoML) process, which adapts pre-existing classification models with good performance in standardized databases and optimizes them to be trained on a new database. The image classification was used for inference of lithology, geotechnical state and presence of veins. The second type, corresponding to a *cGAN* (Conditional Generative Adversarial Networks) model, is composed of two neural networks facing each other in a zero-sum game, where if one wins the other loses, obtaining outstanding results in tasks such as the transfer of styles in images. This model was used for the segmentation of mineralization in veins. To quantify the performance of the models, 5 metrics derived from the corresponding confusion matrices were used.

With respect to the lithological classification results, the best performance was achieved by using the binary label corresponding to "lithology of interest" and "other type", together with training a different model specialized in each geotechnical state present (fractured, ground, intact and semi-intact rock). Overall a 79% success rate is obtained and a 59% chance that when recognizing the lithology of interest, it is correctly labeled. The highest performance is obtained within intact rock with 78% and 79% respectively. When analyzing the misclassified data according to the information obtained by direct observation (ground truth), cases are detected that correspond to potential transition zones between the lithology of interest and others. Therefore, although a level of assertiveness necessary to fully automate the *mapping* process is not yet obtained, this methodology presents a potential utility to make a coherent profile of each lithology and to detect errors within previous *mappings*.

With respect to vein estimation, by segmenting the zones of observable mineralization in the photographs, it is possible to estimate their area with an accuracy of 97% and an accuracy of 83% with respect to test data manually segmented by a human. The estimation effectiveness is reduced by factors such as the variability of shades along the rock and the presence of high brightness zones on the surface.

By applying both methodologies, it is possible to make inferences within new drillholes at a speed of over 1 meter per second. It is recommended to use this tool as a support to speed up the *mapping* process, pre-classifying and segmenting the zones with high levels of certainty and pointing out the zones with higher uncertainty for manual validation by the geology team.

KEY WORDS: Machine Learning, Classification, Lithology, Segmentation, Veins.

**Con todo mi corazón a mi madre,
este logro es el fruto del amor que me ha entregado**

INDICE

1.	INTRODUCCIÓN	1
1.1	Descripción del problema.....	3
1.2	Solución propuesta	3
1.3	Objetivos	3
1.3.1	Objetivo general	3
1.3.2	Objetivos específicos.....	3
1.4	Alcances	4
1.5	Metodologías y herramientas utilizadas	4
1.5.1	Preprocesamiento de datos	5
1.5.2	Diferenciación litológica	5
1.5.3	Detección y estimación de porcentaje de vetas	5
1.5.4	Predicción en datos de Prueba.....	5
1.6	Resultados esperados.....	6
1.7	Organización del documento.....	6
2.	Desarrollo.....	7
2.1	Marco teórico	7
2.1.1	exploración de yacimientos	7
2.1.2	Evaluación de yacimientos.....	11
2.1.3	Inteligencia artificial	13

2.1.4	Softwares y herramientas	21
2.2	Caso de estudio.....	22
2.2.1	Contexto geológico	23
2.2.2	Análisis de datos.....	26
2.3	Procedimiento experimental.....	27
2.3.1	Pretratamiento de datos	27
2.3.2	Confección de los modelos para la clasificación litológica.....	28
2.3.3	Confección de modelos para la detección y estimación de vetillas.....	30
2.3.4	Predicción en sondajes de prueba.....	31
3.	Resultados y Análisis de resultados	33
3.1	Predicción de litología.....	33
3.1.1	Etiqueta multi litología (modelo 1.1)	33
3.1.2	Etiqueta binaria (modelo 1.2).....	34
3.1.3	Etiqueta binaria según estado geotécnico.....	34
3.2	Detección y estimación de vetas	35
3.2.1	Detección de vetas (modelo 2)	35
3.2.2	Cálculo de porcentaje de vetas (modelo 3)	36
3.3	Reanálisis de la base de datos	36
3.4	Resultado final.....	38
4.	CONCLUSIONES	40
4.1	Clasificación litológica.....	40

4.2	Detección y estimación de vetas	41
5.	REFERENCIAS	43

<i>Tabla 1 Ejemplo de la predicción de litología y vetas</i>	<i>6</i>
<i>Tabla 2 Litologías presentes dentro de los sondeos utilizados dentro de este trabajo.</i>	<i>23</i>
<i>Tabla 3 Alteraciones presentes dentro de la litología marga.....</i>	<i>25</i>
<i>Tabla 4 Tipos de mineralizaciones presentes dentro de la litología marga.....</i>	<i>25</i>
<i>Tabla 5 Resumen de los resultados obtenidos de los tres tipos de entrenamientos realizados para la clasificación de litología marga.....</i>	<i>33</i>
<i>Tabla 6 Resumen de los resultados obtenidos por cada modelo entrenado según un estado geotécnico</i>	<i>34</i>
<i>Tabla 7 Resumen de los resultados obtenidos en la diferenciación de variaciones dentro de la litología marga</i>	<i>35</i>
<i>Tabla 8 Ejemplo del resultado final obtenido después de unir el resultado de la predicción geotécnica, litológica, variación de litología y segmentación de vetas.....</i>	<i>38</i>

<i>Figura 1 Curva de riesgo de un proyecto minero (Leddy, 2014)</i>	1
<i>Figura 2 Esquema de las clasificaciones de recursos y su correlato en reservas. Fuente: Instituto de Ingenieros de Minas de Perú</i>	12
<i>Figura 3 Esquema del campo de estudio de la inteligencia artificial y sus subáreas. Recuperado de: TerryTaewoongUm, Deep-learning-machine-learning-tutorial-for-beginners</i>	13
<i>Figura 4 Flujo de trabajo estándar para el entrenamiento supervisado de modelos de machine learning (J. Cracknell, 2014)</i>	15
<i>Figura 5 Diagrama de funcionamiento de una red neuronal adversaria (Martinez J. , 2020)</i>	18
<i>Figura 6 Diagrama de funcionamiento de una red neuronal condicional adversaria (Martinez J. , 2020). ...</i>	19
<i>Figura 7 Ejemplo de matriz de confusión con 3 etiquetas (Garcia, 2018)</i>	19
<i>Figura 8 Matriz de casos posibles dentro de cada categoría. Recuperado de: https://rpubs.com/chzelada/275494</i>	20
<i>Figura 9 Recorte del mismo sitio de la caja de sondajes en fotografías con distinta tonalidad de luz</i>	27
<i>Figura 10 Fotografía de referencia con etiquetas de dos tipos de litologías diferentes (Elaboración propia)</i>	28
<i>Figura 11 Fotografía de referencia con etiquetas de dos tipos de litologías separadas en cuadrados uniformes</i>	28
<i>Figura 12 Ejemplos de las categorías del elemento geotécnico presente en la roca</i>	29
<i>Figura 13 Posibles variaciones dentro de la roca marga</i>	30
<i>Figura 14 Segmentación manual de la veta presente dentro de un recorte de roca marga</i>	31
<i>Figura 15 Frecuencia de la confianza de las predicciones de la categoría marga</i>	38

1. INTRODUCCIÓN

El negocio minero es un área de inversión particularmente riesgosa, autores como Naranjo (2005) plantean que además de la necesidad de un enorme capital de inversión y de largos plazos de ejecución, este rubro se destaca por tener una particular y cuantiosa incerteza proveniente de la estimación del yacimiento mineral, lo cual se hace a través de un proceso largo y costoso. Dentro de los esfuerzos por disminuir esta incerteza de estimación, Verdugo (2016) estudió la factibilidad de tomar un mayor número de muestras de la corteza terrestre, concluyendo que, a pesar de sus altos costos asociados, en etapas de alto riesgo como en la exploración (Figura 1) esta inversión se ve recompensada por la disminución del riesgo del proyecto. No obstante, esto será cierto siempre y cuando la descripción de los sondajes o *mapeo* sea correcta.

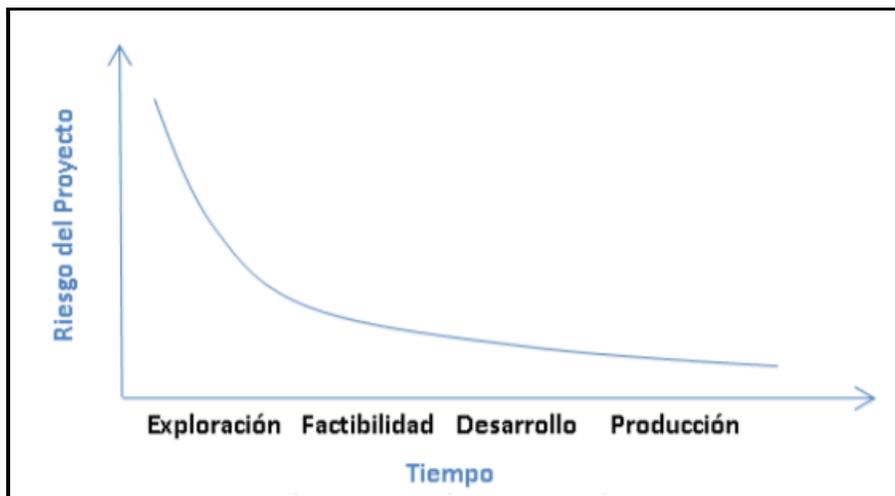


Figura 1 Curva de riesgo de un proyecto minero (Leddy, 2014)

Actualmente el método más efectivo de exploración de yacimientos de minerales es la extracción de testigos de roca a través de sondajes (Gandhi & Sarkar, 2016), sin embargo, la calidad de la información obtenida a través de estas muestras dependerá de la metodología de descripción o *mapeo* que se utilice. Desde los inicios de la exploración geológica, la metodología más empleada es la medición manual y sistemática de la mayor cantidad de información posible, determinando aspectos del yacimiento como la litología, mineralogía, estructuras geotécnicas y zonas de alteración. Por lo tanto, a pesar de que se aumente el número de sondajes, la confianza de la información obtenida tendrá directa relación con la calidad y rigurosidad del *mapeo* realizado por el equipo de geología. Habitualmente esta descripción se complementa con el análisis de muestras puntuales en laboratorios.

Si bien el análisis manual de la roca es una costumbre desde los inicios de la minería, esta metodología tiene importantes desventajas ya que requiere el análisis exhaustivo de un geólogo competente, volviéndolo un proceso lento y dependiente de la experiencia del geólogo (Bonnici, 2008). Junto a esto, si bien se han desarrollado diversas herramientas para el *mapeo* manual, existen dificultades técnicas intrínsecas del trabajo humano, como es el mantener un criterio constante a lo largo del tiempo sin importar de quien lo realice o la incapacidad de estimar rápidamente características cuantitativas.

Debido a esto es posible encontrar diversos esfuerzos para automatizar y estandarizar estos procesos, dentro de los cuales se destaca el uso del análisis hiperespectral en testigos, permitiendo obtener un mapa de píxel preciso de los minerales presentes a través de su espectro electromagnético característico (Martínez, Voisin, & McFarlane, 2015; Contreras y otros, 2018; Contreras y otros, 2020). Sin embargo, los altos costos asociados al uso de esta tecnología hacen que su aplicación sea acotada a zonas del yacimiento de gran interés, a yacimientos con minerales particularmente valiosos o a proyectos de gran tamaño. Los costos de inversión cobran mayor importancia si consideramos que según Sernageomin (2019), en Chile la pequeña y mediana minería extrae desde el 4 hasta el 43% de la producción metálica dependiendo del elemento, empresas que a su vez, suelen tener una capacidad de inversión baja en comparación a la gran minería.

Debido a todo lo planteado anteriormente, es que el uso de inteligencia artificial se ha masificado dentro de los últimos años como una herramienta prometedora, ya que ha demostrado obtener buenos resultados en distintas tareas sin requerir grandes costos de inversión o implementación. Algunos ejemplos de la aplicación de estos algoritmos en el *mapeo* son, el uso de información geofísica y el escaneo de las paredes durante la realización del sondaje para predecir las características de la roca (Merembayev, Yunussov y Yedilkhan, 2018; Kexiong y Laibin, 2008; Singh, Seol, & Myshakin, 2020; Caté y otros, 2017; Chen y otros, 2020), la interpretación del espectro infrarrojo de ondas largas para la predicción de la abundancia del mineral dentro del yacimiento (Barker y otros, 2020) y la clasificación de rocas y minerales a través de imágenes digitales, sobre lo cual se tratará este trabajo.

Dentro de los estudios de la clasificación de rocas a través de imágenes digitales se suele hacer énfasis en la textura, ya que además de ser una característica muy estudiada dentro del tratamiento de imágenes, a su vez, tiene un importante correlato con significado geológico. De esta forma, es posible diferenciar los distintos tipos de rocas a través de métodos como el análisis de la coocurrencia de los niveles de grises y la longitud de niveles de grises (Pérez-Barnuevo, Lévesque, & Bazin, 2018), utilizándose también para la detección y clasificación de granos y texturas vistas a través de

microscopios (Tiu, 2017). El estudio reciente realizado por Alzubaidi y otros (2020), demuestra que modelos pre entrenados basados en redes neuronales convolucionales (CNN) presentan buenos resultados clasificando litologías a macro escala. Siguiendo esta misma línea, es que dentro de este informe se propone estudiar la factibilidad de usar herramientas desarrolladas recientemente como el Auto Machine Learning y la arquitectura cGan en el contexto minero.

Dentro de este informe se describirá el procedimiento y resultados obtenidos de la aplicación de herramientas de *machine learning* para describir automáticamente recortes de sondajes con respecto a su litología y mineralización.

1.1 DESCRIPCIÓN DEL PROBLEMA

Actualmente, el método más común de la caracterización de cuerpos geológicos es a través de la descripción manual de sondajes, la cual es una tarea que depende directamente de la interpretación y experiencia del equipo de geología. Junto a esto, se debe ponderar el balance entre la velocidad de avance y el detalle de la descripción, por lo cual, suelen quedar fuera datos importantes debido a la dificultad de extraer rápidamente aspectos cuantitativos, como es el porcentaje de vetas por tramos para poder ser agregados a las bases de datos.

1.2 SOLUCIÓN PROPUESTA

Se propone describir la litología y calcular el porcentaje de vetas presente a través de modelos de *machine learning* entrenados específicamente para cada una de estas tareas. Por lo tanto el proceso constará de dos pasos, en el primero se detectará a que litología pertenece cada sector y en el segundo se detectará la presencia de mineralización en forma de vetas y se segmentarán según su área visible.

1.3 OBJETIVOS

1.3.1 OBJETIVO GENERAL

Desarrollar una metodología de clasificación y estimación de vetas en sondajes tipo diamantino utilizando *machine learning* y visión computacional.

1.3.2 OBJETIVOS ESPECÍFICOS

- Desarrollar una metodología que permita diferenciar entre diferentes tipos de litologías.

- Desarrollar una metodología para desagregar la mineralización tipo veta y separarla de la roca caja.
- Advertir aspectos importantes en la confección de la base de datos para las metodologías propuestas.

1.4 ALCANCES

Dentro de este trabajo no se busca desarrollar nuevos modelos de *machine learning* para la detección o segmentación, sino más bien, se busca evaluar la aplicabilidad de modelos prediseñados en contexto minero y geológico. Cabe destacar que tampoco se realizará el modelo geoestadístico, no se estimarán los recursos, no se desarrollará un modelo de bloques y tampoco se realizará una interpretación geológica profunda de los resultados.

Debido a que los datos utilizados en esta memoria fueron suministrados por una empresa la cual pidió anonimato para evitar la filtración de información sensible, no se describirá la ubicación ni a qué tipo de yacimiento pertenece. Junto a esto, cabe destacar que las descripciones y fotografías fueron recopiladas de sondeos antiguos, por lo que no se pudo acceder a ellos para corroborar la información descrita, sin embargo, es comunicado por parte del equipo de geología, que la litología marga es una de las *mapeadas* con mayor detalle, ya que presenta principal interés debido a su relación con un tipo de mineralización que impacta posteriormente en los siguientes procesos. Por lo tanto, dentro de todo este trabajo nos centraremos en esta litología, a la cual llamaremos “litología de interés”.

