

UNIVERSIDAD DE CONCEPCIÓN

Facultad de Ingeniería

Departamento de Ingeniería Metalúrgica

Profesores Patrocinantes

Daniel Sbárbaro Hofer

Eduardo Balladares Varela

**ANÁLISIS QUÍMICO CUANTITATIVO
DE CONCENTRADOS DE COBRE
UTILIZANDO MICROSCOPIA ÓPTICA
MULTIESPECTRAL**

GERMÁN ANDRÉS HENRÍQUEZ VÉJAR

Informe de Memoria de Título

para optar al Título de

Ingeniero Civil Metalúrgico

Enero de 2020

AGRADECIMIENTOS

A mis docentes guía, Dr. Daniel Sbárbaro y Dr. Eduardo Balladares, quienes me apoyaron y ayudaron durante toda la investigación, aportando tanto en ideas como discusiones y detalles que permitieron enfocarme en ciertos detalles.

Al encargado de laboratorio, Dr. Rodrigo Fuentes, quien me ayudó con mucha teoría, montaje de equipos, resolución de dudas, además de enseñarme a utilizar el equipamiento necesario. Sumo a Benjamín y José, quienes trabajan en el laboratorio, tanto por su ayuda en problemas que tenía como por la gran compañía que fueron durante largas horas.

A Walter Díaz, quien me ayudó con el análisis de datos usando PCA, lo que posteriormente me sirvió para continuar con todas las aplicaciones de estos, le deseo mucho éxito en sus proyectos de investigación.

A todos los que durante toda mi estadía por la universidad me han ayudado estudiando conmigo, facilitándome apuntes relevantes o derechamente ayudándome a entender contenidos, los que son demasiados como para nombrar, pero, aun así, los recuerdo a todos.

A muchos de mis amigos por el apoyo, Rocío, Rodrigo, Paula, Alejandra y varios más, quienes me entregaron su apoyo en este proceso. Además, gracias Matías por animarme a la distancia.

Agradecer a Catalina ♥, quien estuvo conmigo durante todo este proceso, apoyándome enormemente durante momentos de trabas.

A mi hermano Esteban, quien pudo proveerme de horas de distracción, quizás demasiadas.

David, gracias por nada y al mismo tiempo gracias por tanto.

A mis padres.

Por último, pero no menor, a todos los voluntarios de Bomberos de Chile, por su gran labor, muchas veces no tan reconocida como debería serlo.

RESUMEN

En el presente informe se propone un método que permitiría realizar análisis mineralógico de concentrados de cobre basado en reflectancia óptica multiespectral, usando iluminación especular (campo claro) e iluminación difusa (campo oscuro).

El presente estudio se dividió en tres etapas, la primera consistió en la obtención de longitudes de onda discriminantes a través del uso de un espectrómetro en el intervalo VISNIR, entre 400 y 920 nm. Se determinó que las longitudes de onda capaces de clasificar muestras de mineral corresponden a 600, 920 y 400 nm, en ese orden de importancia.

En una segunda etapa se crearon modelos de ML con algoritmos KNN y SGD con los mismos minerales. Se probó la validación de los modelos y se determinó que:

- En BF las longitudes de onda discriminantes corresponden a 400, 600 y 900 nm, donde la validación máxima fue de un 71,45% con el algoritmo KNN.
- En DF no se pudo utilizar luz infrarroja, pero dentro del espectro de luz visible, las longitudes de onda que maximizaron la validación del modelo corresponden a 400, 430 y 560 nm.
- Se identificó un efecto producido por la resolución en lo que se cree que es importante la razón entre el tamaño de partícula y el área ocupada por un píxel, sin embargo, esto requiere más estudio.

En una tercera y última etapa se probaron los modelos obtenidos anteriormente en las condiciones que maximizan su validación sobre concentrados de cobre y se obtuvieron los siguientes resultados

- El error promedio asociado a cada elemento corresponde a 18,9%, 60,7% y 40,1% para el cobre, hierro y azufre, respectivamente.
- Se determina que el modelo no está en condiciones de utilizarse para determinar composiciones mineralógicas de concentrados de cobre.

Se sugiere continuar la investigación enfocándose en el efecto del tamaño de partícula y la resolución de las imágenes utilizando iluminación difusa (microscopía de campo oscuro), agregando luz infrarroja además de la luz visible.

ABSTRACT

In this report a method that would allow mineralogical analysis of copper concentrates using multispectral optical reflectance by employing specular illumination (bright field) and diffuse illumination (dark field) microscopy techniques is proposed.

This research was divided in three stages, first is to obtain discriminant wavelengths through the use of a spectrometer in VISNIR light range, between 400 and 920 nm. It was determined that wavelengths capable of classifying ore samples are 600, 920 and 400 nm.

In the second stage, ML models were created by using KNN and SGD algorithms with the same minerals. A cross-validation test was used and it was determined that:

- In bright field microscopy, discriminant wavelengths are 400, 600 and 900 nm. The maximum cross-validation score was 71,45% using the KNN algorithm.
- In dark field microscopy, due to filter limitations, only visible light was tested. Discriminant wavelengths are 400, 430 and 560 nm.
- There are different variance ratios by using different image resolutions, it is believed that the ratio between the particle size and the area occupied by a pixel is an important factor. Nevertheless, this requires further study.

In the last stage, ML models obtained in the second stage were tested on copper concentrates, getting the following results:

- The mean relative error for each element corresponds to 18,9%, 60,7% and 40,1% for copper, iron and sulfur, respectively.
- The current model is not ready to use as a precise way of determining a copper concentrate mineralogical composition.

It is suggested to continue the study focused on the ratio between the particle size and the photography resolution, also using dark field microscopy using visible and infrared light.

ÍNDICE

1	INTRODUCCIÓN	1
1.1.	Generalidades	1
1.2.	Objetivos	2
1.2.1.	Objetivo general	2
1.2.2.	Objetivos específicos	2
2	ANTECEDENTES	3
2.1.	Espectroscopía de reflectancia	3
2.1.1.	Fuentes de luz	3
2.1.2.	Microscopio de reflectancia	5
2.1.3.	Filtros y monocromadores	8
2.1.4.	Espectrómetro y curvas de reflectancia espectral.....	11
2.2.	Cota superior asintótica y orden de complejidad	13
2.3.	Análisis de componentes principales (PCA)	13
2.3.1.	PCA usando SVD	13
2.3.2.	PCA como técnica de exploración	14
2.4.	Aprendizaje supervisado y modelos de clasificación	15
2.4.1.	Conceptos relevantes.....	15
2.4.2.	Árbol de decisiones (DT)	16
2.4.3.	KNN	17
2.4.4.	SVM.....	18
2.4.5.	SGD.....	20
2.4.6.	Clasificación con múltiples clases	21
2.5.	Investigaciones previas	22
2.5.1.	Estudio de automatización de análisis de sulfuros	22
2.5.2.	Estudio de reconocimiento automatizado de menas metálicas.....	23

3	Metodología	24
3.1.	Equipamiento y muestras.....	24
3.1.1.	Primera etapa.....	24
3.1.2.	Segunda etapa	25
3.2.	Montaje experimental.....	25
3.3.	Detalle de muestras	28
3.4.	Desarrollo experimental.....	29
3.4.1.	Primera etapa.....	29
3.4.2.	Segunda etapa (BF).....	30
3.4.3.	Segunda etapa (DF)	31
3.4.4.	Análisis de concentrados	31
4	RESULTADOS Y DISCUSIONES	33
4.1.	Determinación de curvas de reflectancia espectral	33
4.2.	Análisis de curvas de reflectancia espectral y preparación de modelos.....	36
4.3.	Toma y análisis de imágenes de minerales (BF).....	40
4.3.1.	Ajuste de modelos.....	40
4.3.2.	Efecto resolución.....	41
4.4.	Toma y análisis de imágenes de minerales (DF).....	42
4.4.1.	Ajuste de modelos.....	44
4.4.2.	Efecto resolución.....	44
4.5.	Resumen de ajuste de modelos.....	45
4.6.	Resultados de análisis de concentrados	46
5	CONCLUSIONES Y RECOMENDACIONES	49
5.1.	Conclusiones y recomendaciones generales	49
5.2.	Recomendaciones para la continuidad del estudio	50
5.3.	Línea de investigación complementaria	51

6	BIBLIOGRAFÍA Y REFERENCIAS	53
7	ANEXOS	56
7.1.	Abreviaciones de esquemas de luz en microscopios.....	56
7.2.	Figuras	57



NOMENCLATURA

- ML: *Machine learning* (aprendizaje computacional).
- PCA: *Principal Component Analysis* (análisis de componentes principales).
- KNN: *K-Nearest Neighbors*.
- SVM: *Support Vector Machine*.
- SGD: *Stochastic Gradient Descent*.
- BF: *Bright Field [Microscopy]* (Microscopía en campo claro).
- DF: *Dark Field [Microscopy]* (Microscopía en campo oscuro).
- SVD: *Singular Value Decomposition* (Descomposición de valores principales, factorización utilizada en álgebra lineal).
- VISNIR: Espectro de luz visible e infrarrojo cercano. Corresponde a longitudes de onda de 380nm hasta 900nm.
- R[300 nm]: Reflectancia a 300 nm.