1.5 METODOLOGÍAS Y HERRAMIENTAS UTILIZADAS

Debido a que los modelos de *machine learning* que se utilizarán en este trabajo emplean entrenamiento supervisado, el trabajo iniciará analizando, adaptando y etiquetando los datos que se utilizarán tanto para el entrenamiento como para la evaluación. Luego de esto, se procederá al entrenamiento para la clasificación de litología y la segmentación de vetas. Finalmente, los modelos entrenados serán aplicados en nuevos sondeos, dando como resultado la diferenciación de la litología y el área visible de vetas.

A continuación se presenta un resumen del procedimiento realizado:

1.5.1 PREPROCESAMIENTO DE DATOS

- Recopilar la descripción de los sondeos realizada por el equipo de geología y sus respectivas imágenes.
- Realizar un tratamiento previo de las imágenes.
- Etiquetar manualmente la litología dentro de las imágenes.
- Seleccionar los sondeos de entrenamiento, validación y prueba.
- Recortar y generar una base de datos con las imágenes y su litología correspondiente.

1.5.2 DIFERENCIACIÓN LITOLÓGICA

- Separar los recortes de entrenamiento según su litología usando categorías binarias.
- Separar los recortes de entrenamiento según su litología usando múltiples categorías.
- Entrenar modelos de clasificación litológica.
- Separar una pequeña muestra de recortes según el elemento geotécnico presente.
- Entrenar un modelo de clasificación para el estado geotécnico.
- Predecir el elemento geotécnico dentro de todos los recortes de entrenamiento.
- Entrenar un modelo de clasificación litológica de categoría binaria para cada estado geotécnico.

1.5.3 DETECCIÓN Y ESTIMACIÓN DE PORCENTAJE DE VETAS

- Separar recortes según la alteración y mineralización dentro de la litología de interés.
- Entrenar modelo de clasificación de alteración y mineralización dentro de la litología de interés.
- Segmentar manualmente las zonas de vetillas.
- Entrenar un modelo de segmentación.

1.5.4 PREDICCIÓN EN DATOS DE PRUEBA

- Predecir la litología utilizando cada modelo entrenado para esta tarea.
- Detectar la presencia de vetas.
- Segmentar y calcular el área superficial de las vetas.

Para la diferenciación de los elementos geotécnicos, la litología y la mineralización, se utilizará el entorno de entrenamiento de Google Cloud Platform. Para la estimación de porcentaje de vetas se utilizarán librerías de Python como keras y TensorFlow, implementados dentro del entorno virtual Google Colab. La predicción con los modelos entrenados se realizará a través de Docker.

1.6 RESULTADOS ESPERADOS

Con este trabajo se busca obtener dos grupos de modelos entrenados de *machine learning*, uno para diferenciar la litología de interés y mineralización, y otro tipo para el cálculo de porcentaje de vetas dentro de esta litología. En la Tabla 1 se puede observar un ejemplo del resultado final esperado.

Tabla 1

Ejemplo de la predicción de litología y vetas

Desde (m)	Hasta (m)	Litología	Veta (cm ²)
0	1	Litología de interés	5
1	1.5	Otra litología	-
1.5	1.7	Litología de interés	30

Nota. Esta tabla representa un ejemplo ficticio de la recopilación esperada de la predicción de la litología y la estimación de las vetas dentro de la litología de interés.

1.7 ORGANIZACIÓN DEL DOCUMENTO

El documento continúa con el capítulo de Desarrollo, en donde, dentro de la sección Marco teórico se abordan todos los conceptos que sustentan este trabajo. Luego de esto se aborda la sección de Caso de estudio, en el cual se exponen todos los datos específicos que serán utilizados para la aplicación de las metodologías y se explicarán las decisiones tomadas, por ejemplo, sobre cómo se realizará la separación de las bases de datos. Dentro del siguiente capítulo se presentan los resultados obtenidos producto a la metodología planteada. El documento termina con la presentación de las conclusiones generales y luego las conclusiones propias de cada objetivo específico.

2. DESARROLLO

Este capítulo inicia abordando los conceptos básicos relacionados a la minería y a los algoritmos necesarios para poder entender el trabajo realizado y cuál es su potencial impacto. Luego de esto se expone un caso de estudio en donde es necesario diferenciar una litología en específico y estimar la presencia de vetas.

2.1 MARCO TEÓRICO

Para poder asimilar el impacto que tiene la descripción de sondajes o *mapeo* dentro del negocio minero y poder entender cómo es que la inteligencia artificial puede ayudar a solucionar algunos de estos problemas, necesitaremos abordar primero conceptos sobre la exploración minera y la estimación de minerales. Junto a ello, es importante conocer que es un algoritmo de inteligencia artificial, que tipos existen y cuales se utilizarán con la finalidad de cumplir con los objetivos planteados.

A continuación, se abordarán los conceptos relacionados con la minería, los cuales están fuertemente ligados a la exploración y evaluación de yacimientos.

2.1.1 EXPLORACIÓN DE YACIMIENTOS

La industria minera es sin lugar a duda un negocio de alto riesgo de inversión, sobre todo dentro de la primera etapa correspondiente a la exploración geológica (Figura 1). Esto se debe principalmente al alto nivel de incertidumbre que existen dentro de la obtención de información y estimaciones del cuerpo mineralizado.

Se ha estudiado el impacto de las campañas de exploración en el valor del plan minero, obteniendo como conclusión que es conveniente utilizar una malla de sondajes lo más pequeña posible, ya que el costo de inversión es sustancialmente menor al valor del negocio que se pone en riesgo por falta de información (Verdugo H., 2016). Otra forma de disminuir el riesgo es minimizar el error en la obtención de los datos. Es importante destacar que la estimación visual de vetas obliga a utilizar intervalos, lo cual tiene un gran error asociado.

2.1.1.1 Exploración de yacimientos

La exploración de yacimientos suele utilizarse para buscar nuevos cuerpos mineralizados, para expandir una minera o para confirmar los datos estimados a medida que avanza el proyecto. Esta

exploración se basa principalmente en dos metodologías en paralelo, por un lado, se tiene la obtención de muestras muy pequeñas de roca, las cuales son enviadas a laboratorio para su análisis químico. Junto a esto, se realiza la descripción de los sondajes a través de un proceso denominado *mapeo*, el cual puede ser realizado por un humano o una maquina y que consiste en la descripción de anomalías geotécnicas como la presencia de fracturas o roca molida, y también la descripción litológica como el tipo de roca, la presencia de alteraciones o mineralización.

2.1.1.2 Sondaje

Corresponden a muestras cilíndricas de roca con dimensiones que usualmente van desde los 27 mm (AQ) a los 85 mm (PQ) de diámetro y de longitud variable que puede ir desde los cientos hasta los miles de metros. Por lo regular, para su obtención son empleadas maquinarias de perforación por diamantina, martillo rotatorio o por percusión (Valenzuela, 2015).

Una vez extraídos los sondajes del suelo, estos son almacenados y etiquetados en cajas planas de madera o aluminio para su posterior análisis (Fotografía 1). Estas cajas se distinguen por estar divididas a lo largo en compartimientos con anchos acorde al diámetro de los testigos y por tener un largo menor a los 1.5 metros de longitud, para facilitar su manipulación debido al peso de la roca (Gandhi & Sarkar, 2016).



Fotografía 1 Sondajes diamantino almacenado en la caja porta testigos. Recuperado de: <https://www.induambiente.com/especial/mineria/impactos-de-exploracion>

2.1.1.3 Descripción geológica manual

La descripción manual es ampliamente utilizada en las faenas mineras. Esta metodología consiste en el estudio exhaustivo por más de un geólogo experto para discutir las apreciaciones, debido a que

cada uno puede tener su propia interpretación de la misma alteración o falla geológica. Este análisis consiste en la extracción de un amplio espectro de información, como es la determinación de la litología, las estructuras geotécnicas y las potenciales zonas de alteración del depósito mineral (Gandhi & Sarkar, 2016). Esta labor no solo conlleva una gran inversión de tiempo, sino que también se basa en el conocimiento y experiencia de cada geólogo (Bonnici, 2008), lo cual puede conllevar a problemas en las bases de datos de no seguir un criterio fijo a lo largo de los años.

La velocidad de avance por geólogo puede ser muy variable dependiendo de la experiencia previa, conocimiento del yacimiento en específico o el nivel de detalle de la descripción, pero en promedio la capacidad de descripción no suele superar los 100 m por día. Junto a esto, hay parámetros difíciles de describir manualmente como son los factores cuantitativos, por ejemplo, el número de vetillas o su porcentaje de presencia.



Fotografía 2 Geólogos *mapeando* sondajes diamantinos en el proyecto Rockhaven's Klaza en Yukon.
Recuperado de: Rockhaven Resources Ltd

2.1.1.4 Litología

La litología es la rama de la geología que estudia las características de las rocas que constituyen una determinada formación geológica. Este dato presenta fundamental importancia para la geotecnia, ya que proporciona una base para entender cómo se ejercieron las fuerzas de empujes tectónicos, los agentes de erosión, transporte y el clima (E-CONSTRUIR, s.f.).

Desde la mirada litológica, los materiales se clasifican de acuerdo con su origen de formación en (Suárez, 1998):

Rocas sedimentarias: Se forman por la sedimentación y cementación de partículas de arcilla, arena, grava o cantos. Poseen planos de estratificación (Suárez, 1998).

Rocas volcánicas o ígneas extrusivas: También conocidas como rocas Piroclásticas, son el producto de la cristalización de los materiales expulsados por los volcanes (Suárez, 1998).

Rocas ígneas intrusivas: Son producto del enfriamiento del Magma bajo la superficie terrestre. Si bien las rocas ígneas corresponden al 98% del volumen de la corteza terrestre, son menos comunes en la superficie que las rocas sedimentarias (Suárez, 1998).

Rocas metamórficas: Son el resultado de la recrystalización de rocas ígneas y sedimentarias. Dentro de este proceso se producen cambios de textura y mineralógicos, dando como resultado la alteración de las características originales o la pérdida completa de éstas (Suárez, 1998).

Depósitos orgánicos: Son depósitos de materiales orgánicos, como la acumulación de turba o de material orgánico que no se ha descompuesto totalmente debido a su alto contenido de agua. Los depósitos orgánicos en ocasiones se encuentran estratificados con otros elementos tales como limos o arenas o entremezclados con arcilla (Suárez, 1998).

2.1.1.5 Textura

Dependiendo del contexto, nos podremos referir a dos tipos de texturas

En geología, el concepto de textura hace referencia a la manera en la cual los granos individuales o minerales se encuentran dispuestos en la roca. Dentro de los tipos de textura tenemos cristalina, hipocristalina, hridralina y clástica (Suárez, 1998). Si bien las texturas geológicas no suelen presentar patrones uniformes, si existe una repetición no casual de partes elementales que poseen características similares (Marulanda, 2016).

En imagenología, la textura es considerada una de las características principales para la clasificación de imágenes, la cual representa la variación del brillo local de píxel a píxel en un pequeño vecindario en una imagen digital (Bharati, Liu, & F.MacGregor, 2004). Al clasificar imágenes de textura se busca diferenciar patrones característicos que pertenezcan a una categoría o clase.

2.1.1.6 Alteraciones

En la geología, se denomina alteración a el conjunto de modificaciones de las propiedades fisicoquímicas de los minerales, y por lo tanto, de las rocas.

2.1.1.7 Mineralización

La mineralización es producto a la alteración de las rocas con elementos económicamente importantes. Su proceso puede ser hidrotermal, por sedimentación, entre otras.

Debido a su relación con respecto a los elementos de interés, suele tener gran importancia en el mapeo geológico.

2.1.2 EVALUACIÓN DE YACIMIENTOS

Este es el proceso de utilizar la información recopilada durante la exploración, para poder estimar aspectos del yacimiento como el volumen, concentración de minerales, etc, con la finalidad de evaluar económicamente la extracción y procesamiento de ese yacimiento.

2.1.2.1 Modelo de bloques

El modelo de bloques es la cubicación del yacimiento mineral. Esto se realiza para ubicar espacialmente los datos recolectados en la etapa de exploración, con datos como la densidad, leyes, litología y elementos geotécnicos. Con la información inicial y puntualizada, son estimadas las zonas desconocidas con un cierto valor de confianza. Cada sector tendrá un grado de confianza asociado a cuanto información se tenga disponible para estimar ese punto y a que distancia se encuentran los datos de muestreo.

2.1.2.2 Reservas y recursos minerales

Una vez estimado todo el cuerpo geológico a través del modelo de bloques, se cataloga cada sector correspondientemente al nivel de confianza que se tiene sobre la información y con la varianza asociada al modelo que se utilizó para estimarlo. Cuando el dato posee el grado mínimo de confianza se denomina recurso inferido, para un nivel medio de confianza se denomina indicado y medido para su grado de máxima confianza (Figura 2).

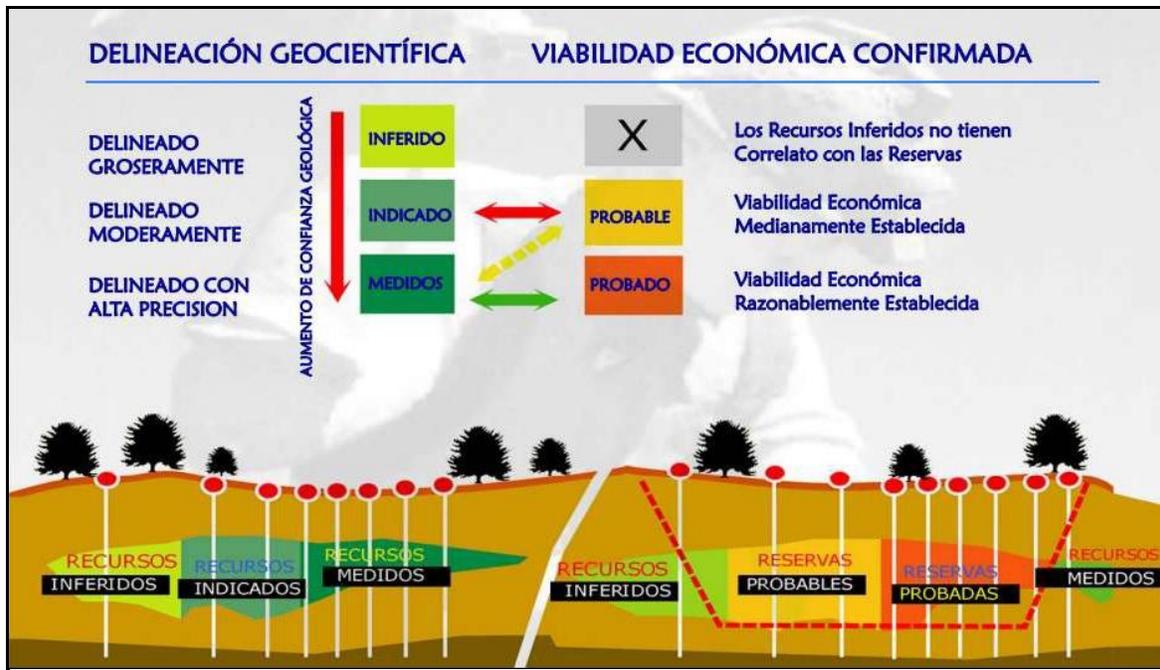


Figura 2 Esquema de las clasificaciones de recursos y su correlato en reservas. Fuente: Instituto de Ingenieros de Minas de Perú.

Las reservas minerales corresponden a la parte económicamente explotable de un recurso medido o indicado, el cual debe ser demostrado al menos por un estudio de prefactibilidad y debe cumplir con los estándares que regulan la categorización de recursos como son los códigos CIM, JORC, UNFC, SEC y SAMREC. Dentro del estudio de prefactibilidad se deben considerar factores económicos, ambientales, legales, sociales, entre otros que justifiquen la factibilidad técnica y económica de la extracción del mineral.

No necesariamente todos los recursos medidos e indicados pasarán a reservas. Como se puede observar en la Figura 2, parte de los recursos medidos quedaron fuera del rajo final debido a que no son económicamente rentables a pesar de tener un alto nivel de confianza. Por otro lado, como no es posible pasar de recursos inferidos a reservas, a pesar de que pueda haber leyes altas en estos sectores, sus leyes no son consideradas para la estimación del rajo óptimo debido a su bajo nivel de confianza.