ÍNDICE DE FIGURAS

Figura 2.1: espectro emitido por una lámpara de tungsteno-halógena	4
Figura 2.2: esquema de una lámpara de tungsteno-halógena para microscopía	4
Figura 2.3: representación de microscopios utilizados en microscopía de reflectancia ..	5
Figura 2.4: esquema simplificado de una sección transversal de un microscopio	6
Figura 2.5: esquema de una imagen formada por microscopía en campo claro	7
Figura 2.6: esquema de imagen formada por iluminación difusa	8
Figura 2.7: esquema de funcionamiento de un monocromador	9
Figura 2.8: efecto del tamaño de apertura en un monocromador	10
Figura 2.9: esquema de funcionamiento de filtros	11
Figura 2.10: reflectancia espectral difusa de calcopirita y pirita	12
Figura 2.11: dos posibles hiperplanos de separación	19
Figura 2.12: resultados en clasificación para estudio de Chopard	22
Figura 2.13: curva que recoge la variación del plano de enfoque óptimo	23
Figura 3.1: montaje experimental	26
Figura 3.2: espectrómetro Ocean Optics HR2000+ES, monocromador Oriel 7240	27
Figura 4.1: reflectancia de pellets en BF con un área de integración de 40 μm	34
Figura 4.2: reflectancia DF en luz visible, área de integración de 40 μm	34
Figura 4.3: reflectancia VISNIR en campo oscuro	35
Figura 4.4: efecto de la granulometría en pellets de calcopirita.	36
Figura 4.5: <i>scores</i> de tres componentes principales	37
Figura 4.6: <i>loadings</i> en el sentido de varianza explicada.	38
Figura 4.7: árbol de decisiones para las 1125 muestras de minerales	39
Figura 4.8: efecto de la escala de la resolución sobre la validación del modelo BF	41
Figura 4.9: <i>scores</i> de los dos primeros componentes principales en DF	43
Figura 4.10: <i>loadings</i> de los primeros cuatro componentes principales en DF	43
Figura 4.11: escala de la resolución sobre la validación del modelo en DF	45
Figura 5.1: espectros de reflectancia de concentrados de cobre	52
Figura 5.2: espectros de reflectancia de pellets de minerales	52
Figura 7.1: separación de muestras de calcopirita.	57
Figura 7.2: micrografía de calcopirita bajo luz monocromática a 600 nm	57
Figura 7.3: árbol de decisiones detallado	58

ÍNDICE DE TABLAS

Tabla 1: espectros de emisión de distintas fuentes de luz.....	3
Tabla 2: composición de las muestras.....	28
Tabla 3: composición química de muestras de concentrado.....	28
Tabla 4: fracciones elementales de cada mineral.....	32
Tabla 5: Máximos locales de cada componente principal.....	37
Tabla 6: resultados de modelos ML.....	39
Tabla 7: puntajes de validación para cada técnica con distintos modelos.....	45
Tabla 8: Resultados de modelo KNN en iluminación BF.....	47



1 INTRODUCCIÓN

1.1. Generalidades

Uno de los primeros productos obtenidos en una planta metalúrgica cuprífera es el concentrado de cobre, proveniente directamente de etapas de flotación. Este concentrado está compuesto principalmente de calcopirita, seguido de calcosina, covelina, bornita, además de pirita y cuarzo. Dependiendo de la ubicación del yacimiento, puede incluso, contener enargita, como es el caso de División Ministro Hales [1]. El proceso continúa hacia una etapa de fundición, donde finalmente se produce un ánodo de cobre. Conocer composiciones mineralógicas del concentrado es importante tanto en la etapa de flotación, para controlar las variables operacionales, como en la fundición para ajustar el uso de oxígeno.

En la actualidad, se usan técnicas de fluorescencia de rayos X (Outotec Courier) y difracción y espectroscopía de rayos X (FEI QEMSCAN), las cuales poseen un costo elevado y exigen preparaciones de la muestra que pueden ser exigentes en tiempo y/o complejas. Es por esto que se ve la reflectancia óptica multiespectral como una alternativa de menor costo y que podría requerir una preparación de muestra simple y rápida o, potencialmente, nula.

1.2. Objetivos

1.2.1. Objetivo general

Como objetivo general se busca desarrollar un método que obtenga la composición mineralógica de forma cuantitativa para un concentrado de cobre, utilizando fenómenos de reflectancia espectral.

1.2.2. Objetivos específicos

- Evaluar la eficacia de utilizar la reflectancia espectral de cada mineral como una forma de identificarlos en pellets.
- Evaluar métodos para clasificar una partícula de mineral utilizando un número razonable de mediciones.
- Investigar y/o crear un algoritmo que permita clasificar de manera simple la partícula mineral con base en mediciones realizadas.
- Evaluar la eficacia del método, analizando todas las partículas de un concentrado de cobre para obtener una composición química del mismo.
- Cuantificar márgenes de error del sistema de clasificación con respecto a composiciones del concentrado de cobre utilizando otros métodos.

2 ANTECEDENTES

2.1. Espectroscopía de reflectancia

La espectroscopía de reflectancia es un área de la quimiometría que investiga la composición espectral de la radiación reflejada con respecto a la composición espectral incidente primaria. Existen dos casos límite importantes: reflexión regular (especular), o sea, reflectancia de una superficie pulida, y reflexión difusa [2].

2.1.1. Fuentes de luz

Una fuente de luz (también llamada lámpara) provee de la iluminación en forma de rayos de luz incidentes a la muestra. Las más utilizadas corresponden a la halógena, deuterio, xenón y LED. Dependiendo del origen de emisión se producen distintos espectros de luz aplicables dentro de un determinado intervalo, según la Tabla 1, las luces LED presentan distintos rangos de emisión dependiendo de la utilidad, la más utilizada es la LED blanca que emite sólo dentro del espectro visible (400-720 nm).

Tabla 1: espectros de emisión de distintas fuentes de luz. (*) Observado experimentalmente. (S/I) Sin información.

Tipo de fuente	Intervalo de emisión [nm]	Intervalo útil [nm] (*)
Halógena [3]	360 – 2500	400 – 920
Deuterio [4]	190 – 700	300 – 600
Xenón [5]	250 - 1750	(S/I)
LED	Variable	(S/I)

2.1.1.1. Fuente de luz halógena

Una lámpara de tungsteno-halógena, es una fuente de luz incandescente que produce un espectro desde el ultravioleta cercano hasta la luz infrarroja, dependiente de la temperatura del filamento (Figura 2.1) [6,7].

Para su funcionamiento, una corriente eléctrica calienta el filamento de tungsteno, produciendo una serie de reacciones llamadas "ciclo regenerativo de los halógenos",

donde se oxida y vuelve a depositar el tungsteno del filamento en fase gaseosa hasta alcanzar una temperatura estable, donde se emite un espectro de luz. Adicionalmente, una fuente de luz utilizada en microscopía, debe tener lentes colectores y reflectores que permitan darle dirección para enfocar la luz en una muestra (Figura 2.2, donde se destacan partes como el filtro de calor (filtro infrarrojo), sistema de lentes colectores y sistemas de enfriamiento por convección) [7].

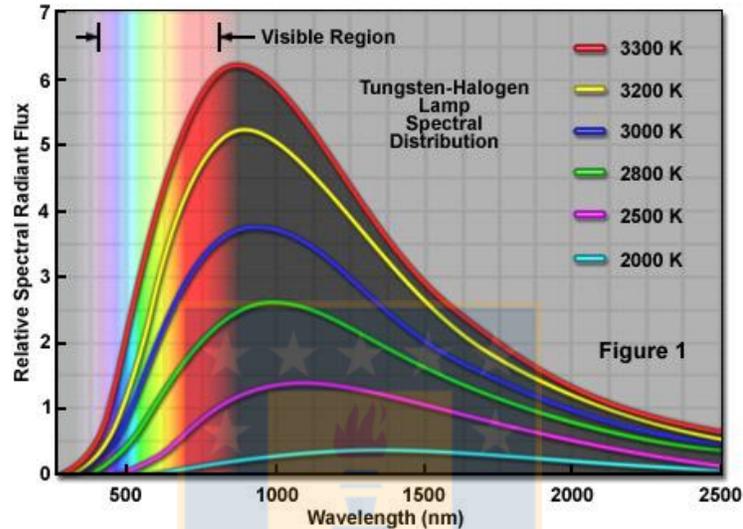


Figura 2.1: espectro emitido por una lámpara de tungsteno-halógena para diferentes temperaturas [7].

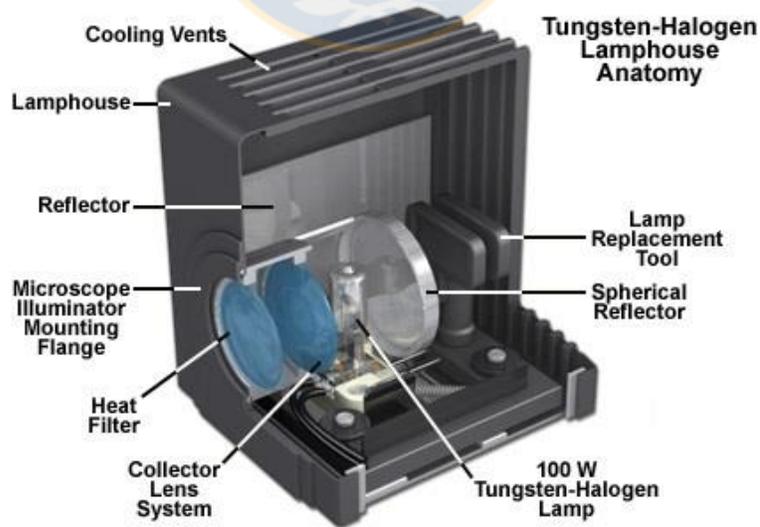


Figura 2.2: esquema de una lámpara de tungsteno-halógena para microscopía [7].

2.1.2. Microscopio de reflectancia

Para el análisis de muestras petrográficas¹ se puede utilizar un microscopio de reflectancia, que diferencia de un microscopio de transmisión (comúnmente utilizados en biología para la visualización de células), pues trabaja con muestras opacas y solo visualiza la superficie [8]. Algunos de estos microscopios se muestran en la Figura 2.3: (a) Zeiss Axioplan, (b) Leitz Ottoplan-Pol, (c) Nikon Optiphot-Pol.

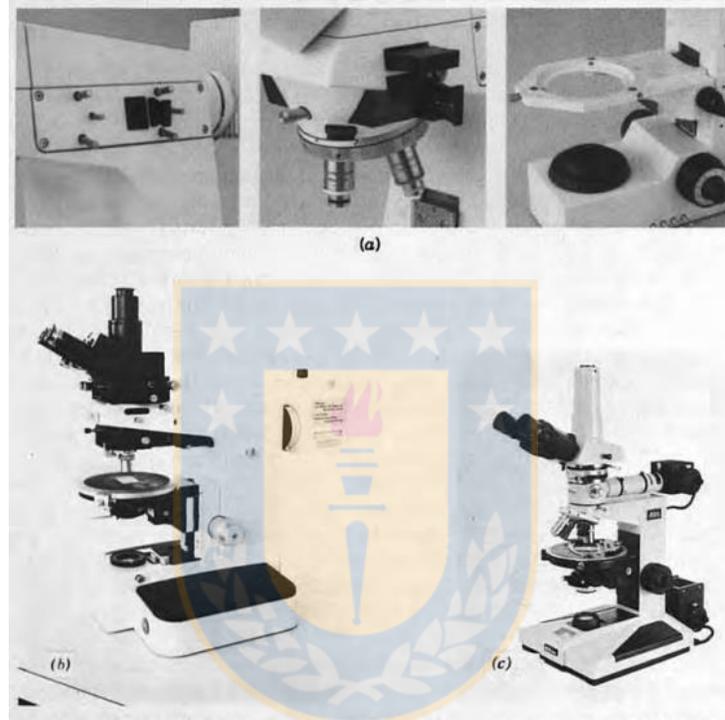


Figura 2.3: representación de microscopios utilizados en microscopía de reflectancia de minerales [8].

En la Figura 2.4 se presenta un esquema que representa los rayos principales y sus partes: sección de lente ocular (*eyelends, fixed diaphragm, field lens*), analizador (*analyzer*), prisma (*half field prism*), filamento de la fuente de luz (*filament*), lente colector (*collector lens*), diafragmas (*illuminator field diaphragm, illuminator aperture diaphragm*) y lente iluminador (*illuminator lens*).

¹ Petrografía: área que estudia la composición y propiedades de rocas y minerales.

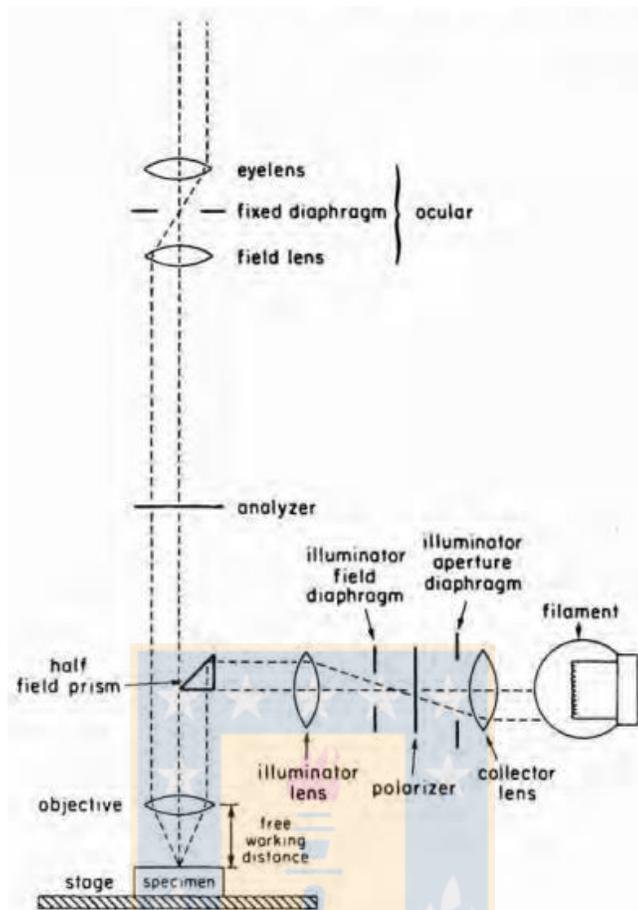


Figura 2.4: esquema simplificado de una sección transversal de un microscopio de reflectancia, los rayos de luz más importantes y sus partes [8].

2.1.2.1. Formación de imágenes en campo brillante (Bright Field)

Analizando la Figura 2.5, si se consideran los rayos de luz divergentes de un punto de la superficie de la fuente luminosa de la lámpara (L), se observa que el colector (Col) los reorganiza y los intersecta a una determinada distancia (en L'), formando una imagen de la superficie luminosa. En este plano se ubica un condensador apertura-diafragma (CAD), en el cual los rayos que pasan por el centro de este condensador se llaman rayos principales. Este condensador hace que los rayos sean paralelos entre sí, de esta forma, llegan a la muestra en forma cilíndrica (rayos paralelos que proyectan una circunferencia) lo que permite delimitar el área iluminada del mismo [9]. Notar que los rayos de luz impactan perpendicularmente a la muestra.

De otra forma, si se consideran rayos paralelos provenientes de la superficie luminosa, estos serán procesados por el colector hacia un plano focal, donde se ubica el diafragma

de campo luminoso (LFD), luego los condensadores (Con) enfocan estos rayos en la superficie de la muestra [9].

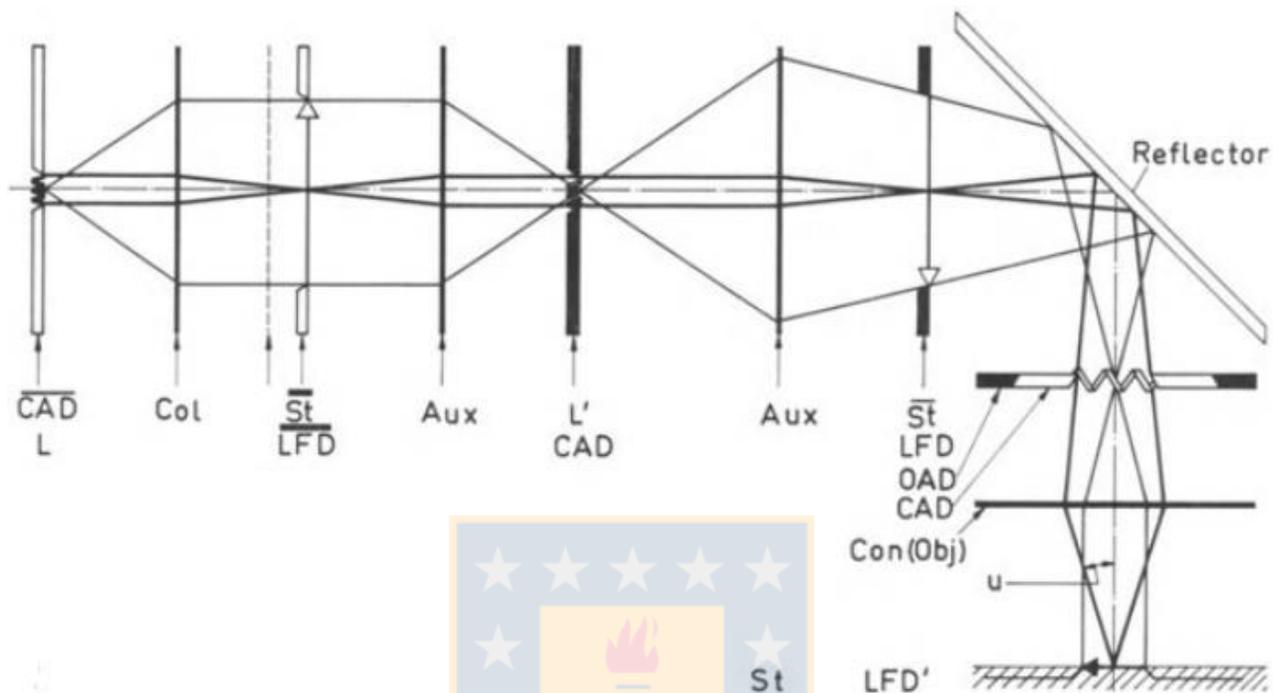


Figura 2.5: esquema de una imagen formada por microscopía en campo claro [9].