Otro punto importante es que los recursos se dividen en probables y probados dependiendo de la certeza que se tenga de la viabilidad de la explotación de ese mineral, no obstante, estas categorías no son fijas, ya que a medida que avanza el proyecto minero se obtiene nueva información, lo que puede modificar la viabilidad económica de extracción haciendo que cambie de categoría o puede variar por agentes económicos externos como el precio del mineral en la bolsa.

2.1.3 INTELIGENCIA ARTIFICIAL

La inteligencia artificial es el macro área (Figura 3) que busca crear algoritmos con el propósito de crear máquinas que presenten las mismas capacidades que el ser humano (Ibedrola, s.f.).

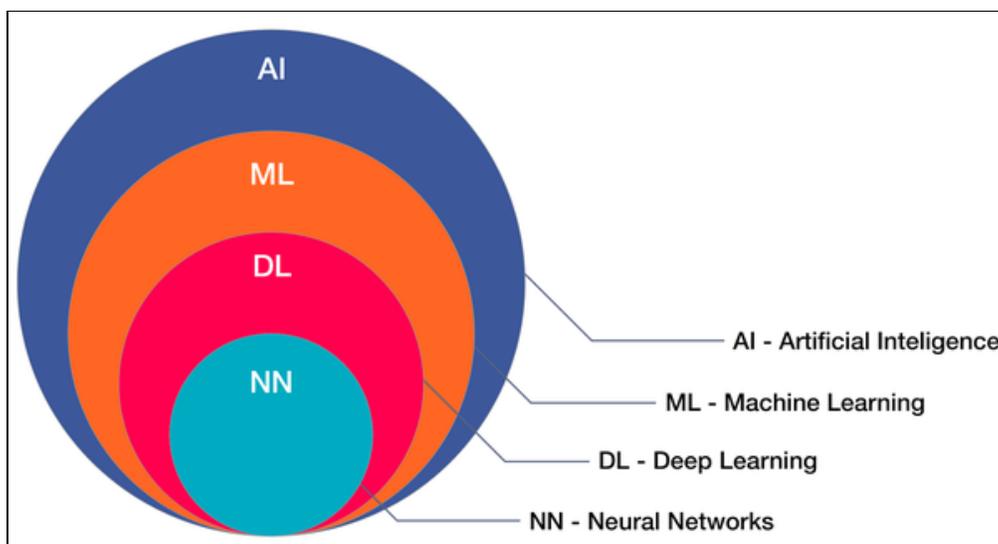


Figura 3 Esquema del campo de estudio de la inteligencia artificial y sus subáreas. Recuperado de: TerryTaewoongUm, *Deep-learning-machine-learning-tutorial-for-beginners*

2.1.3.1 Machine Learning

El aprendizaje de máquinas o *machine learning* es una disciplina científica del ámbito de la inteligencia Artificial que crea sistemas que aprenden automáticamente (González, 2021). Aplicado en este contexto, aprender quiere decir saber identificar patrones en cantidades inmensas de datos de forma concreta, además de que sean capaces de comprender y estructurar comportamientos a partir de información no estructurada (Espinoza, 2019). El modelo debe ser capaz de aprender gracias a la comprensión de la información. Se dice que aprende de unos determinados datos o experiencias E , con respecto a una tarea T , un medidor de eficiencia EF , si el rendimiento de una tarea T , medido por medidor de eficiencia EF , logra mejorar con la experiencia E (Rodríguez-Sahagún, 2018).

2.1.3.2 Tipos de tareas de Machine Learning

En base a la definición de aprendizaje de *machine learning* mencionada anteriormente, existen numerosos tipos de tareas que se pueden abordar utilizando *machine learning*. A continuación, se describe brevemente alguna de ellas (Rodríguez-Sahagún, 2018):

Clasificación: Este tipo de tarea consiste en especificar a cuál de las x categorías posibles pertenecen los datos de entrada. Para realizar esta tarea, usualmente se crea una función $y = f(x)$, en donde el dato de entrada corresponde a x y la clasificación del dato corresponde a y (Rodríguez-Sahagún, 2018).

Clusterización: Consiste en dividir una población de elementos o puntos en varios grupos de modo que los datos del mismo grupo sean similares entre sí y diferentes a los puntos de datos de otros grupos. Si bien esta tarea suele realizarse como un método no supervisado para la exploración preliminar de los datos, en este trabajo se realizará la clusterización de elementos específicos dentro de imágenes a través de métodos semi-supervisados.

Regresión: El modelo debe predecir un valor numérico dado cierto dato de entrada y , en vez de clasificarlos con una variable categórica, la salida debe ser una variable numérica (Rodríguez-Sahagún, 2018).

Detección de anomalías: El modelo busca patrones dentro de una serie de eventos u elementos considerados anómalos o atípicos (Rodríguez-Sahagún, 2018).

2.1.3.3 *Aprendizaje supervisado, semi-supervisado y no supervisado*

Dentro de la definición de *machine learning*, se dice que el modelo aprende de cierta experiencia E a través de un proceso de aprendizaje del modelo, el cual puede ser de distintos tipos (Rodríguez-Sahagún, 2018):

Aprendizaje supervisado: Los modelos que utilizan este tipo de aprendizaje obtienen como entrada datos que contienen una cierta característica junto a una etiqueta asociada, siguiendo el esquema presentado en la Figura 4. El concepto de supervisado proviene del hecho de que se conoce previamente la etiqueta correcta de cada dato, por lo cual se ajusta el aprendizaje con respecto a la tasa de aciertos y errores de las predicciones de validación (Rodríguez-Sahagún, 2018). Debido a esta dinámica de supervisión, es posible que este método presente escenarios de underfitting u overfitting.

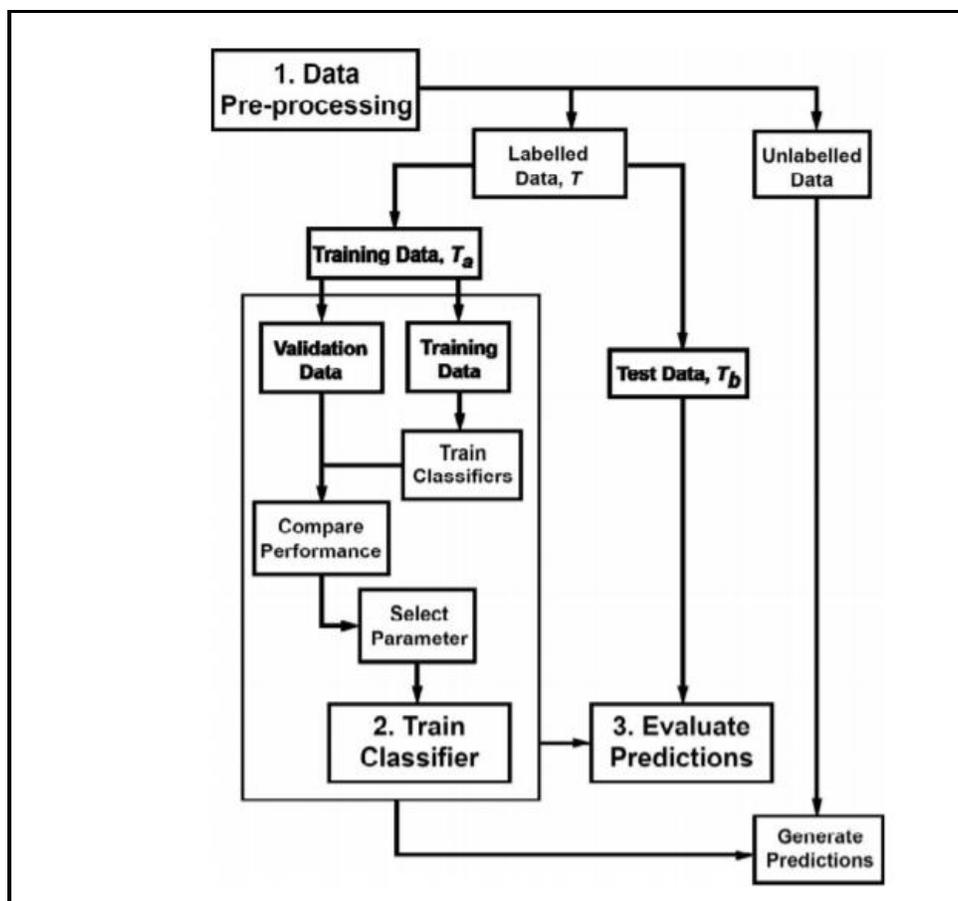


Figura 4 Flujo de trabajo estándar para el entrenamiento supervisado de modelos de *machine learning* (J. Cracknell, 2014)

Aprendizaje no supervisado: Este tipo de modelo analiza una serie de datos que contienen una cantidad indefinida de características, de las cuales el propio modelo debe aprender a extraer las propiedades para poder diferenciar patrones de interés o separar grupos de datos (Rodríguez-Sahagún, 2018).

Aprendizaje semi-supervisado: Los modelos que utilizan este tipo de aprendizaje se caracterizan por utilizar tanto datos etiquetados como no etiquetados. Se ha descubierto que haciendo uso de una pequeña cantidad de datos etiquetados es posible mejorar considerablemente la exactitud del aprendizaje (Rodríguez-Sahagún, 2018). Dentro de los modelos destacados que utilizan esta metodología se encuentra el tipo específico de arquitectura denominada GAN (Generative Adversarial Network) y que dio paso a la arquitectura cGAN (Conditional Generative Adversarial Network), en las cuales se profundizará más adelante.

2.1.3.4 *Overfitting y underfitting*

El principal desafío del *machine learning* es que los modelos no solo tengan buen rendimiento con los datos con los cuales ha sido entrenado, sino que también pueda ser capaz de rendir frente a nuevos datos de entrada. Se denomina generalización a la habilidad de obtener un buen rendimiento con datos no observados durante el entrenamiento (Rodríguez-Sahagún, 2018).

Típicamente, cuando se entrena un modelo de *machine learning* el grupo de datos es dividido en tres partes: un grupo de entrenamiento, un grupo de validación y un grupo de prueba.

Los grupos de entrenamiento y validación son utilizados para probar un gran número de procesamientos, arquitecturas e hiper parámetros, dando como resultado diversos modelos entrenados que luego serán evaluados en el grupo de validación para guiar la búsqueda de la combinación óptima. Esta búsqueda irá guiada con el objetivo de minimizar el error de entrenamiento.

Una vez que deja de reducirse el error de entrenamiento, es considerada la combinación con mejor rendimiento como la versión final, la cual es utilizada para medir la capacidad de generalización del modelo al utilizar entradas nunca vistas durante el entrenamiento (Google, s.f.).

El rendimiento del modelo de *machine learning* se medirá según su capacidad de minimizar el error de entrenamiento y de generalización a la vez. Estos dos factores se encuentran ligados a los principales desafíos del *machine learning*: *overfitting* y *underfitting*. El principal parámetro que afecta a estos dos desafíos es la complejidad del modelo, si el modelo es muy simple se tenderá a un escenario de *underfitting* en donde ambos errores se mantienen altos. Si el modelo resulta ser demasiado complejo para esa tarea, se tenderá hacia un escenario de *overfitting*, en donde el error de entrenamiento sigue disminuyendo mientras el de generalización aumenta, esto provoca que el modelo de *machine learning* esté sesgado hacia los datos de entrenamiento y validación (Rodríguez-Sahagún, 2018).

2.1.3.5 *Redes neuronales*

Las redes neuronales son redes interconectadas de forma masiva en paralelo de elementos simples y con cierta organización jerárquica, las cuales interactúan entre sí emulando el sistema nervioso del ser humano. Las redes, al interconectar las neuronas, generan 3 tipos de capas importantes: la capa de entrada, capas ocultas y capa de salida. Cada neurona se denomina nodo y existe funciones básicas las cuales determinan su comportamiento (Espinoza, 2019).

2.1.3.6 Automated machine learning (AutoML)

AutoML es el proceso de automatizar los procesos de desarrollo iterativos de los modelos de *machine learning*, aplicados principalmente a las tareas de clasificación, regresión y series de tiempo. Normalmente, el desarrollo de estos modelos requería un gasto importante de recursos y la necesidad de un gran dominio en distintas áreas del *machine learning*. Sin embargo, con la implementación de AutoML permite bajar la barrera mínima de conocimiento de entrada al crear automáticamente modelos con un grado adecuado de complejidad y altos niveles de eficiencia (Microsoft, 2020).

Dentro de los procesos automatizados se encuentran (Hutter, y otros, 2018):

- Preparación y gestión de los datos
- Detección de tipo de datos (booleano, numérico discreto, numérico continuo o texto)
- Detección de intento (objetivo/etiqueta, función numérica, función de texto categórico, etc)
- Detección de tarea (clasificación binaria, regresión, series de tiempo)
- Ingeniería de características
- Selección de características
- Extracción de características
- Meta aprendizaje y aprendizaje por experiencia
- Detección y manejo de datos atípicos o perdidos
- Selección del modelo
- Optimización de los hiper parámetros del algoritmo de aprendizaje y caracterización
- Selección de tiempo máximo, memoria y limitaciones de complejidad
- Selección de las métricas de evaluación y procedimiento de validación
- comprobación de problemas
- Detección de fugas
- Detección de configuración incorrecta
- Análisis de resultados obtenidos
- Matriz de confusión

2.1.3.7 Redes generativas adversarias (GAN)

Estos modelos se destacan por estar compuesto por dos redes neuronales profundas que son adversarias entre sí. Con esto nos referimos a que compiten dentro de un juego de suma cero, donde si una de las redes gana, la otra pierde. Estas redes son llamadas Generador y Discriminador (Figura 5). La tarea del Generador es aprender a crear imágenes que parezcan reales con respecto a

las imágenes de ejemplo. Por otro lado, la tarea del Discriminador es diferenciar si una imagen es real, en donde conocemos que la respuesta correcta será “real”, o por el contrario, si la imagen entregada es un producto del Generador, en donde la respuesta correcta será “falso” (Martinez J. , 2020).

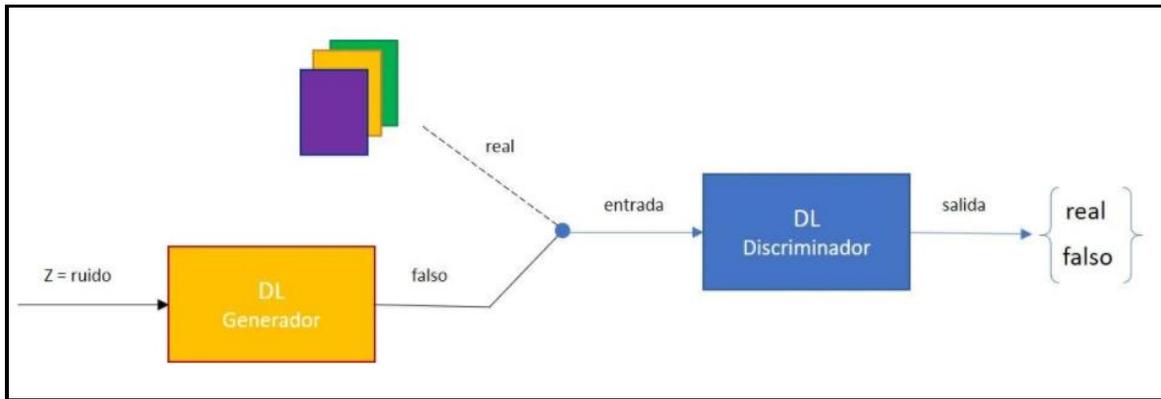


Figura 5 Diagrama de funcionamiento de una red neuronal adversaria (Martinez J. , 2020)

De esta forma, con una pequeña porción de ejemplos se puede entrenar un modelo tipo GAN, en donde el aprendizaje se desarrollará en la dinámica de competición entre ambas redes. Al inicio será sencillo que el Discriminador pueda diferenciar entre una imagen real o generada, pero a medida que el Generador mejora, la tarea de discernir será más difícil, haciendo que el Discriminador deba ir mejorando simultáneamente (Martinez J. , 2020).

Las entradas iniciales de estos modelos suelen ser vectores aleatorios, el cual generará una salida con respecto a los datos de entrenamiento. Por ejemplo, podríamos estar generando imágenes de caras o paisajes. Usando otro vector obtendremos otra cara, paisaje o con lo que se le haya entrenado (Martinez J. , 2020).