2.1.2.2. Formación de imágenes en campo oscuro (*Dark Field*)

En microscopía de campo oscuro, ningún rayo especular forma una imagen, solo la luz dispersada por la muestra es visible. De esta forma, las características visibles corresponden solo a inhomogeneidades en la muestra. En la iluminación de campo oscuro, no es posible controlar directamente la incidencia de luz, por lo que es posible que el objeto no esté iluminado homogéneamente, incluso usando arreglos simétricos. La iluminación DF en reflectancia se obtiene mediante un espejo en forma de anillo elíptico junto con un condensador que permite que la luz incida desde el ángulo requerido (Figura 2.6, notar que los rayos de luz inciden con un ángulo respecto a la normal de las muestras) [9].

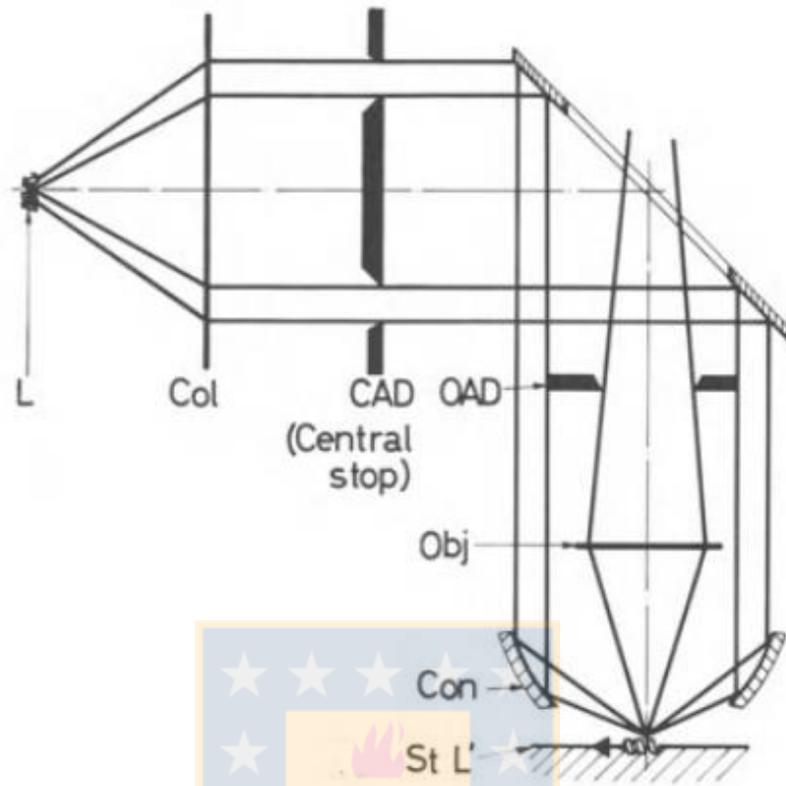


Figura 2.6: esquema de imagen formada por iluminación difusa (DF) [9].

2.1.3. Filtros y monocromadores

Las propiedades ópticas de varios materiales y minerales son directamente dependientes de la longitud de onda de la luz incidente, por lo que es de interés, estudiar dichas propiedades para una longitud de onda en específico. Para esto se puede utilizar un monocromador luego de la fuente de luz o un filtro, el cual puede estar luego de la fuente de luz o antes de un analizador. El intervalo de monocromadores y filtros disponibles comercialmente varía entre los 250 nm (luz ultravioleta) y 1100 nm (luz infrarroja) [9].

2.1.3.1. Monocromador

Un monocromador es un dispositivo de dispersión óptica utilizado para elegir una banda delgada de luz a partir de un mayor intervalo de longitudes de onda disponibles de una fuente de luz. La luz blanca es proveída desde la fuente de luz, entrando a través de la apertura, donde un lente colimador refleja la luz hacia una red de difracción, la que es reflejada hacia un lente de enfoque, el cual proyecta la luz hacia la salida. El ángulo de

la red de difracción determina la longitud de onda de la luz monocromática [10], como se muestra en la Figura 2.7.

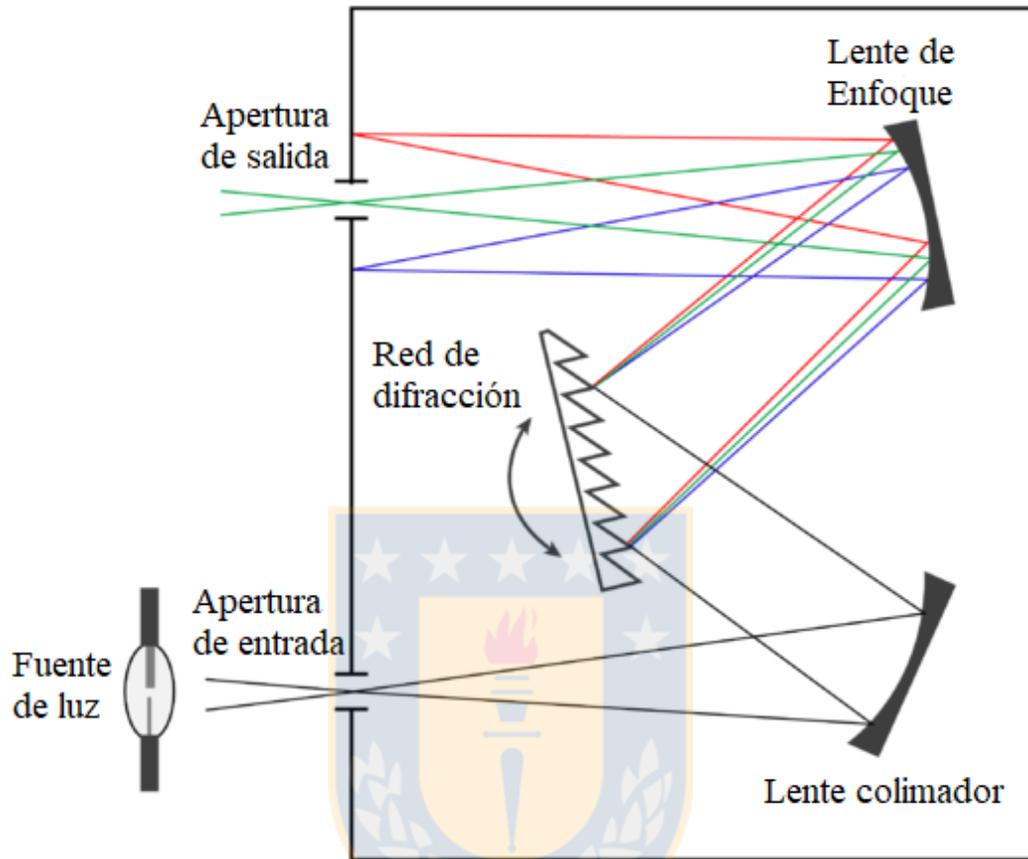


Figura 2.7: esquema de funcionamiento de un monocromador [11].

Los monocromadores funcionan con base en la dispersión de luz usando una red de difracción o prisma, donde una longitud de onda en particular atraviesa el equipo mientras todo el resto es filtrado. Debido a esto, es parte del funcionamiento de los equipos que exista una banda pasante de luz, *Pass band profile*, según su nombre en inglés, el cual comúnmente es definido por el usuario (Figura 2.8: Izquierda: Efecto del tamaño de apertura de entrada y salida en la luz producida por un monocromador asumiendo luz blanca en entrada (a) Tamaño de entrada ajustado con el tamaño de salida, (b) Entrada pequeña con respecto al tamaño de salida. Derecha: Distintos perfiles de ajuste para el ancho de banda en un monocromador, puede ser de tipo triangular (a), de banda estrecha en un filtro de interferencia (b), o de banda ancha (c)) [9].