2.1.3.8 Redes generativas adversarias condicionales (cGAN)

Para mejorar los modelos generativos se puede influenciar el resultado utilizando una condición que se le da tanto al Generador como el Discriminador (Figura 6). Esta condición puede ser cualquier tipo de información auxiliar como una clase o una imagen con un estilo en particular.

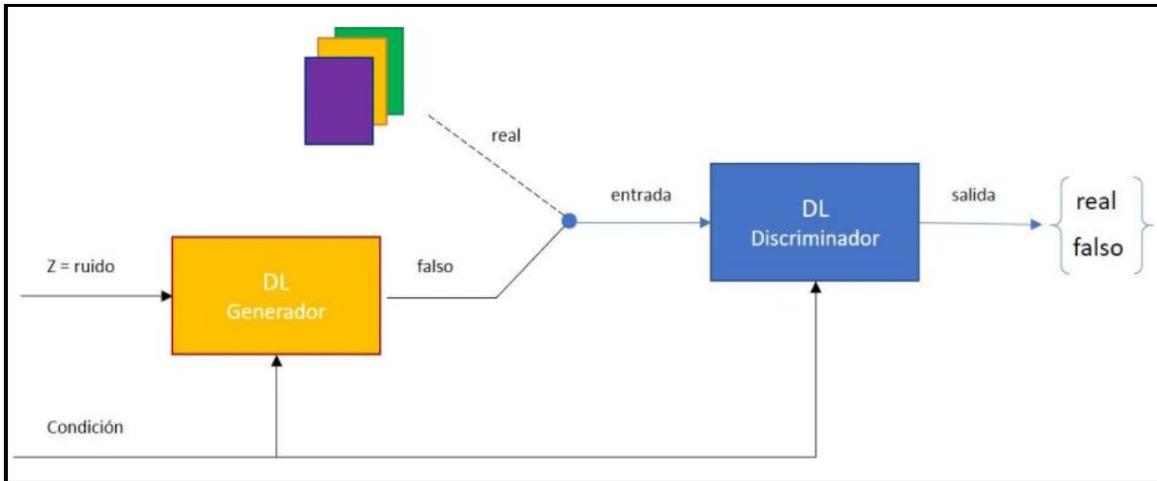


Figura 6 Diagrama de funcionamiento de una red neuronal condicional adversaria (Martinez J. , 2020).

Gracias a esta condición se consigue resolver tareas como el cambio de estilo de imagen, fotografías a partir de bocetos, aumentar la resolución de fotografías, entre otras aplicaciones. En este trabajo utilizaremos un modelo cGAN para que aprenda a segmentar únicamente la zona de vetas dentro de los recortes de sondajes.

2.1.3.9 Matriz de confusión

La matriz de confusión es una herramienta sumamente útil para valorar el desempeño de un modelo de clasificación de *machine learning*. Dentro de la Figura 7, cada columna de la matriz representa el número de predicciones de cada clase, mientras que la sumatoria de cada fila representa la totalidad de los casos reales por cada clase.

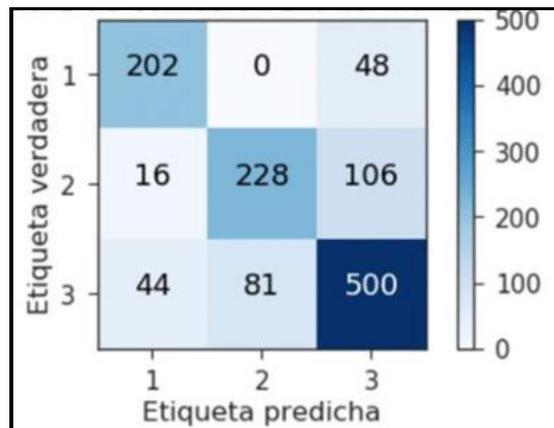


Figura 7 Ejemplo de matriz de confusión con 3 etiquetas (Garcia, 2018).

Según Alonso, dentro de la matriz de confusión tendremos 4 posibles escenarios (Figura 8):

Verdaderos positivos (VP): Es el número de casos positivos que fueron clasificados correctamente como positivos por el modelo

Verdaderos negativos (VN): Es el número de casos negativos que fueron clasificados correctamente como negativos por el modelo

Falsos positivos (FP): Es el número de casos positivos que fueron clasificados incorrectamente como negativos por el modelo

Falsos negativos (FN): Es el número de casos negativos que fueron clasificados incorrectamente como positivos por el modelo

		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Figura 8 Matriz de casos posibles dentro de cada categoría. Recuperado de: <https://rpubs.com/chzelada/275494>.

Utilizando el número de casos en cada escenario, se estima la efectividad del modelo con respecto a 5 métricas:

Exactitud: Clasifica los datos correctamente. Los modelos con mayor número de etiquetas pueden tender a tener un rango de exactitud grande, debido a que el número de verdaderos negativos suele ser considerablemente mayor que el de verdaderos positivos, siendo este segundo factor mucho más importante. Por lo tanto, la exactitud debe ser contextualizada con el resto de las métricas.

$$Exactitud = \frac{VP+VN}{Total} \tag{1}$$

Sensibilidad: La cantidad de aciertos catalogados como positivos dentro del número real de positivos. Entre mayor es la sensibilidad, el modelo logrará reconocer todas las etiquetas presentes en los datos de prueba.

$$Sensibilidad = \frac{VP}{Total\ positivos} \quad (2)$$

Especificidad: La cantidad de aciertos catalogados como negativos dentro del número total de casos negativos. Si la especificidad es baja significa que están ocurriendo varios casos de falsos positivos o falsos negativos. Cuando la etiqueta no esté presente dentro del entrenamiento del modelo, la especificidad alcanzará el valor de 100%, debido a que no puede haber falsos positivos, el cual es su factor complementario.

$$Especificidad = \frac{VN}{Total\ negativos} \quad (3)$$

Precisión: La cantidad de aciertos catalogados como positivos dentro del número total de casos catalogados como positivos. Como lo dice su nombre, con esta métrica mediremos cuantas veces acierta a la etiqueta positiva, con respecto al total de veces que predice positivo. Entre mayor sea el porcentaje, la predicción de esa etiqueta será más confiable.

$$Precisión = \frac{VP}{Total\ clasificados\ positivos} \quad (4)$$

Valor de predicción negativo (VPN): La cantidad de aciertos catalogados como negativos dentro del número de casos total catalogados como negativo. Al igual que la precisión, esta métrica medirá la confiabilidad de la predicción negativa.

$$VPN = \frac{VN}{Total\ clasificados\ negativos} \quad (5)$$

2.1.4 SOFTWARES Y HERRAMIENTAS

2.1.4.1 Base de datos

Una base de datos corresponde a un conjunto de datos que comparten un mismo contexto y los cuales son almacenados sistemáticamente para su posterior uso.

2.1.4.2 Python

Python es un lenguaje de programación interpretado y multi paradigma, que soporta la orientación a objetos, programación imperativa y, en menor medida, la programación funcional (Luca AI Powered

Decisions, s.f.). A pesar de ser un lenguaje creado para uso general, ha sido adoptado por la comunidad del procesamiento de datos debido a sus importantes librerías.

2.1.4.3 *Google Cloud Platform (GCP)*

El servicio GCP es la integración de todos los servicios ofrecidos hasta la actualidad por la empresa Google. Dentro de estos servicios cuenta con una sección enfocada al *machine learning* y la visión computacional. Esta herramienta se enmarca en los modelos de AutoML, en donde permite cargar los datos de entrenamiento, validación y prueba, para luego ajustar automáticamente el modelo e hiper parámetros, entrenar el modelo y obtener métricas de efectividad (Acronis International, s.f.).

2.1.4.4 *Google Colaborador*

Google Colaborador o *Google Colab* es un entorno virtual que facilita el acceso gratuito a GPUs, permitiendo ejecutar y programar en Python desde el navegador (Google, s.f.). Entre una de las ventajas que tiene trabajar con *machine learning* en Google Colaborador, es la disposición previa de las librerías más utilizadas como Tensorflow y Keras.

2.1.4.5 *Spyder*

Spyder es un entorno científico y de código abierto escrito en Python, el cual ofrece un diseño orientado a científicos, ingenieros y analistas de datos (Spyder Website Contributors, s.f.).

2.1.4.6 *Graiph Map*

Esta es una herramienta diseñada para definir regiones en una imagen y adjudicarle una descripción en texto, ideal para la creación de las bases de datos de entrenamiento y revisión de predicciones.

2.1.4.7 *Safexa*

Safexa es una herramienta para la segmentación semi automática y de análisis de imagen, con la cual se confeccionará la base de entrenamiento del modelo de segmentación de vetas.

2.2 CASO DE ESTUDIO

Dentro de este apartado se aborda la aplicación de la metodología planteada a un caso de estudio en particular. Se comienza con una descripción superficial del contexto geológico del yacimiento para crear una idea de que se puede encontrar dentro de los sondajes. Es preciso recordar que fue omitido

todo tipo de información que pudiese conectar este trabajo con la empresa que facilitó la base de datos. Continuando con el desarrollo del caso de estudio, se realiza el análisis de las observaciones de las bases de datos e imágenes y se finaliza esta sección con las acciones descritas paso a paso realizadas dentro de este estudio.

Cabe destacar que dentro de todas las litologías presentes existe una en particular, llamada marga, que recibe especial interés a pesar de no poseer elementos de valor, ya que por el contrario, su estrecha relación con mineralizaciones de calcita en forma de vetas obliga a reajustar los reactivos dentro de los procesos metalúrgicos para no disminuir la efectividad de éstos. Por lo tanto, se decide enfocar este estudio a la litología tipo marga.

2.2.1 CONTEXTO GEOLÓGICO

El yacimiento mineral se encuentra hospedado en ambientes previamente marítimos, con secuencias volcano-sedimentarias tipo manto y con alteraciones asociadas a procesos de tipo skarn, observándose zonas de transición entre ambos tipos de mineralizaciones. Por lo tanto, es frecuente la presencia de rocas altamente calcáreas, brechas y tobas. Se presentan zonas bien definidas de mantos de origen principalmente terrígeno, carbonero y orgánico.

2.2.1.1 Litologías presentes

En la descripción manual de los sondajes utilizados en este trabajo se distinguen 12 litologías diferentes (Tabla 2), de las cuales 4 corresponden a transiciones entre arenisca y otras litologías.

Tabla 2

Litologías presentes dentro de los sondajes utilizados dentro de este trabajo.

Litología	abreviación	Imagen de ejemplo	Litología	abreviación	Imagen de ejemplo
Marga	MAR		Arenisca tobácea	ART	

Conglomerado	CG		Arenisca calcárea	ARCA	
Dique porfídico	DP		Arenisca	ARE	
Toba cristalina	TCR		Arenisca cornea	ARCO	
Cornea	COR		Brecha volcánica	BVC	
Coquina	COQ		Arenisca conglomerádica	ARC	

Nota. La apariencia de las litologías puede variar en comparación a los casos de ejemplo producto a las diferentes alteraciones.

Como fue mencionado en los alcances, debido a que la entidad que facilitó estas bases de datos mencionó que la roca margácea es de suma importancia, posee una descripción más detallada con respecto a otras litologías. La importancia de esta litología es que está asociada a el mayor porcentaje de la mineralización de calcita en vetas, por lo cual, todo el desarrollo posterior fue realizado con respecto a la roca margácea, a la cual nos referiremos como litología de interés.

2.2.1.2 Alteraciones

Las alteraciones serán estudiadas dentro de la litología seleccionada anteriormente como de interés, por lo cual, en base a la descripción realizada en las observaciones entregadas por el equipo de geología, se diferenciaron 3 categorías principales (Tabla 3); marga oscura, marga clara o cornificada y marga con materia orgánica.

Tabla 3

Alteraciones presentes dentro de la litología marga

Etiqueta	observación	características	Recorte de ejemplo
Oscura	Corresponde a la litología marga sin alteraciones	Color negro sin granos visibles y con alta presencia de vetas de calcita	
Clara/Cornificada	Presencia de alteraciones, principalmente cronificada o parecida a la arenisca calcárea	Es de color gris y presenta eventualmente mineralización de calcita en vetas	
Materia orgánica	Presencia de materia orgánica o fósiles estratificados	Color azulado con presencia vetilleo cebrado de tonalidades más oscuras	

2.2.1.3 Mineralización

Dentro de estos sondajes encontramos principalmente mineralizaciones de tipo estrato ligados debido a su formación sedimentaria (Tabla 4). Eventualmente se pueden encontrar alteraciones de pirita y calcopirita producto de la alteración tipo skarn.

Tabla 4

Tipos de mineralizaciones presentes dentro de la litología marga

Etiqueta	Observación	Características	Recorte de ejemplo
Vetillada	Presencia de calcita, la cual puede estar en forma de veta o molida	Vetas blancas	
Masiva	Es marga principalmente brechizada, en donde se tiene una matriz de calcita	La calcita representa el mayor porcentaje del segmento recortado	

Dentro de la base de datos, la mineralización es descrita en dos secciones, por un lado, son anotadas observaciones dentro de la descripción geológica relatando brevemente si es que existe alguna mineralización utilizando frases como “muy vetillada” o “vetas masivas entre los metros...”. En paralelo, se tiene una tabla en donde especifica la abundancia de mineralización por tramo, utilizando 4 categorías; traza, débil, moderada y fuerte.

2.2.2 ANALISIS DE LA BASE DE DATOS

Los sondajes utilizados provienen de una campaña de 80 sondajes con sus respectivas descripciones e imágenes de respaldo. Debido a que las fotografías tenían como única finalidad tener un respaldo gráfico de los sondajes, estas no presentan un estricto rigor en factores como la iluminación ambiental y la presencia de brillo en la superficie de la roca.

Se distinguen más de 24 litologías en total, de las cuales al menos la mitad corresponde a transiciones dentro de las litologías principales. Dentro de estas, existen 7 litologías que representan más del 60% de la longitud de los sondajes.

Antes de etiquetar los datos fue necesario revisar todas las observaciones dentro de las bases de datos, en busca de menciones sobre la presencia de pequeños segmentos de litologías dentro de tramos más extensos.

Se observa también que se presentan zonas descritas como transición que suelen ser, por ejemplo, litología *A* con presencia de litología *B*, o litología *B* con presencia de litología *A*, etiquetadas sin una constancia aparente. Debido a la imposibilidad de acceder a los sondajes, se tomó la decisión de seguir

el criterio del geólogo, sin embargo, luego de obtener los resultados de las predicciones se analizarán los casos mal clasificados (falsos positivos y falsos negativos) para analizar la probabilidad a que corresponda a una roca de transición con presencia de marga, no obstante, no será posible asegurar la veracidad de esas estimaciones.

2.3 PROCEDIMIENTO EXPERIMENTAL

En la figura A.1 se presenta el esquema del flujo de trabajo seguido dentro del procedimiento experimental, con el objetivo de diferenciar mejor en qué circunstancias fue entrenado cada modelo de *machine learning* y cuáles fueron sus salidas.

2.3.1 PRETRATAMIENTO DE DATOS

Recopilación de información: Fueron facilitadas las imágenes de los sondajes diamantino junto a su correspondiente descripción litológica por tramos, realizadas en terreno por el equipo de geología.

Tratamiento previo de las imágenes: Fueron seleccionados los sondajes con tonalidades de luz blanca, descartando los sondajes con tonalidades más oscuras producto a la luz ambiental, como puede verse en la separación de colores de la caja recopilados en la Figura 9. Luego se ajustó la perspectiva de las imágenes para corregir la proporción de píxel por distancia en los extremos superior e inferior con respecto al centro.

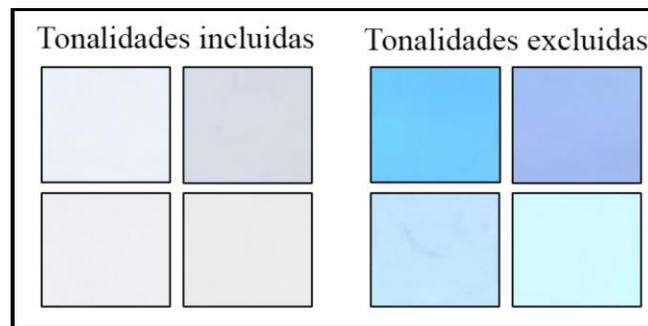


Figura 9 Recorte del mismo sitio de la caja de sondajes en fotografías con distinta tonalidad de luz.

Etiqueta manual de las imágenes: Se etiquetó la litología con rectángulos a lo largo de toda la superficie visible de los sondajes de entrenamiento y prueba, como puede observarse en la Figura 10, con el objetivo de poder recortar únicamente la superficie de la roca.

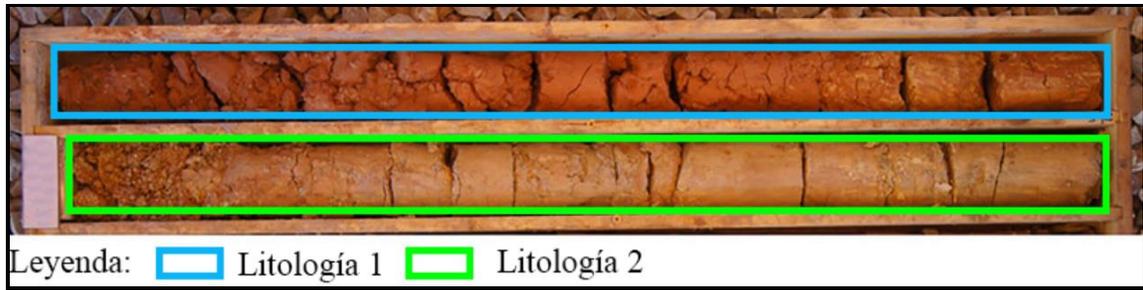


Figura 10 Fotografía de referencia con etiquetas de dos tipos de litologías diferentes (Elaboración propia).

Selección de datos de entrenamiento y de prueba: Para el grupo de entrenamiento se eligieron 7 sondajes con un total de 724 [m] lineales aproximados y con la presencia de 11 litologías diferentes (Tabla B1). El grupo de prueba estuvo compuesto por 4 sondajes con un total de 436 [m] lineales aproximados y 12 litologías presentes (Tabla B2).

Recorte y clasificación de las secciones del sondaje: Una vez etiquetados todos los sondajes, los rectángulos fueron divididos en pequeños cuadrados (Figura 11), los cuales fueron recortados y guardados en carpetas con el nombre de la litología correspondiente a esa sección.

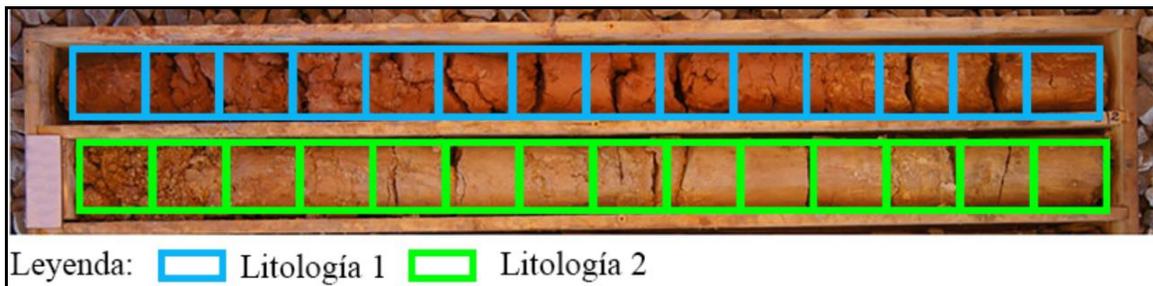


Figura 11 Fotografía de referencia con etiquetas de dos tipos de litologías separadas en cuadrados uniformes (Elaboración propia).

2.3.2 CONFECCIÓN DE LOS MODELOS PARA LA CLASIFICACIÓN LITOLÓGICA

Separar datos de entrenamiento según clasificación Multi litología (modelo 1.1): En este caso el modelo fue entrenado con todas las etiquetas de litologías presentes en esos sondajes, sin embargo, se realizó la comparación únicamente con respecto a la litología de interés. Cabe recordar que los resultados del resto de las litologías presentes no son realmente representativos, ya que no se tiene seguridad acerca de si están correctamente descritas.

Separar datos de entrenamiento según clasificación binaria (modelo 1.2): Para el entrenamiento de este modelo la base de datos fue separada en dos grupos principales “litología de interés” y “otra litología”.

Entrenar modelo de clasificación de litologías: Dentro de este paso se probaron las 2 formas distintas de etiquetado anteriormente descritas para comparar si existe alguna diferencia en los resultados.

Separar recortes por el elemento geotécnico presente en la roca: Dentro de todos los recortes, fue separada manualmente una pequeña muestra dentro de 5 categorías; roca intacta, roca semi intacta, roca fracturada, roca molida y roca borde, como puede observarse en la Figura 12. Esto se realiza con el objetivo de que cada categoría sea más consistente entre sus ejemplos.

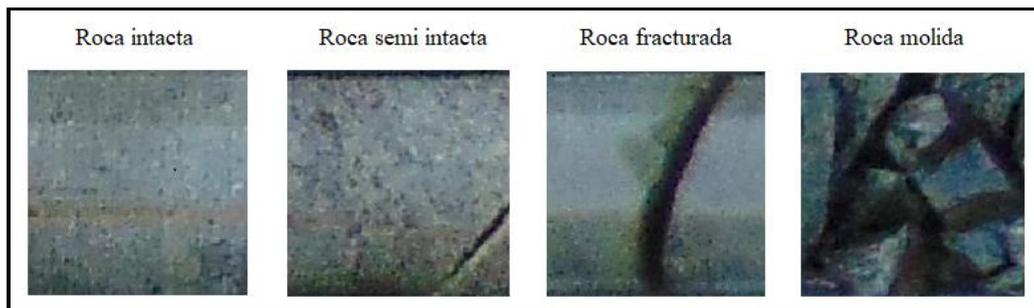


Figura 12 Ejemplos de las categorías del elemento geotécnico presente en la roca (Elaboración propia).

Entrenar modelo de clasificación para el estado geotécnico (modelo 1.3.F): Fue entrenado el modelo de AutoML para la clasificación del estado geotécnico, utilizando como entrada los ejemplos obtenidos en el paso anterior.

Predecir el elemento geotécnico: Una vez entrenado el modelo del paso anterior, fue utilizado para clasificar el estado geotécnico del resto de los recortes del grupo de entrenamiento. Al revisar manualmente los resultados, fue necesario modificar la etiqueta de una pequeña cantidad de casos.

Entrenamiento de clasificación binaria separada por estado geotécnico (modelos 1.3): Para estudiar la influencia del estado geotécnico de los recortes en la clasificación, se separaron los grupos de datos definidos para la clasificación binaria, utilizando las predicciones del modelo 1.3.F, obteniendo los modelos:

- Roca intacta (modelo 1.3.A)
- Roca casi intacta (modelo 1.3.B)

- Roca fracturada (modelo 1.3.C)
- Roca molida (modelo 1.3.D)
- Roca borde (modelo 1.3.E)

Esto fue realizado únicamente con el método de clasificación binaria, ya que con el método de etiquetado multi litología no se obtiene una cantidad mínima por cada etiqueta, por lo cual no se podría realizar un entrenamiento efectivo.

2.3.3 CONFECCIÓN DE MODELOS PARA LA DETECCIÓN Y ESTIMACIÓN DE VETILLAS

Con estos pasos se busca entrenar un modelo para diferenciar cuales recortes contienen presencia de vetas y otro para obtener la segmentación de éstas.

Separar recortes de la litología de interés: Debido a que necesitaremos buscar la presencia de vetas dentro de esta litología, fue separada una muestra manualmente en las categorías de; marga clara, marga oscura, marga vetillada, marga vetillada masiva y marga borde (Figura 13). Esta última etiqueta se especifica principalmente porque en la segmentación generan mayor número de falsos positivos, por lo cual es conveniente diseñar un protocolo diferente para esos casos.

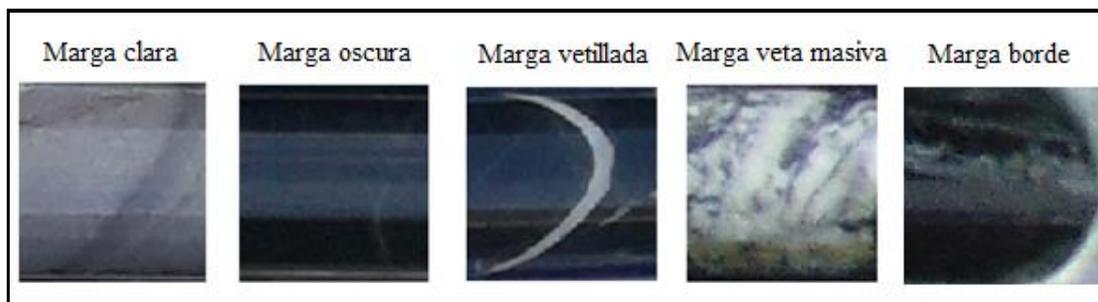


Figura 13 Posibles variaciones dentro de la roca marga.

Entrenar modelo de clasificación de variaciones de la litología de interés (modelo 2): Fue entrenado un modelo de clasificación de imágenes con los ejemplos obtenidos en el paso anterior.

Segmentar manualmente las zonas de vetillas: Fueron segmentadas las imágenes con presencia de vetillas (Figura 14) con el objetivo de generar la base de datos de entrenamiento para la detección y cálculo de porcentaje presente de éstas.

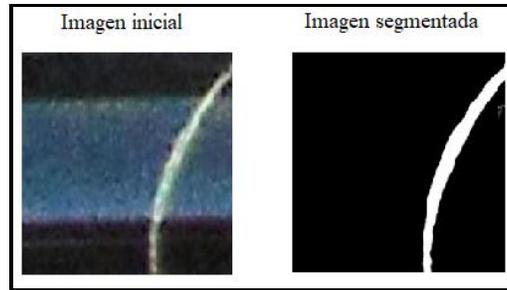


Figura 14 Segmentación manual de la veta presente dentro de un recorte de roca marga.

Entrenar modelo de segmentación (modelo 3): Con la base de datos generada en el paso anterior, se entrenó un modelo de redes generativas adversarias con el objetivo de que separe automáticamente las zonas correspondientes a vetas. Al principio se hizo el entrenamiento con una pequeña muestra para probar el modelo y al aplicarlo se fueron diferenciando los casos en donde no tenía un buen rendimiento, por lo que se fue reentrenando a medida que se incorporaban ejemplos de esos casos a la base de entrenamiento, se agregaron cronológicamente los siguientes grupos:

- Oscura vetillada
- Oscura vetillada con alta presencia de brillo
- Oscura no vetillada con alta presencia de brillo
- Clara vetillada, no vetillada, con y sin brillo
- Marga con veta masiva
- Diversos ejemplos en donde no hay vetas

2.3.4 PREDICCIÓN EN SONDAJES DE PRUEBA

Para comprobar la efectividad de los modelos entrenados anteriormente, se realizará la predicción en los sondeos de prueba simulando el proceso propuesto para la descripción litológica de sondeos a través de *machine learning*.

Predicción de litología: Se procedió a predecir por separado en los datos de prueba utilizando los modelos 1.1, 1.2 y los 6 correspondiente al modelo 1.3.

Predecir la variación dentro de la litología de interés: Este modelo fue utilizado solo dentro de la litología de interés del modelo 1.3, debido a que fue el que obtuvo el mejor rendimiento en la clasificación de litologías.

Calcular el porcentaje de vetas: Para calcular el porcentaje que corresponde la veta dentro de un segmento de sondaje, se tomó una muestra aleatoria de 354 recortes correspondientes a 18 [m] lineales de roca marga para un 95% de confianza. Se calculó el número de píxel correspondiente a la veta presente y fue dividido por el número de píxel totales que corresponde a ese segmento de sondaje. Al multiplicar este porcentaje por el área visible de ese segmento, se obtiene el área correspondiente a la veta.

3. RESULTADOS Y ANÁLISIS DE RESULTADOS

Los resultados presentados a continuación corresponden a las métricas calculadas sobre cada matriz de confusión resultante de las predicciones realizadas en la Subsección 3.3.4 dentro del procedimiento experimental.

3.1 PREDICCIÓN DE LITOLOGÍA

Como el objetivo del entrenamiento de estos modelos es el diferenciar la roca marga, por lo cual, si bien en casos como el de multi litología tiene más de una categoría, la comparación se hará con respecto únicamente a la litología marga. En la Tabla 5 se resumen los resultados finales de cada método de entrenamiento. En caso de querer conocer los datos exactos, puede consultar el Apéndice C.

Tabla 5

Resumen de los resultados obtenidos de los tres tipos de entrenamientos realizados para la clasificación de litología marga

Factor	Multi litología	Binaria	Binaria por elemento geotécnico
Exactitud	71	76	79
Sensibilidad	61	52	59
Especificidad	74	84	86
Precisión	45	54	59
VPN	85	84	86

Nota. Los números expresados en esta tabla representan porcentajes (%).

3.1.1 ETIQUETA MULTILITOLOGÍA (MODELO 1.1)

A pesar de haber múltiples categorías como se observa dentro de los resultados en bruto expresados en las Tabla C1 y Tabla C2, nos centraremos en los resultados de las métricas con respecto a la litología marga resumidas en la Tabla 5. Cabe recordar que el bajo rendimiento de las otras etiquetas distintas a marga no es realmente representativo, ya que no se tiene real seguridad de que estén correctamente etiquetadas.

Con respecto a la categoría marga, el resultado más destacable de este método de etiqueta es su sensibilidad, alcanzando el mejor resultado de los tres métodos con un 61%, logrando predecir correctamente imágenes con alta presencia de brillo, en comparación a sus competidores. Sin embargo, obtuvo la menor sensibilidad con un 45%, haciendo menos confiable su predicción, confundándose con las categorías ARCA, CG, ARCO, TCR, BVC, ARC, especificadas en la Tabla C10.

3.1.2 ETIQUETA BINARIA (MODELO 1.2)

En este caso la predicción posible corresponde únicamente a litología de interés u otra litología. En comparación al método multi litología, aquí obtenemos un aumento en la especificidad en 10% y la precisión en un 9%, resultando en un menor número de errores al etiquetar como verdadero positivo y una menor cantidad de falsos positivos. Sin embargo, se redujo un 11% en la sensibilidad, haciendo que varios casos de etiqueta marga sean catalogados como otra categoría. Las litologías más confundidas con la de interés corresponden a la ARCA, CG y ARCO especificadas en la Tabla C10.

3.1.3 ETIQUETA BINARIA SEGÚN ESTADO GEOTÉCNICO

Para este caso, fue utilizado el modelo 1.3.F para la predicción previa del estado geotécnico para luego aplicar el modelo correspondiente a cada caso.

3.1.3.1 Predicción de la litología por cada estado geotécnico

El resultado final se muestra en la columna *General* dentro de la tabla 6, en donde además para poder hacer una comparación, se filtraron los resultados de los dos modelos anterior con respecto a los mismos tramos que los utilizados en esta sección (Tabla C9), para poder comprobar si ayuda separar esa categoría o no. El resumen de los resultados obtenidos por cada elemento geotécnico se puede observar en la Tabla 6, junto a su respectivo puesto con respecto al rendimiento de los otros dos métodos de etiquetado. Dentro de los tramos catalogados como intacta, el resultado fue significativamente superior, mientras que en las categorías como casi intacta, fracturada y molida se tienen mejoras más leves y en el caso puntual de la categoría borde se tuvo una disminución del rendimiento.

Tabla 6

Resumen de los resultados obtenidos por cada modelo entrenado según un estado geotécnico

Factor	Intacta		Casi Intacta		Fracturada		Molida		Borde		General	
	R	S (%)	R	S (%)	R	S (%)	R	S (%)	R	S (%)	R	S (%)
Exactitud	1	78	1	80	2	79	1	86	1	73	1	79
Sensibilidad	2	62	2	50	1	62	2	44	3	22	2	59
Especificidad	1	87	2	85	2	82	1	91	1	94	1	86
Precisión	1	73	1	36	1	40	1	40	1	60	1	59
VPN	1	80	1	91	1	92	1	92	2	74	1	86

Nota. La columna *R* representa el ranking general de los modelos entrenados según el estado geotécnico con respecto a los otros dos métodos. *S* es el porcentaje de acierto obtenido.