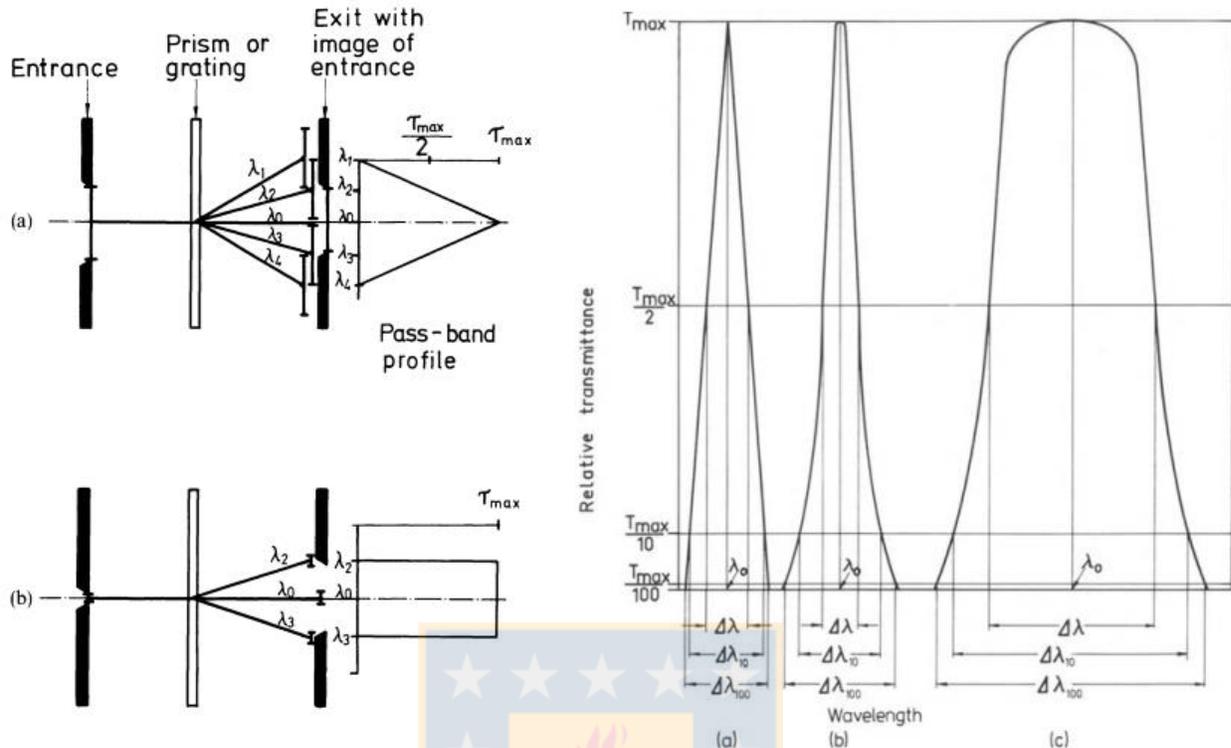


Figura 2.8: efecto del tamaño de apertura de entrada y salida de la luz producida por un monocromador y distintos perfiles producidos [9].

2.1.3.2. Filtro de luz variable

Los dos tipos de filtro variable más conocidos son el filtro variable de cristal líquido (LCTF, según su sigla en inglés) y el filtro acústico-óptico variable (AOTF).

Los LCTF están basados en el filtro Lyot (inventado por Bernard Lyot), el cual aprovecha la anisotropía de la birrefringencia para producir una banda delgada de luz pasante, con la diferencia de que el filtro Lyot utiliza una serie de cristales de cuarzo, mientras que un LCTF utiliza cristal líquido para producir el mismo efecto [12,13]. Ejemplos de este tipo de filtro corresponden a VariSpec (Perkin Elmer, discontinuado) y Kortum (Thor Labs).

Un AOTF, se basa en el efecto elasto óptico, y genera una red de difracción variable a partir de ondas sónicas sobre un cristal, las que van desde 428 hasta 990 MHz [14,15].

Esquemas de funcionamiento de ambos filtros se visualizan en la Figura 2.9.

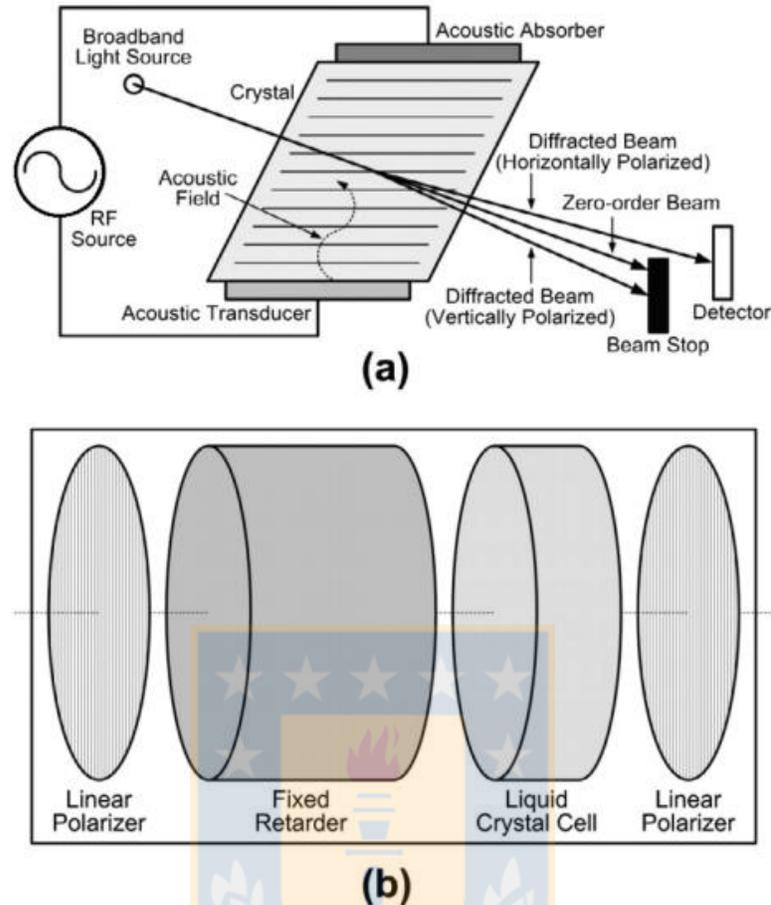


Figura 2.9: esquema de funcionamiento de (a) filtro acusto óptico, y (b) filtro de cristal líquido [16].

2.1.4. Espectrómetro y curvas de reflectancia espectral

Un espectrómetro óptico es un instrumento diseñado para medir propiedades de la luz dentro de una determinada porción del espectro electromagnético, típicamente utilizado para identificar materiales y composiciones [17], y para efectos del presente informe se utilizará para medir reflectancias de luz dentro del espectro VIS-NIR (400 nm hasta 900 nm).

En el contexto de reflectancia, un espectrómetro mide conteos de luz (en unidades arbitrarias) para cada longitud de onda que permite según sus especificaciones, por lo que para obtener una reflectancia se debe aplicar la ecuación (2.1) donde I es el conteo de luz en unidades arbitrarias medidas en la muestra o en una referencia [8].

$$R(\%) = \frac{I_{medida}}{I_{referencia}} \times 100\% \quad (2.1)$$

Una muestra (o blanco) de referencia es un material el cual teóricamente refleja la totalidad de la luz incidente dentro de un intervalo de longitudes de onda. Los blancos de referencia, pueden ser propietarios [18] o de materiales conocidos, como dióxido de titanio (TiO_2) u óxido de aluminio (Al_2O_3), pero debe tomarse en cuenta que distintas características, como tamaño de partícula o rugosidad de la superficie, pueden darle mejor utilidad a reflectancia difusa o especular [8]. Una curva de reflectancia espectral exhibe la reflectancia a todas las longitudes de onda medidas con respecto a la luz proveniente de la fuente (ver Figura 2.10: la pirita presenta colores agrisados debido a que la diferencia de reflectancia entre 400 nm y 700 nm no es tan pronunciada, por otro lado, la calcopirita posee colores más amarillos debido a una mucho mayor reflectancia en longitudes de onda a partir de 600 nm en comparación a la reflectancia a 400 nm), donde esta propiedad depende fuertemente de las características superficiales de la muestra tales como orientaciones mineralógicas, existencia de birrefringencia, pleocroísmo, etc. [19].

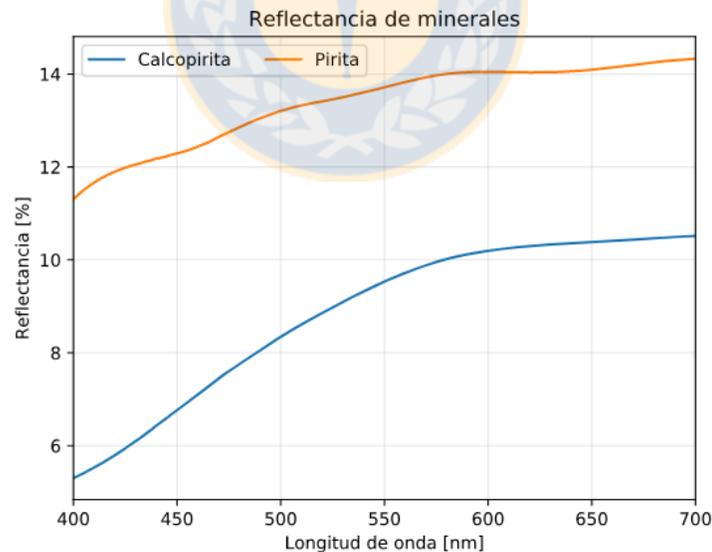


Figura 2.10: reflectancia espectral difusa de calcopirita y pirita en el espectro de luz visible. (Figura de elaboración propia).