Si consideramos el rendimiento general, se obtuvo una mejoría en todas las categorías con respecto a los otros dos modelos, con la excepción de la sensibilidad, la cual estuvo 3% por debajo de la categoría multi litología.

3.2 DETECCION Y ESTIMACION DE VETAS

3.2.1 DETECCIÓN DE VETAS (MODELO 2)

El principal objetivo de este paso es tener la mínima cantidad de falsos negativos posibles en las etiquetas de vetillada y vetillada masiva, con el objetivo de que no se quede fuera ningún recorte que contenga la presencia de vetas para que pasen a la segmentación.

Existe un rendimiento promedio sobre el 90% (Tabla 7), sin embargo, existen casos puntuales que inducen a errores. Dentro de los factores que influyen negativamente está el exceso de brillo en la superficie del sondaje y la roca clara que se tiende a confundir con zonas de vetas masivas.

Tabla 7

Resumen de los resultados obtenidos en la diferenciación de variaciones dentro de la litología marga

Factor	clara	masiva	oscura	vetillada	General
Exactitud (%)	95	97	97	95	96
Sensibilidad (%)	86	100	92	97	92
Especificidad (%)	98	97	100	94	97

Precisión (%)	95	34	99	91	94
VPN (%)	95	100	96	98	97

3.2.2 CÁLCULO DE PORCENTAJE DE VETAS (MODELO 3)

El procesamiento del cálculo de vetas toma aproximadamente 6 segundos por metro, obteniendo los resultados expresados en la Tabla 8.

Tabla 8

Resumen de los resultados obtenidos de la segmentación aplicada en el grupo de prueba.

Factor	Segmentación (%)
Exactitud	97
Sensibilidad	68
Especificidad	99
Precisión	83
VPN	97

Dentro de los casos vistos durante la realización de este trabajo, se pueden distinguir algunas características que hacen más probable obtener un comportamiento inusual de las predicciones. Dentro de la roca marga oscura u oscura vetillada, la segmentación presenta niveles altos de aciertos, sin embargo, existen casos puntuales en donde el brillo es tan intenso que es confundido con el blanco de las vetas. Por otro lado, cuando la roca es clara o con tonalidades muy cercanas a la de las vetas se produce problemas al segmentarla, sobre todo si hay presencia de brillo. La roca masiva es confundida con roca clara o vice versa. Otro caso que genera error dentro de la predicción es la presencia de canaleta, debido a su color cercano al blanco. Se pueden ver ejemplo de los casos en la Tabla D3.

3.3 REANÁLISIS DE LA BASE DE DATOS

Si observamos la frecuencia de certeza dentro de las predicciones de litología con etiqueta binaria y separada según su estado geotécnico (*Figura 15*), podemos diferenciar los cuatro escenarios presentes dentro de la matriz de confusión. Cuando la certeza es superior al 50% puede ser un verdadero

positivo o un falso positivo. Si la certeza es menor al 50% puede corresponder a un verdadero negativo o un falso negativo.

La mayoría de los verdaderos positivos tienen altos niveles de certeza (entre 80 y 100%), sin embargo dentro de este mismo rango se concentra también la mayor cantidad de supuestos falsos positivos, no obstante, al observar los casos dentro de esta categoría, es posible encontrar muchos ejemplos de roca potencialmente marga, pero también se observa la presencia de una roca granulada muy distinta (Tabla D4). Volviendo a analizar los datos de entrenamiento, se observan una docena de casos de roca con una mineralización muy distinta a todo lo visto anteriormente que pudo haber confundido al modelo (Tabla D5).

El rango entre 50% y 80% se compone principalmente de casos de roca fracturada, casi intacta o con mayor presencia de brillo. A menor certeza es posible observar casos de verdaderos positivos en donde evoca duda si es marga u otra litología que se diferencia por pequeños detalles como la toba cristalina. Dentro de los falsos positivos se continúan viendo patrones granulados que claramente no deberían ser catalogados como margas, sin embargo, hay muchos otros casos de roca similar a la marga clara y con materia orgánica.

Dentro del rango entre 20 a 50%, los supuestos verdaderos negativos se observan casos de potencial marga clara y patrones nuevamente parecidos a los de la Tabla D5. Dentro de los falsos negativos se observa roca oscura con vetas finas, marga clara con tonalidades más oscuras de lo habitual y recortes con baja resolución.

Dentro del rango 0 a 20% no se pudo encontrar ningún caso que evocara duda sobre si es un verdadero negativo. Sin embargo, aquí también se concentra la mayor cantidad de casos de falsos negativos, siendo generalmente roca marga que fue catalogada como otra. Las características principales de esta categoría es roca con excesiva presencia de brillo y roca con pequeños granos, muy parecida a la descripción que se tiene de la toba cristalina. También se encuentran aquí varios casos de segmentos macizos de calcita, pudiendo confundirse con una roca completamente blanca. Igualmente hay un grupo considerable de roca marga clara con un aspecto más cercano a la roca córnea.

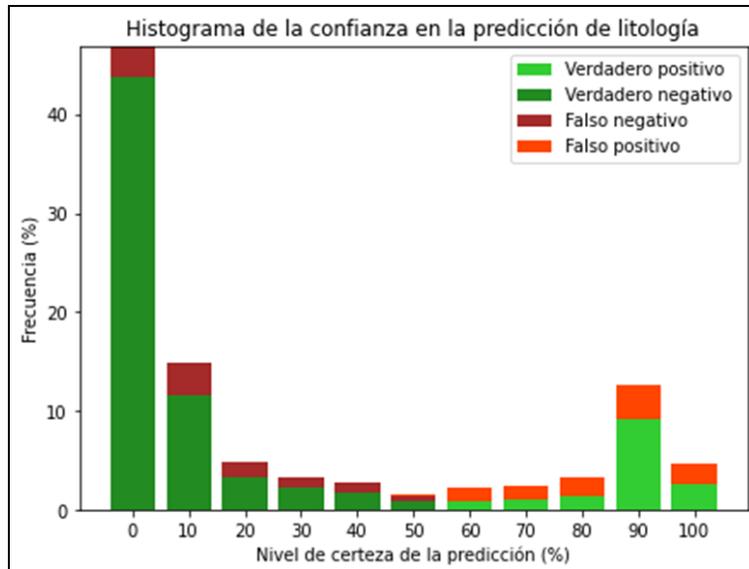


Figura 15 Frecuencia de la confianza de las predicciones de la categoría marga (Elaboración propia).

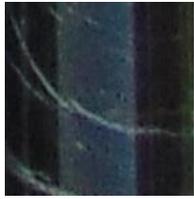
3.4 RESULTADO FINAL

La combinación de todos los resultados expuestos permite el diseño de tablas como la expuesta a continuación (Tabla 9), en donde se describe la litología, su alteración, la presencia de mineralización y el estado geotécnico en un nivel muy básico.

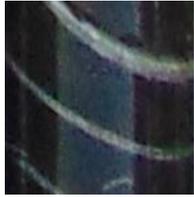
Tabla 9

Ejemplo del resultado final obtenido después de unir las inferencias de la predicción geotécnica, litológica, variación de litología y segmentación de vetas.

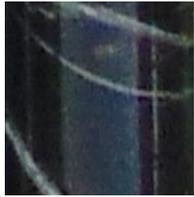
Imagen de entrada	Segmentación	Veta (cm ²)	Mineralización	Estado geotécnico	Litología predicha	Nivel de confianza (%)
		-	Oscura	Fracturada	Marga	98
		1.55	Vetillada	Fracturada	Marga	99



0.49 Vetillada Intacta Marga 90



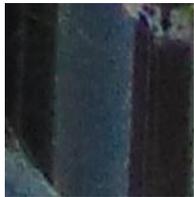
2.65 Vetillada Intacta Marga 90



1.77 Vetillada Intacta Marga 90



1.69 Vetillada Fracturada Marga 99



0.34 Vetillada Casi intacta Marga 98

4. CONCLUSIONES

La aplicación de estas metodologías demostró ser una potencial herramienta para realizar un perfil coherente de cada litología presente. Debido a que la tarea de crear una base de datos de entrenamiento conlleva la incorporación de la mayor representatividad de la variabilidad dentro de los posibles casos, resulta sencillo ver que dentro de la descripción manual de los sondajes no se sigue una descripción coherente a través del tiempo.

Luego de analizar y separar cada descripción litológica entregada por el equipo de geología, es posible concluir que existe un pequeño nivel de inconsistencia que no permite describir correctamente cada sector. Dentro de la descripción es difícil captar los tramos pequeños o presencias de transiciones litológicas. Las limitaciones de la estimación visual traen errores al categorizar zonas de transición debido a la presencia de zonas con porcentajes variables de cada litología transicional. Esto provoca que la descripción geológica normalmente utilizada dentro de este proyecto minero no sea apta como insumo para el entrenamiento de un modelo efectivo de *machine learning*, ya que es requerida mayor precisión dentro de la categorización de cada segmento.

Al comparar la estimación de vetas a través de segmentación con respecto a la estimación visual del equipo de geología, se encuentra una correspondencia dentro de los rangos estipulados. No obstante, con la metodología manual incorpora incerteza sobre la posición y la magnitud, ya que la variación del rango más alto va desde 4 hasta 20 cm² por cada 4 cm.

A continuación, se detallarán conclusiones dentro del desarrollo de cada objetivo específico.

4.1 CLASIFICACIÓN LITOLÓGICA

Con respecto al método de entrenamiento, es posible concluir que cuando se utiliza la etiqueta binaria se tiene un número de casos suficientes para ambas etiquetas en comparación al método multi litología, ya que es común que algunas categorías se vean subrepresentadas debido a su presencia esporádica en los sondajes seleccionados. Sin embargo, la etiqueta binaria suele tener una menor sensibilidad que el caso multi litología, aunque con mayor precisión, significando que la etiqueta binaria reconoce menos casos que el multi litología, pero cuando es seleccionada esa etiqueta, tiene mayor probabilidad de estar correcta. Por lo tanto, si no se tiene certeza en la definición de las categorías, es aconsejable realizar la metodología de etiqueta binaria, con el objetivo de poder ver cuáles son los casos que mejor capta y en cuales se equivoca, pudiendo así, corregir la base de datos para entrenar un modelo multi litología bien definido.

El separar patrones generales que son independientes de la litología como lo es el estado geotécnico, ayuda considerablemente a aumentar el rendimiento general, esto puede deberse a que se acota la variación de los ejemplos de prueba. Un punto para destacar es que la predicción del estado geotécnico tuvo buenos resultados a pesar de la variación de litologías, dando luces de que puede ser utilizado para obtener macro características independiente del yacimiento, cosa que puede resultar muy útil para obtener información del recorte sin tener que adaptarlo a ese yacimiento en específico.

Como esta es una metodología de entrenamiento supervisado, se recomienda obtener la información de entrenamiento de forma detallada y orientada específicamente para el entrenamiento de un modelo de redes neuronales, ya que requiere de un detalle en escalas de centímetros para poder catalogar correctamente cada zona y que no se mezcle la litología de interés dentro de la categoría negativa. Normalmente la descripción litológica se hace en tramos de varios metros debido a la necesidad de balancear el detalle de descripción y la velocidad de avance, es común que anomalías encontradas o zonas de pocos centímetros incrustadas dentro de una litología predominante sean ser mencionadas de forma escrita y, por lo tanto, traerán problemas al etiquetar los datos de entrenamiento.

Al analizar detalladamente los resultados obtenidos de las predicciones, es posible observar que existen varios tramos que potencialmente pertenecen a roca de tipo marga, pero fueron catalogadas como otro tipo de roca, por lo tanto esta característica puede servir para hacer pruebas de calidad del *mapeo* realizado anteriormente.

4.2 DETECCIÓN Y ESTIMACIÓN DE VETAS

Luego de detectar la litología de interés, se obtuvieron buenos resultados en la diferenciación de tipos de alteraciones y presencia de mineralización, con estadísticas de acierto superiores al 92%. La principal fuente de error se encuentra en la roca clara sin contraste, confundiendo con vetas masivas de calcita. Sin embargo, logra cumplir su objetivo de seleccionar los recortes que contengan calcita para ser segmentados.

Al momento de entrenar el modelo para segmentar las vetas, es importante tener en consideración las variaciones que se pueden presentar, sobre todo si alguna de estas tiene similitudes de color o patrones con respecto a lo que se busca separar. En este caso nos encontramos con que los principales errores se encuentran en la roca con exceso de brillo, roca clara sin contrastes, roca borde y en los números escritos en plumones naranjos y rojos, no así los escritos en azul.

Debido a que el resultado de esta estimación es el área de vetilla visible, no se recomienda ser utilizado de forma directa, ya que es diferente tener calcita en forma de un trozo esférico de roca, a que tener esa misma área de una veta que corte transversalmente el sondaje. Por lo tanto, se recomienda ponderar esa área con respecto a otros factores como a si el recorte figura como molida, entera, vetillada u otro aspecto que pueda aproximar mejor el resultado a la realidad.

5. REFERENCIAS

- Acronis International. (s.f.). *Acronis*. Recuperado el marzo de 2021, de <https://www.acronis.com/es-mx/articles/google-cloud-platform/>
- Alzubaidi, F., Mostaghimi, P., Swietojanski, P., Clark, S., & Armstrong, R. (2020). Automated lithology classification from drill core images using convolutional neural networks. Sydney: Journal of Petroleum Science and Engineering.
- Barker, R., Barker, S., Cracknell, M., Stock, E., & Holmes, G. (2020). *Quantitative Mineral Mapping of Drill Core Surfaces II: Long-Wave Infrared Mineral Characterization Using μ XRF and Machine Learning*.
- Barreto, R., & Lizarraga, D. (2019). *Modelo de Sistema de Reconocimiento Facial para el Control de la Trata de Personas*. Arequipa: Universidad Tecnológica Del Perú.
- Bharati, M., Liu, J., & F.MacGregor, J. (2004). *Image texture analysis: methods and comparisons*. McMaster University, Department of Chemical Engineering. Ontario: Science Direct.
- Bonnici, N. H. (2008). Relating textural attributes to mineral processing. *developing a more effective approach for the CadiaEast Cu-Au porphyry deposit* (págs. 4-5). Proceedings of the Ninth International Congress for Applied Mineralogy Conference.
- Caté, A., Perozzi, L., Gloaguen, E., & Blouin, M. (2017). *Machine learning as a tool for geologists*.
- Chen, G., Chen, M., Hong, G., Lu, Y., Zhou, B., & Gao, Y. (2020). *A New Method of Lithology Classification Based on Convolutional Neural Network Algorithm by Utilizing Drilling String Vibration Data*.
- Contreras, C. K., Tusa, L., & Gloaguen, R. (2020). *A supervised technique for drill-core mineral mapping using Hyperspectral data*. Online: EGU General Assembly .
- Contreras, C., Khodadadzadeh, M., Tusa, L., Ghamisi, P., & Gloaguen, R. (2018). *A Machine Learning Technique for Drill Core Hyperspectral Data Analysis*. Amsterdam: Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS).

Contreras, C., Khodadadzadeh, M., Tusa, L., Ghamisi, P., & Gloaguen, R. (2018). *A MACHINE LEARNING TECHNIQUE FOR DRILL CORE HYPERSPECTRAL DATA ANALYSIS*. Amsterdam.

Contreras, I., Khodadadzadeh, M., Tusa, L., & Gloaguen, R. (2019). A Machine Learning Framework for Drill-Core Mineral Mapping Using Hyperspectral and High-Resolution Mineralogical Data Fusion. *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING*, 1-2.

E-CONSTRUIR. (s.f.). *E-CONSTRUIR*. Recuperado el 01 de abril de 2021, de <http://e-construir.com/geotecnia/litologia.html>

Espinoza, B. (2019). *EVALUACION DEL USO DE REDES NEURONALES PARA LA ESTIMACION DE LEYES COMO ALTERNATIVA A LA GEOESTADISTICA*. Universidad de Talca, Curicó.

Gandhi, S., & Sarkar, B. (2016). Drilling. En *Essentials of mineral exploration and evaluation* (págs. 199-210). New York, NY, USA: Elsevier.

García, F. (2018). *CLASIFICACIÓN DE FALLAS EN RODAMIENTOS DE MÁQUINAS ROTATIVAS UTILIZANDO APRENDIZAJE DE MÁQUINAS*. UNIVERSIDAD NACIONAL DE INGENIERÍA.