2.2. Cota superior asintótica y orden de complejidad

La cota superior asintótica, también conocida como notación O grande, es una notación matemática que describe el comportamiento asintótico de una función cuando el argumento tiende tanto a un valor particular como al infinito. En ciencias computacionales la notación se utiliza para calificar y comparar la complejidad y escalamiento de algoritmos con respecto al tamaño de los valores de entrada [20]. Como tamaño se considera el sentido de memoria computacional, para un algoritmo de clasificación, esto se traduce al número de datos de entrada.

Se dice que un algoritmo es de orden $O(g(n))$ cuando la complejidad computacional escala según una función $g(n)$, donde n es el tamaño del valor de entrada.

Órdenes de complejidad comunes incluyen orden constante $O(1)$ (no depende del número de valores de entrada), orden logarítmico $O(\log(n))$, orden lineal logarítmico $O(n\log(n))$ y orden potencial de grado c $O(n^c)$.

2.3. Análisis de componentes principales (PCA)

El objetivo de realizar un PCA es crear una reducción de dimensiones, aproximando una matriz de datos X en el producto de dos matrices, una de *scores* T y una de *loadings* P , según la ecuación (2.2) si se utiliza la descomposición a través del método NIPALS (*Nonlinear Iterative Partial Least Squares*) [21].

$$X = T \cdot P \quad (2.2)$$

Otro método para la descomposición es usar SVD [22] según la ecuación (2.3):

$$X = U \Gamma V^T \quad (2.3)$$

Para efectos de este informe se utilizó descomposición mediante SVD.

2.3.1. PCA usando SVD

Se busca descomponer la matriz para reducir su orden, pues cualquier matriz evaluada con valores reales puede descomponerse según la ecuación (2.3), si X es una matriz de dimensiones $(n \times m)$, U es una matriz ortonormal de dimensiones $(n \times m)$, V una matriz cuadrada ortonormal de dimensiones $(m \times m)$ y Γ una matriz de dimensiones

$(m \times m)$ con valores singulares en su diagonal y valores nulos en el resto de sus posiciones. La matriz de covarianzas, de dimensiones $(m \times m)$, queda definida según la ecuación (2.4):

$$C = \frac{1}{n} XX^T = \frac{1}{n} U \Gamma^2 U^T \quad (2.4)$$

Si se define una matriz W , con dimensiones $(m \times m)$, de vectores propios de la matriz de covarianzas C y si se quieren obtener los k primeros componentes principales, usando la ecuación (2.5), se obtiene una matriz X_k con los scores de los elementos principales de la matriz original de datos X . Se observa que el número de componentes principales no puede superar el número de variables que inicialmente tienen los datos.

$$X_k = XW_k \quad (2.5)$$

La varianza explicada por cada componente principal está representada en cada valor propio de la matriz de covarianzas C [23]. Un símil de los *loadings* se obtiene a través de la varianza de todas las mediciones en cada variable.

2.3.2. PCA como técnica de exploración

Usar PCA en exploración de datos, para utilidad de esta investigación, tiene dos objetivos:

1. Reducir las variables a componentes principales que tengan la mayor varianza explicada, de esta forma, se pueden graficar los *scores* para determinar si se forman clústers de datos relacionados.
2. Estudiar los *loadings* para determinar que variables son más relevantes para la detección de puntos de interés entre las mediciones.

La existencia de clústers de datos similares, señala que existe una relación entre los datos de cada clúster (o sea, podrían ser clasificados independientemente). La existencia de *peaks* en un gráfico de variables/*loadings* indica variables de interés para el estudio de los datos, como saber cuáles producen un mayor efecto [21].

2.4. Aprendizaje supervisado y modelos de clasificación

El aprendizaje computacional (ML, *machine learning* en inglés) se define como un conjunto de métodos computacionales para realizar predicciones basadas en *experiencia* (datos previamente conocidos). ML consiste en diseñar y aplicar algoritmos predictivos precisos, basados en información obtenida previamente. Aplicaciones de ML incluyen detección de *spam* en mails, reconocimiento y síntesis de voz y detección de fraudes bancarios, entre otros.

El aprendizaje supervisado es un tipo de aprendizaje utilizado en regresión y clasificación donde el modelo recibe un *set* de ejemplos como datos de aprendizajes para ajustar el modelo (*fit*), lo que posteriormente permite hacer predicciones.

Un clasificador se define como un modelo que asigna categorías a un valor de entrada en la cual su categoría es desconocida [24].

En el presente informe se utilizaron cuatro clasificadores: árbol de decisiones, KNN, SVM y SGD.

2.4.1. Conceptos relevantes

- *Fitting*: proceso en el que se ajusta el modelo dado el *set* de aprendizaje.
- *Overfitting*: cuando el modelo se sobre ajusta a los datos del *set* de aprendizaje y no generaliza de forma correcta.
- Función de decisión: función que determina numéricamente el resultado de un algoritmo clasificador.
- *Cross validation*: prueba para validar la precisión de un modelo, usando una fracción de los datos obtenidos que no son usados para ajustar el modelo. Se puede utilizar para detectar *overfitting*.
- Función de pérdida (*loss function*): función que determina la diferencia (o pérdida de precisión) entre la clase predicha y la real.
- Función signo: determina el signo de la función, formalmente está definida por

$$\text{sgn}(x) = \begin{cases} -1 & \text{si } x < 0 \\ 1 & \text{si } x > 0 \end{cases}$$

- *Outlier*: punto que difiere significativamente del resto de las mediciones.

- Clase: clasificación de la muestra.

2.4.2. Árbol de decisiones (DT)

Un árbol de decisiones, [25] es un método de aprendizaje supervisado utilizable en clasificación y regresión. En el caso de clasificación, el objetivo es crear un modelo a través del aprendizaje de reglas simples en el *set* de entrenamiento.

Ventajas incluyen:

- Fáciles de entender e interpretar; los modelos pueden ser visualizados.
- La cantidad de mediciones para crear un modelo puede ser baja (menor a 50 mediciones), sin embargo, también es aplicable a grandes cantidades (mayores a cien mil).
- Puede utilizar variables tanto numéricas como categóricas.
- Complejidad computacional baja, de tipo lineal logarítmico $O(n \log(n))$ para creación del modelo y su aplicación es trivial.

Desventajas incluyen:

- Se pueden crear modelos complejos que no generalizan bien los datos (*overfitting*). Esto se puede evitar, limitando el número de resultados que el modelo entrega.
- El modelo puede ser sesgado, si el número de mediciones no está equilibrado (existe una gran diferencia en el número de distintos tipos de mediciones).

2.4.2.1. Formulación matemática

Dado un *set* de vectores de entrenamiento $x_i \in \mathbb{R}^n$ y un vector de resultados $y \in \mathbb{R}^l$, un árbol de decisiones crea particiones de tal forma que los ejemplos con el mismo resultado queden agrupados. Sean los datos de un nodo m representados por Q . Para cada candidato a separación $\theta = (j, t_m)$, donde j es una característica y t_m un límite, se crea una partición para los conjuntos $Q_1(\theta)$ y $Q_2(\theta)$ según las ecuaciones (2.6) y (2.7).

$$Q_1(\theta) = (x, y) \mid x_j \leq t_m \quad (2.6)$$

$$Q_2(\theta) = Q - Q_1(\theta) \quad (2.7)$$

La impureza total $G(Q, \theta)$, dada una función de impureza $H(X)$ en el nodo m , es la suma de las impurezas normalizadas de cada partición según la ecuación (2.8). Por lo que se debe elegir un valor θ tal que minimice dicha ecuación.

$$G(Q, \theta) = \frac{n_1}{N_m} H(Q_1(\theta)) + \frac{n_2}{N_m} H(Q_2(\theta)) \quad (2.8)$$

Este proceso se repite hasta alcanzar la cantidad de nodos deseada. La función de impureza utilizada en esta investigación fue entropía en información, definido en la ecuación (2.9), donde el operador $P(x)$ corresponde a la fracción de x con respecto al total de los elementos.

$$H(X) = -\sum_{i=1}^n P(x_i) \log P(x_i) \quad (2.9)$$

2.4.3. KNN

El algoritmo KNN es uno de los métodos más simples para clasificación y no requiere ajuste de un modelo. Compara la posición de un elemento de clase desconocida con relación a los de clase conocida. Se asume que si un elemento es "cercano" a una clase, entonces corresponde a esta [26].