González, A. (05 de Marzo de 2021). *Claver Data*. Obtenido de <https://cleverdata.io/que-es-machine-learning-big-data/>

Google. (s.f.). *Google Cloud*. Recuperado el Marzo de 2021, de <https://cloud.google.com/automl-tables/docs/prepare>

Google. (s.f.). *Google Colab*. Recuperado el marzo de 2021, de https://colab.research.google.com/notebooks/intro.ipynb?hl=es#scrollTo=5fCEDCU_qrC0

Greenspan, H. (2009). *Super-Resolution in Medical Imaging*. Israel: The British Computer Society.

Hutter, F., Caruana, R., Bardenet, R., Bilenko, M., Guyon, I., Kegl, B., & Larochelle, H. (2018). AutoML 2014 @ ICML". AutoML 2014 Workshop @ ICML.

Iberdrola. (s.f.). *Iberdrola*. Recuperado el 02 de abril de 2021, de <https://www.iberdrola.com/innovacion/que-es-inteligencia-artificial>

- J. Cracknell, M. (2014). *MACHINE LEARNING FOR GEOLOGICAL MAPPING: ALGORITHMS AND APPLICATIONS*. University of Tasmania, School of Physical Sciences.
- Kappeler, A., Yoo, S., Dai, Q., & Katsaggelos, A. (2016). Video Super-Resolution With Convolutional Neural Networks. *IEEE Transactions on Computational Imaging*.
- Kexiong, W., & Laibin, Z. (2008). *Predicting formation lithology from log data by using a neural network*. Beijing.
- Leddy, P. (2014). *ANÁLISIS DEL RIESGO DE INVERSIÓN DE UN PROYECTO DE EXPLORACIÓN DE VETAS ANGOSTAS DE ORO*. UNIVERSIDAD DE CHILE, DEPARTAMENTO DE INGENIERIA DE MINAS, Santiago.
- Luca AI Powered Decisions. (s.f.). *WayBackMachine*. Recuperado el marzo de 2021, de <https://web.archive.org/web/20200224120525/https://luca-d3.com/es/data-speaks/diccionario-tecnologico/python-lenguaje>
- Maitre, J., Bouchard, K., & Bédard, L. (2019). Mineral grains recognition using computer vision and machine learning. *Computers & Geosciences*, 84-93.
- Martinez, J. (10 de 2020). *Iartificial*. Obtenido de <https://www.iartificial.net/redes-neuronales-generativas-adversarias-gans/>
- Martinez, P., Voisin, L., & McFarlane, A. (2015). Hyperspectral logging for mineral deposits and its applications to mining. *XIV Congreso Geológico Chileno*.
- Marulanda, L. (2016). *ALGORITMO DE MAPEO DE TEXTURAS GEOLÓGICAS EN TESTIGOS DE SONDAJES UTILIZANDO CLASIFICADOR SECUENCIAL EN CADENA BASADO EN TÉCNICAS DE TRANSFORMADA DE SEÑALES*. UNIVERSIDAD DE CHILE, DEPARTAMENTO DE INGENIERÍA DE MINAS, Santiago.
- Matich, D. (2001). *Redes Neuronales: Conceptos Básicos y Aplicaciones*. Rosario: Universidad Tecnológica Nacional.
- Merembayev, T., Yunussov, R., & Yedilkhan, A. (2018). *Machine Learning Algorithms for Classification Geology Data from Well Logging*. Kaskelen.

- Microsoft. (10 de 2020). Obtenido de <https://docs.microsoft.com/en-us/azure/machine-learning/concept-automated-ml>
- Müller, M. U., Ekhtiari, N., Almeida, R. M., & Rieke, C. (2020). *Super-resolution of multispectral satellite images using convolutional neural networks*. Nice: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences.
- Naranjo, D. (2005). *Modelo de riesgo para la evaluación económico financiera de proyectos mineros*. Universidad Politecnica de Madrid, Madrid.
- New Pacific Metals Corp. (2018). *New Pacific Metals*. Obtenido de <https://newpacificmetals.com/mining-101/mineral-estimates-reserves-vs-resources>
- Pérez-Barnuevo, L., Lévesque, S., & Bazin, C. (2018). *Automated recognition of drill core textures: A geometallurgical tool for mineral processing prediction*. Minerals Engineering.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). *You Only Look Once: Unified, Real-Time Object Detection*. University of Washington.
- Riesenhuber, M., & Poggio, T. (2000). *Models of Object Recognition*. Nature America Inc.
- Rodríguez-Sahagún, P. (2018). *Aplicación de redes neuronales convolucionales y recurrentes al diagnóstico de autismo a partir de resonancias magnéticas funcionales*. Universidad Politécnica de Madrid, Escuela Técnica Superior de Ingenieros, Madrid.
- SERNAGEOMIN. (2019). cuadro 49. En *Anuario de la Minería de Chile 2018* (pág. 66). Santiago.
- Simeone, O. (2018). *A Very Brief Introduction to Machine Learning With Applications to Communication Systems*. United Kingdom: IEEE.
- Singh, H., Seol, Y., & Myshakin, E. (2020). *Automated Well-Log Processing and Lithology Classification by Identifying Optimal Features Through Unsupervised and Supervised Machine-Learning Algorithms*. SPE Journal.
- Spyder Website Contributors. (s.f.). *Spyder*. Recuperado el marzo de 2021, de <https://www.spyderide.org/>
- Suárez, J. (1998). *Deslizamientos y estabilidad de taludes en zonas tropicales*. Universidad Industrial de Santander, UIS.

- Tiu, G. (2017). *Classification of Drill Core Textures for Process Simulation in Geometallurgy*. Luleå University of Technology, Department of Civil, Environmental and Natural Resources Engineering.
- Valenzuela, A. (2015). *Equipo de sondaje de aire reverso para la investigación de minerales*. Valparaiso: Escuela de Ingeniería Mecánica.
- van Gerven, M., & Bohte, S. (eds. 2018). *Artificial Neural Networks as Models of Neural Information Processing*. Netherlands: Centrum Wiskunde & Informatica.
- Verdugo H., J. J. (2016). *APORTE DE LAS CAMPAÑAS DE RECONOCIMIENTO GEOLÓGICO AL VALOR DEL PLAN MINERO CODELCO-ANDINA*. UNIVERSIDAD DE CHILE, DEPARTAMENTO DE INGENIERÍA INDUSTRIAL, Santiago.
- Zhang, L., Zhang, H., Shen, H., & Li, P. (2010). A super-resolution reconstruction algorithm for surveillance images. Wuhan: Signal Processing.

Apéndice A

La figura A.1 presenta el flujo de trabajo realizado en el procedimiento experimental, observándose principalmente 4 fases, el pretratamiento de los datos, el entrenamiento de los modelos de clasificación de litologías, el entrenamiento de los modelos de detección y clasificación de vetas, y la aplicación de estos modelos en nuevos sondajes.

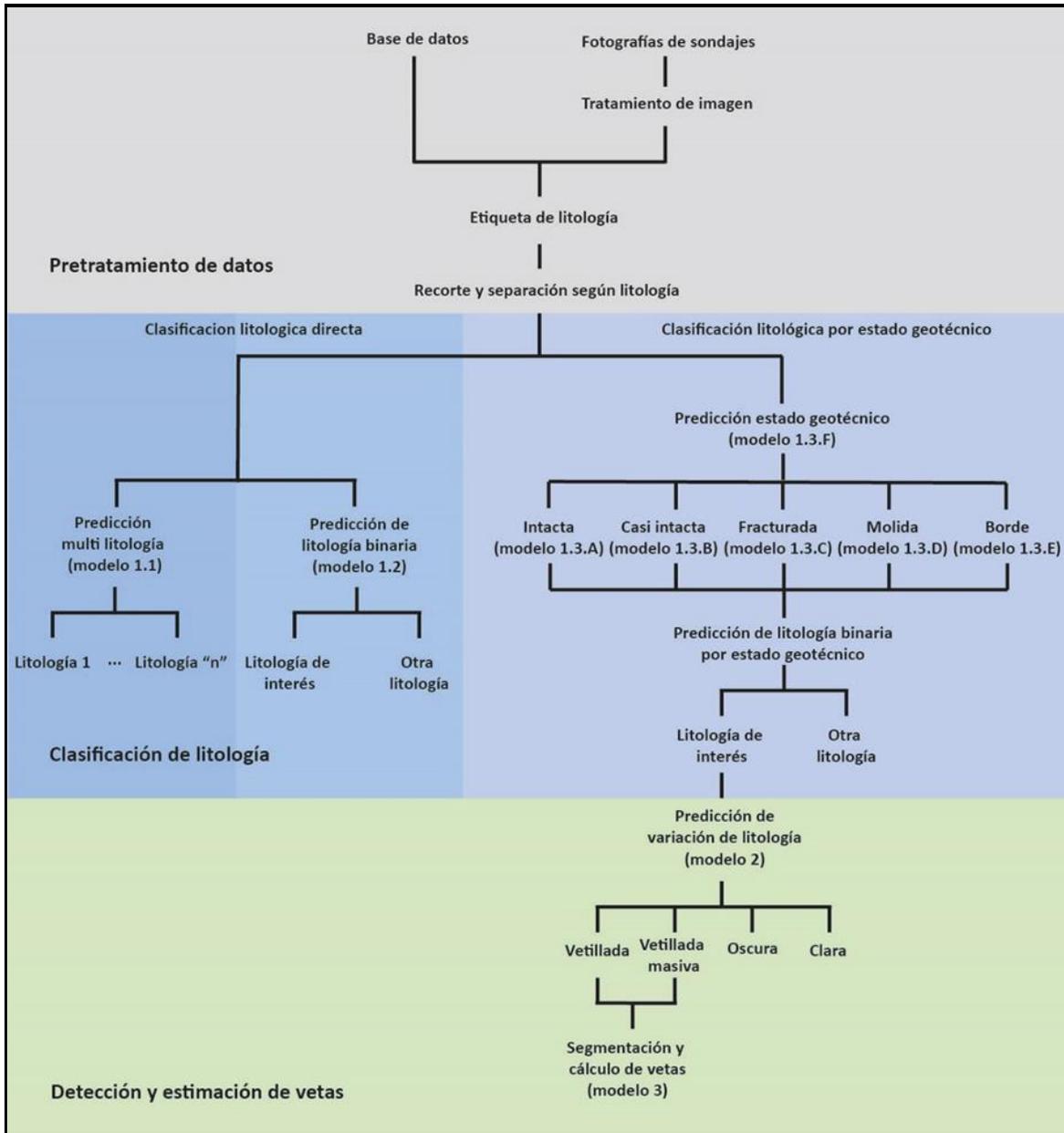


Figura A.1 Flujo de trabajo realizado en el procedimiento experimental

Apéndice B

A continuación, se muestran las litologías presentes dentro de los datos de entrenamiento (Tabla B1) y los de prueba (Tabla B2).

Tabla B.1

Longitud lineal de las litologías presentes dentro de los datos de entrenamiento

Sondaje	MAR	TCR	ARCA	ARCO	COR	ARC	ARE	DP	ART	CG	Longitud total
A	45.1	0	0	0	0	0	0	0	0	0	45.1
B	16.3	0	0	0	0	0	0	0	12.5	5.3	34.2
C	13.1	63.9	35.6	3.6	41.1	20.4	2.6	1.3	0	17.9	199.9
D	19.5	33.2	40.3	22.3	3.4	29.6	6.0	4.0	8.9	9.3	176.8
E	25.3	0	0	0	0	0	0	0	0	0	25.31
F	17.3	25.6	25.1	11.4	1.2	5.8	27.3	11.4	10.1	1.7	137.3
G	12.0	49.4	12.9	17.1	5.6	0	0	0.8	7.4	0	105.4
Presencia total	148.8	172.2	114.1	54.5	51.3	55.9	36.0	17.7	39.0	34.4	724.3

Nota. Los números presentados están expresados en metros. MAR: Marga; TCR: Toba cristalina; ARCA: Arenisca calcárea; ARCO: Arenisca cornea; COR: Cornea; ARC: Arenisca conglomerídica; ARE: Arenisca; DP: Dique porfídico; CG: Conglomerado.

Tabla B.2

Longitud lineal de las litologías presentes dentro de los datos de prueba

Sondaje	MAR	TCR	ARCA	ARCO	COR	ARC	ARE	DP	ART	CG	COQ	Longitud total
A	50.1	0	0	0	0	0	0	0.8	0	12.0	0	62.9
B	9.5	33.7	51.9	0	0	19.7	34.6	10.8	1.2	11.0	0.6	173.0
C	42.1	0	0	2.9	0	0	2.4	0.6	0	0	0	48.0
D	16.1	24.1	27.7	27.1	8.8	7.7	8.4	17.5	9.8	1.9	3.3	152.4
Presencia total	117.8	57.8	79.6	30.0	8.8	27.4	45.4	29.7	11.0	24.9	3.9	436.3

Nota. Los números presentados están expresados en metros. MAR: Marga; TCR: Toba cristalina; ARCA: Arenisca calcárea; ARCO: Arenisca cornea; COR: Cornea; ARC: Arenisca conglomerídica; ARE: Arenisca; DP: Dique porfídico; CG: Conglomerado; COQ: Coquina.

Apéndice C

Dentro de este apartado se encuentran los resultados obtenidos dentro de la clasificación litológica. Dentro de los tres métodos de etiquetado probado se presentan dos tablas, la primera exponiendo la matriz de confusión con las predicciones obtenidas (Tablas C1, C3, C4, C5, C6, C7 y C8), luego se presentan las métricas de rendimiento de los modelos (Tablas C2 y C9).

Tabla C1

Matriz de confusión producto de la clasificación de los datos de prueba con etiqueta multi litología

	MAR	TCR	COQ	ART	ARCA	COR	ARCO	ARE	CG	ARC	DP	BVC	Total real
MAR	1516	386	3	82	93	43	2	331	7	6	11	0	2480
TCR	182	434	6	29	155	20	10	203	35	5	57	0	1136
COQ	33	8	14	1	1	0	0	3	1	0	0	0	61
ART	39	46	2	11	5	16	0	75	0	14	0	0	208
ARCA	522	114	10	25	554	124	33	133	38	34	6	0	1593
COR	6	13	0	0	9	28	0	102	0	13	0	0	171
ARCO	287	14	5	2	85	11	57	12	78	2	1	0	554
ARE	181	149	4	22	328	67	9	41	33	17	32	0	883
CG	237	55	4	4	26	59	6	4	97	0	8	0	500
ARC	101	42	10	31	82	80	9	193	5	34	5	0	592
DP	161	203	23	3	86	7	15	31	10	6	48	0	593
BVC	101	226	0	28	144	68	16	21	172	7	65	0	848
Predicciones totales	3366	1690	81	238	1568	523	157	1149	476	138	233	0	9619

Nota. MAR: Marga; TCR: Toba cristalina; ARCA: Arenisca calcárea; ARCO: Arenisca cornea; COR: Cornea; ARC: Arenisca conglomerádica; ARE: Arenisca; DP: Dique porfídico; CG: Conglomerado; COQ: Coquina.

Tabla C2

Métricas de rendimiento calculadas sobre la matriz de confusión de clasificación con etiqueta multi litología

	VP	VN	FP	FN	Ex (%)	Se (%)	Esp (%)	Pre (%)	VPN (%)
MAR	1516	5289	1850	964	71	61	74	45	85
TCR	434	7227	1256	702	80	38	85	26	91
COQ	14	9491	67	47	99	23	99	17	100
ART	11	9184	227	197	96	5	98	5	98
ARCA	554	7012	1014	1039	79	35	87	35	87
COR	28	8953	495	143	93	16	95	5	98
ARCO	57	8965	100	497	94	10	99	36	95
ARE	41	7628	1108	842	80	5	87	4	90
CG	97	8740	379	403	92	19	96	20	96
ARC	34	8923	104	558	93	6	99	25	94
DP	48	8841	185	545	92	8	98	21	94

BVC	0	8771	0	848	91	0	100	0	91
Predicciones totales	3366	1690	81	238	1568	157	1149	476	138

Nota. VP: Verdaderos positivos, VN: Verdaderos negativos; FP: Falsos positivos; FN: Falsos negativos; Ex: Exactitud; Se: Sensibilidad; Esp: Especificidad; Pre: Precisión, VPN: Valor de predicción negativo; MAR: Marga; TCR: Toba cristalina; ARCA: Arenisca calcárea; ARCO: Arenisca cornea; COR: Cornea; ARC: Arenisca conglomerádica; ARE: Arenisca; DP: Dique porfídico; CG: Conglomerado; COQ: Coquina.

Tabla C3

Matriz de confusión producto de la clasificación de los datos de prueba con etiqueta binaria

Observación	Predicción		Total real
	Marga	Otra litología	
Marga	1298	1182	2480
Otra litología	1114	6025	7139
Predicciones totales	2412	7207	9619

Tabla C4

Matriz de confusión producto de la clasificación binaria de los datos de prueba clasificados como roca intacta

Observación	Predicción		Total real
	Marga	Otra litología	
Marga	1071	670	1741
Otra litología	402	2699	3101
Predicciones totales	1473	3369	4842

Tabla C5

Matriz de confusión producto de la clasificación binaria de los datos de prueba clasificados como roca casi intacta

Observación	Predicción		Total real
	Marga	Otra litología	
Marga	100	102	202
Otra litología	178	1011	1189
Predicciones totales	278	1113	1391

Tabla C6

Matriz de confusión producto de la clasificación binaria de los datos de prueba clasificados como roca fracturada

Observación	Predicción		Total real
	Marga	Otra litología	
Marga	251	155	406

Otra litología	377	1768	2145
Predicciones totales	628	1923	2551

Tabla C7

Matriz de confusión producto de la clasificación binaria de los datos de prueba clasificados como roca molida

Observación	Predicción		Total real
	Marga	Otra litología	
Marga	34	43	77
Otra litología	51	523	574
Predicciones totales	85	566	651

Tabla C8

Matriz de confusión producto de la clasificación binaria de los datos de prueba clasificados como roca borde

Observación	Predicción		Total real
	Marga	Otra litología	
Marga	12	42	54
Otra litología	8	122	130
Predicciones totales	20	164	184

Tabla C9

Resultados de los tres métodos de etiqueta comparados entre sí

Factor	Intacta			Casi intacta			Fracturada			Molida			Borde		
	ML	Bi	BG	ML	Bi	BG	ML	Bi	BG	ML	Bi	BG	ML	Bi	BG
Ex	67	72	78	70	80	80	72	80	79	74	83	86	68	68	73
Se	64	55	62	56	46	50	56	49	62	49	39	44	48	26	22
Esp	68	82	87	73	86	85	76	86	82	77	89	91	77	86	94
Pre	56	63	73	29	35	36	32	39	40	26	33	40	50	44	60
VPN	75	77	80	90	90	91	89	90	92	90	92	92	76	74	74

Nota. Todos los números representan porcentajes. Las casillas en verde representan al mejor resultado en esa métrica y dentro de un estado geológico específico. Ex: Exactitud; Se: Sensibilidad; Esp: Especificidad; Pre: Precisión, VPN: Valor de predicción negativo; ML: Multi litología; Bi: Binaria; BG; Binaria espada por estado geológico.

Tabla C10

Resumen de los datos etiquetados como Marga, pero que pertenecen a otra categoría

Etiqueta	CG	DP	ARCO	ARE	ARCA	TCR	COQ	ART	BVC	ARC	COR	total
ML	237	161	287	181	522	182	33	39	101	101	6	1850
Bi	223	57	181	81	360	81	17	31	39	43	1	1114
BG	141	52	132	79	328	126	22	31	59	31	4	1005

Nota. ML: Multi litología; Bi: Binaria; BG; Binaria espada por estado geológico; TCR: Toba cristalina; ARCA: Arenisca calcárea; ARCO: Arenisca cornea; COR: Cornea; ARC: Arenisca conglomerídica; ARE: Arenisca; DP: Dique porfídico; CG: Conglomerado; COQ: Coquina.

Apéndice D

Al igual que en el apartado anterior, a continuación, se presentan los resultados obtenidos dentro de la clasificación de alteraciones y mineralización (Tablas D1 y D2) y los resultados de la segmentación de vetas (Figura D1).

Tabla D1

Matriz de confusión producto de la clasificación según la alteración y mineralización de los tramos catalogados como marga de los datos de prueba

Observación	Predicción				Total real
	Clara	Masiva	Oscura	Vetillada	
Clara	540	48	0	41	629
Masiva	0	37	0	0	37
Oscura	24	1	770	44	839
Vetillada	5	22	4	901	932
Predicciones totales	569	108	774	986	2437

Tabla D2

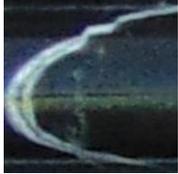
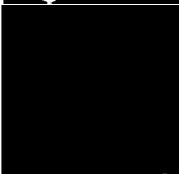
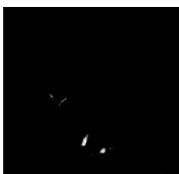
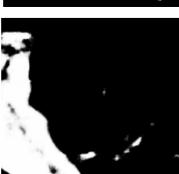
Métricas de rendimiento calculadas sobre la matriz de confusión de clasificación con etiqueta de alteración y mineralización.

Predicción	VP	VN	FP	FN	Ex (%)	Se (%)	Esp (%)	Pre (%)	VPN (%)
Clara	540	1779	29	89	95	86	98	95	95
Masiva	37	2329	71	0	97	100	97	34	100
Oscura	770	1594	4	69	97	92	1,00	99	96
Vetillada	901	1420	85	31	95	97	94	91	98
Predicciones totales	3366	1690	81	238	1568	157	1149	476	138

Nota. VP: Verdaderos positivos, VN: Verdaderos negativos; FP: Falsos positivos; FN: Falsos negativos; Ex: Exactitud; Se: Sensibilidad; Esp: Especificidad; Pre: Precisión, VPN: Valor de predicción negativo;

Tabla D3

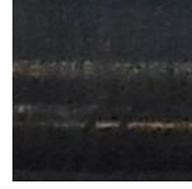
Ejemplos de segmentación de vetas.

Imagen de entrada	Verdad base	Prediccion	Ex (%)	Se (%)	Esp (%)	Pre (%)	VPN (%)
			97	85	100	99	97
			95	62	100	99	95
			90	52	100	100	89
			100	23	100	30	100
			97	89	100	99	96
			88	55	89	13	98
			96	66	99	91	96
			95	94	97	99	86

Nota. Ex: Exactitud; Se: Sensibilidad; Esp: Especificidad; Pre: Precisión, VPN: Valor de predicción negativo;

Tabla D3

Extracto de ejemplos de los escenarios de verdadero negativo y falso negativo, separado según el rango de confianza de la predicción.

[0-10[[10-20[[20-30[[30-40[[40-50[
Verdadero negativo				
				
				
				
				
				
Falso negativo				
				

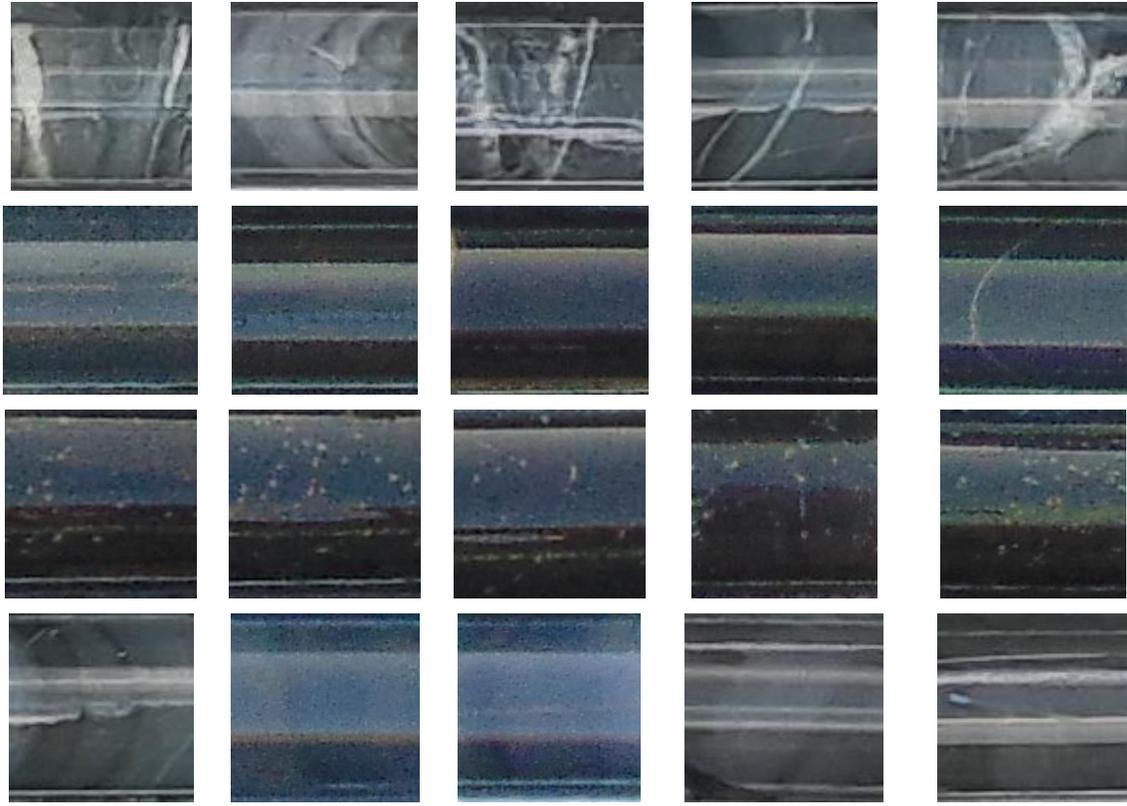
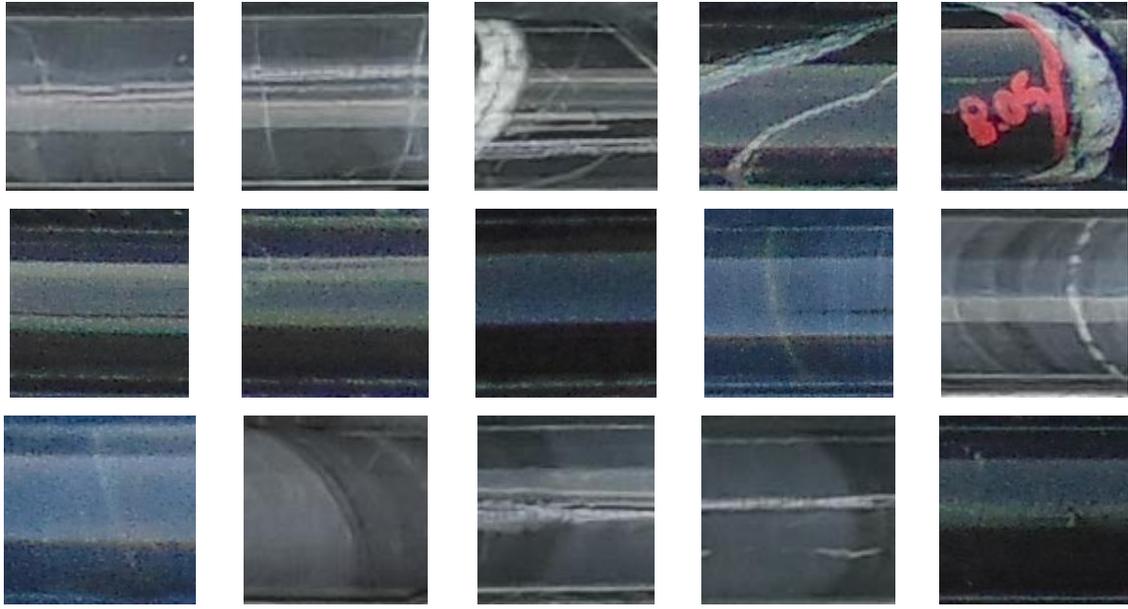


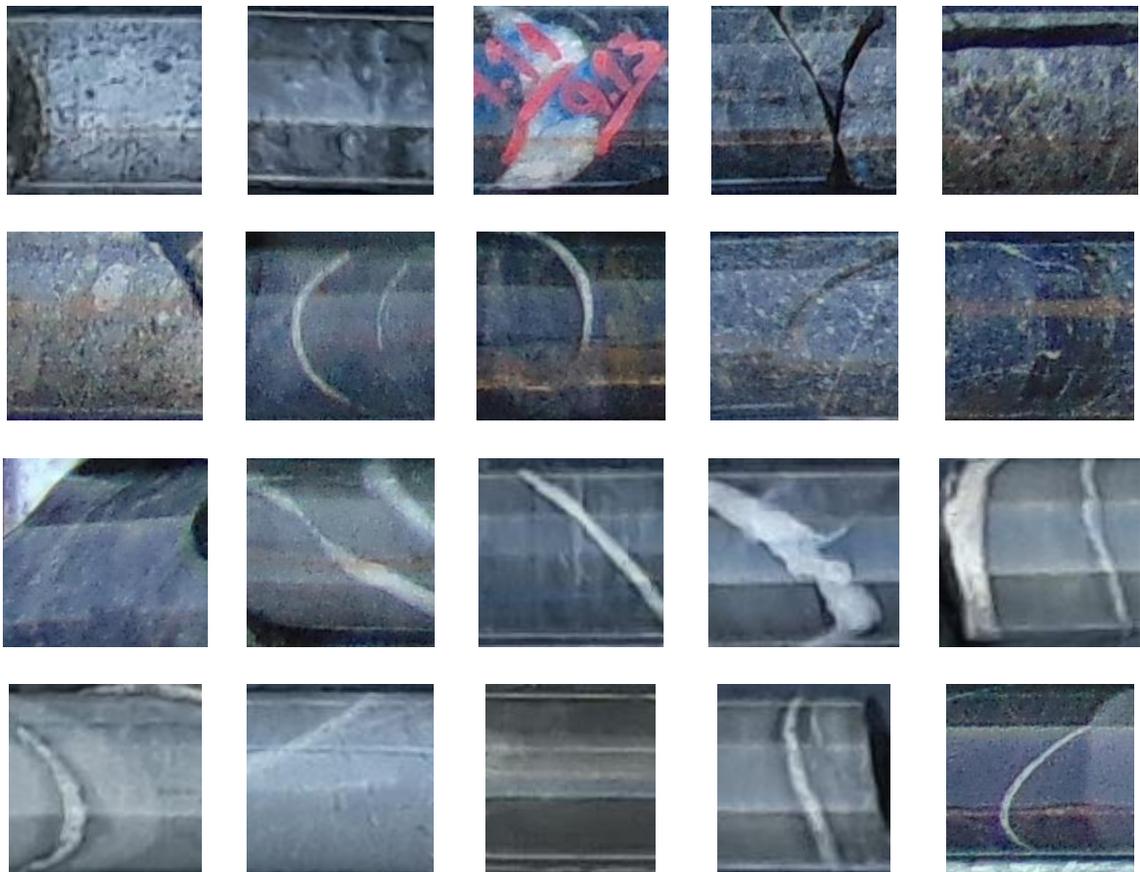
Tabla D4

Extracto de ejemplos de los escenarios de verdaderos positivos y falsos positivos, separado según el rango de confianza de la predicción.

[50-60[[60-70[[70-80[[80-90[[90-100]
Verdadero positivo				



Falso positivo



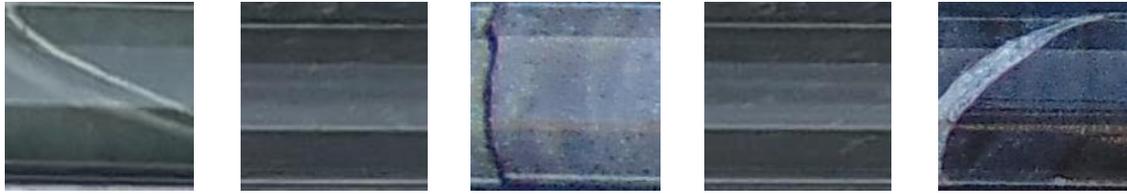


Tabla D5

Casos inusuales dentro de la base de datos de entrenamiento y su posible correlato en las predicciones de falsos positivos

Marga inusual		Falsos positivos	