Ventajas del método incluyen [27]:

- Simple e intuitivo; fácil de entender.
- No necesita supuestos en la información de entrada.
- Puede agregarse nueva información constantemente, debido a que no utiliza un modelo ajustado.
- Fácil de ajustar y sólo requiere dos parámetros: el número k de vecinos a considerar y el tipo de distancia a utilizar.

Desventajas del método son:

- Es un algoritmo que se hace cada vez más lento a medida que aumenta el número de datos. Orden $O(D \cdot n^2)$, para D dimensiones, esto se aplica cada vez que se clasifica un valor.

- Se puede volver impreciso si se agregan muchas variables a los datos.
- No existe un método eficiente para determinar el número de vecinos k a considerar.
- Sensible a *outliers*, especialmente cuando k es pequeño.

2.4.3.1. Metodología y aplicación

Se define la distancia euclidiana según la ecuación (2.10) y la distancia Manhattan según la ecuación (2.11).

$$D_E(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.10)$$

$$D_M(X, Y) = \sum_{i=1}^n |x_i - y_i| \quad (2.11)$$

Dada una lista de datos $(x_1, y_1), \dots, (x_n, y_n)$, con $x_i \in \mathbb{R}^m$ y y_i la clasificación conocida del objeto (numérica o categórica), se elige un valor $k \in \mathbb{N}$ y un tipo de distancia² a considerar. Para determinar la clase de un objeto con valor x_0 , se calcula la distancia a todos los elementos de la lista de datos y se eligen los k más cercanos, el que se repita más veces se considerará la clase del objeto x_0 . Opcionalmente, en vez de simplemente sumar cada objeto cercano se puede considerar el inverso de la distancia como factor de peso. Valores grandes de k pueden reducir el efecto de *outliers*, pero hacer menos precisa la clasificación en regiones intermedias de dos clases. [26]

2.4.4. SVM

El algoritmo SVM (también llamado SVC, *support vector classifier*, cuando se usa para clasificación), puede ser utilizado en regresiones y clasificaciones. Se basa en la determinación de hiperplanos que separen zonas de interés de tal forma que el margen de separación sea el mayor posible (Figura 2.11) [24].

² La distancia euclidiana tiene mejores resultados, pero si el número de datos es alto puede ser conveniente usar distancia Manhattan para reducir el costo computacional.

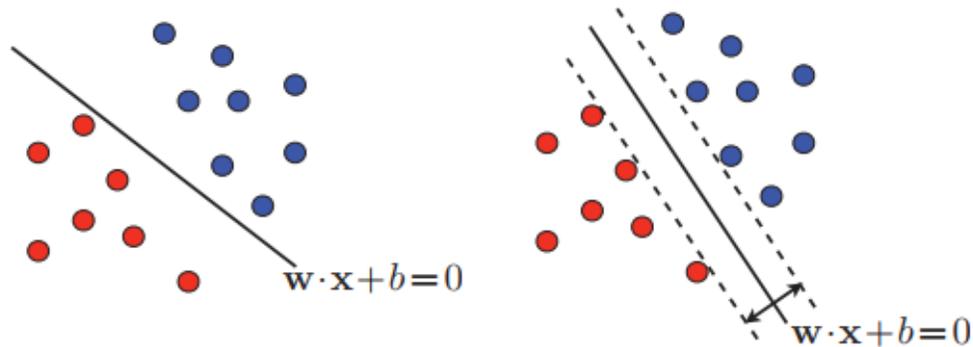


Figura 2.11: dos posibles hiperplanos de separación, el del lado derecho es el que optimiza los márgenes. [24]

Ventajas [28] de usar SVM incluyen:

- Efectivo en casos de altas dimensiones, incluso cuando hay más dimensiones que número de muestras.
- Eficiente con el uso de memoria y recursos durante la clasificación.
- Versátil, puede usar kernels para mejorar la clasificación de datos (no utilizado en este informe).

Desventajas del algoritmo:

- Si el número de variables es mucho mayor al número de muestras es muy probable producir *overfitting*.
- El ajuste del modelo no escala bien a situaciones con una elevada cantidad de datos debido a su complejidad de orden cúbico $O(f \cdot n^3)$, para f dimensiones, en el ajuste del modelo.

2.4.4.1. Formulación matemática

Para el clasificador, dado un *set* de vectores de entrenamiento $x_i \in \mathbb{R}^p$, con $i=1, \dots, n$, y un vector $y \in \{1, -1\}^n$, SVC resuelve el problema de la ecuación (2.12), sujeto a $y^T \alpha = 0$ y $0 \leq \alpha_i \leq C, i=1, \dots, n$.

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \quad (2.12)$$

Donde $C > 0$ es el límite superior, Q una matriz $(n \times n)$ tal que $Q_{ij} \equiv y_i y_j K(x_i, x_j)$, para una función kernel $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$. La función de decisión queda definida según la ecuación (2.13) [28].

$$\text{sgn} \left(\sum_{i=1}^n y_i \alpha_i K(x_i, x) \right) \quad (2.13)$$

2.4.5. SGD

SGD es un clasificador lineal similar a SVM (sin aplicar kernel) que funciona en lotes de información. Cada lote se entrega con datos aleatorios para ir continuamente ajustando el modelo, esto aumenta la velocidad de convergencia y mejora el uso de memoria [29].

Ventajas de SGD incluyen:

- Eficiencia, tanto en aplicación del modelo (correspondiente a funciones lineales) como en el uso de memoria.
- Baja complejidad, de tipo lineal $O(n)$ para el ajuste del modelo.
- Cálculos matemáticos simples, lo que permiten ajustar el código a cada situación.
- Puede ajustarse con un nuevo ingreso de datos sin la necesidad de realizar todos los cálculos nuevamente.

Desventajas incluyen:

- Requiere el ajuste de hiperparámetros y el número de lotes que se va a entregar.
- Sensible al número de mediciones de cada clase, diferencias grandes en el número de mediciones de clases distintas puede generar un sesgo.
- Para converger a una solución se requieren más de 10^5 mediciones previas [30].

2.4.5.1. Formulación matemática

Dado un set de entrenamiento $(x_1, y_1), \dots, (x_n, y_n)$ donde $x_i \in \mathbb{R}^m$, $y_i \in \{-1, 1\}$, el objetivo es encontrar una función lineal $f(x) = w^T x + b$ con parámetros $w \in \mathbb{R}^m$ y $b \in \mathbb{R}$. Donde para hacer predicciones se aplica $\text{sgn}(f(x))$. La forma de determinar parámetros consiste en minimizar la función definida según la ecuación (2.14), donde L es una función de

pérdida, R un término de regularización (penalización de error) y $\alpha > 0$ un hiperparámetro positivo³. [30]

$$E(w, b) = \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i)) + \alpha R(w) \quad (2.14)$$

Funciones R comúnmente usadas son la norma L2 (2.15), L1 (2.16).

$$R(w) = \frac{1}{2} \sum_{i=1}^n w_i^2 \quad (2.15)$$

$$R(w) = \sum_{i=1}^n |w_i| \quad (2.16)$$

2.4.6. Clasificación con múltiples clases

Los métodos SGD y SVM realizan una clasificación binaria, por lo que para extenderlo a más de dos clases se utilizan métodos OvO (*one-versus-one*) y OvR (*one-versus-rest*). Estos métodos transforman un problema multiclase a un set de problemas binarios [31].

2.4.6.1. OvR

En el método *one-vs-rest*, uno contra el resto, también llamado OvA (*one-vs-all*, uno contra los demás), si se tienen K clases se requieren $N = K$ clasificadores binarios, donde cada problema discrimina contra el resto de las $K - 1$ clases restantes. El k -ésimo clasificador se entrena con resultados positivos de dicha clase y resultados negativos de las $K - 1$ clases restantes [32].

2.4.6.2. OvO

En el caso de usar *one-vs-one* (uno contra uno, también llamado AvA, *all-vs-all*) se entrenan modelos para clasificar entre todas las combinaciones de clases posibles, por lo que para K clases se requerirá entrenar $N = K(K - 1) / 2$ clasificadores [32], donde

³ Tanto R como α forman parte de la "penalización" que se le da a la función, a mayor penalización aumenta la posibilidad de *overfitting*, pero penalizaciones muy bajas no entregarán modelos que generalicen correctamente.

después a través de un sistema de votación se elige una clase [31]. Este método entrega mejores resultados que usar OvR [33] a cambio de un mayor costo computacional.

2.5. Investigaciones previas

2.5.1. Estudio de automatización de análisis de sulfuros

Chopard [34] clasificó cuarzo y minerales sulfurados (pirita, calcopirita, esfalerita, arsenopirita, pirrotina, esfalerita y galena) utilizando las bandas 365, 436, 480, 605 y 650 nm con un ancho de banda de 10 nm. Las muestras estaban en forma de polvo con granulometría variando entre 50 y 150 μm con una superficie pulida. Luego de obtener imágenes, con una magnificación de 200X sin usar polarizadores, a las longitudes de onda mencionadas, el estudio procedió a aplicar PCA para comprobar los clústers de datos. La clasificación fue realizada a través de *K-means clustering*, lo que concluye en la producción automatizada de una imagen (Figura 2.12) donde la composición se obtiene mediante conteo de píxeles. En este estudio se destacan bajos errores relativos en relación al análisis mediante microscopía electrónica de barrido, menores al 9% para todos los minerales, con excepción de la arsenopirita (19,6% de error relativo) y la pirita (37,9% de error relativo).

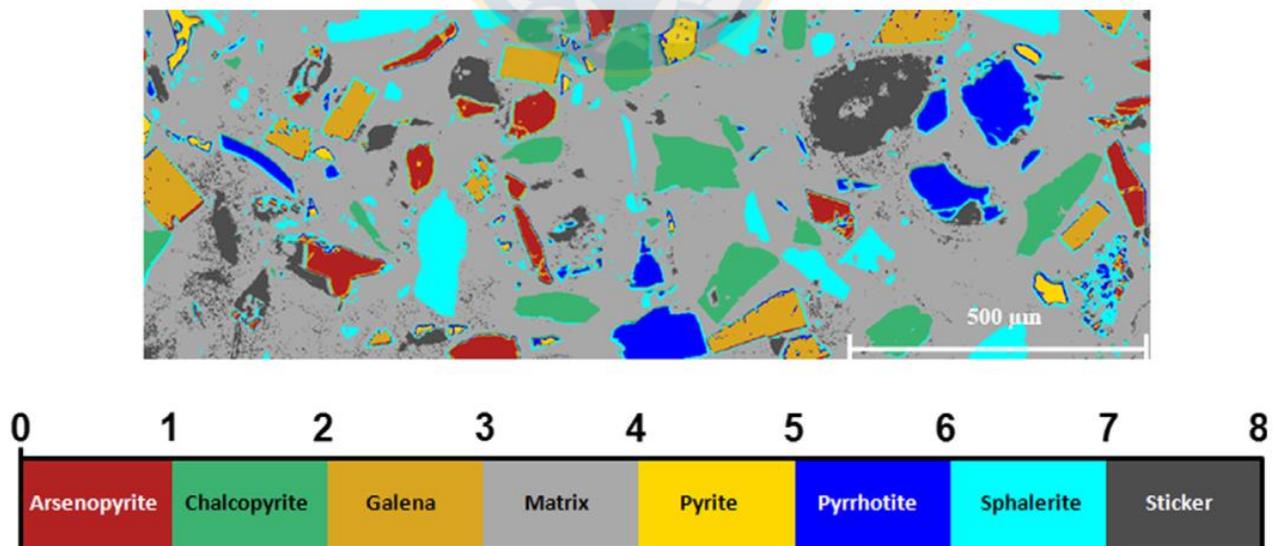


Figura 2.12: resultados en clasificación para estudio de Chopard [34].

2.5.2. Estudio de reconocimiento automatizado de menas metálicas

Castroviejo [35] y Catalina [36] investigaron la automatización de reconocimiento de menas metálicas mediante imágenes multispectrales, lo que terminó en un diseño de una patente a nombre de la Universidad Politécnica de Madrid [37], donde según explican, diseñaron una base de datos que permite discriminar entre "70 minerales de interés" con superficies pulidas e iluminación BF, utilizando longitudes de onda 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950 y 1000 nm con un ancho de banda de 40 nm. Las imágenes fueron tomadas con una lámpara halógena y lentes objetivo de 2,5X, 5X, 10X y 20X. Debido al extenso intervalo de longitudes de onda fue necesario considerar una variación en el plano de enfoque dependiente de cada longitud de onda (Figura 2.13).

La clasificación fue a través del método KNN y obtuvo errores relativos menores al 5% para todos los minerales.

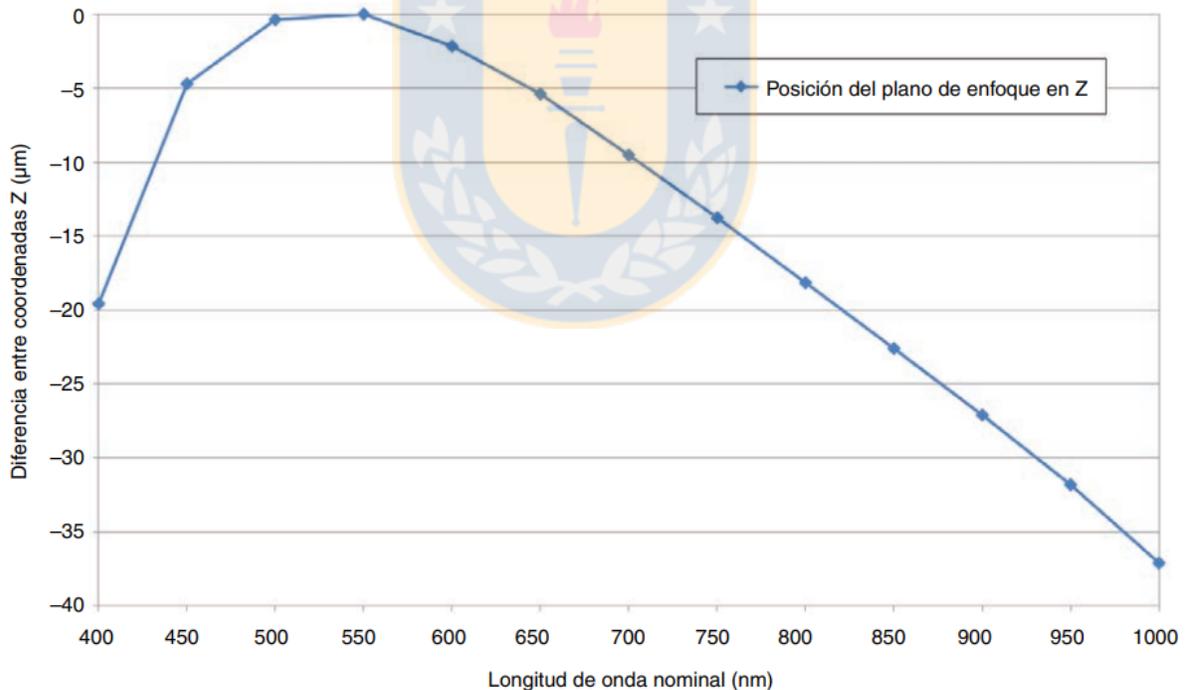


Figura 2.13: curva que recoge la variación del plano de enfoque óptimo [36].

3 Metodología

En la metodología se diferenciaron dos partes: una primera etapa con el uso del espectrómetro para obtención de información y otra utilizando de una cámara monocromática.

El objetivo de la primera etapa fue obtener curvas de reflectancia de minerales para luego, a través de herramientas matemáticas y estadísticas, determinar variables que permitan la clasificación de cada muestra de mineral.

En la segunda etapa, se obtuvieron fotografías de pellets de minerales cuyas composiciones se asumieron puras, para entrenar modelos de ML, que posteriormente se probaron con pellets de concentrados de cobre con composición conocida para determinar la precisión del modelo.

3.1. Equipamiento y muestras

3.1.1. Primera etapa

Con respecto a equipos y materiales, se utilizó:

- Microscopio Nikon Labophot-Pol adaptado para reflectancia (BF y DF).
- Fuente de luz halógena-tungsteno.
- Espectrómetro Ocean Optics HR2000+ES.
- Línea de fibra óptica (con lente opcional).
- Lente objetivo 10X.
- Set A de pellets de minerales: pirita, calcopirita, covelina y calcosina.
- Set B de pellets de minerales: pirita, calcopirita, bornita y enargita.
- Óxido de titanio en polvo (blanco de referencia).

Software utilizado:

- OceanView.
- Python 3.7 con librerías scikit-learn.

3.1.2. Segunda etapa

- Microscopio Nikon Labophot-Pol adaptado para reflectancia (BF y DF).
- Fuente de luz halógena-tungsteno.
- Lente objetivo Nikon 10X.
- Cámara astronómica monocromática QHY-9.
- Monocromador de operación manual Oriel 7240.
- Filtro variable VariSpec (tipo LCTF, intervalo 400-720 nm).
- Set A de pellets minerales: pirita, calcopirita, covelina y calcosina.
- Pellets de muestras de concentrado de cobre.
- Óxido de titanio en polvo (blanco de referencia).

Software:

- VsGui.
- AMIS (Desarrollado por CEFOP).
- Python 3.7 con librerías scikit-learn y skimage.

3.2. Montaje experimental

En el montaje del microscopio varían dos elementos: la fuente de luz utilizada y la herramienta para analizar la luz reflejada. Como fuente de luz se utilizó una lámpara halógena-tungsteno en el cual un monocromador es opcional (ubicado entre la lámpara y la entrada de luz del microscopio). Una fotografía del montaje experimental se muestra en la Figura 3.1.

Durante la primera etapa, en la zona del analizador (cuadrado rojo en Figura 3.1), se utilizó un cable de fibra óptica que dirigió la luz reflejada hacia el espectrómetro HR2000+ES (Figura 3.2, izquierda) con el objetivo de obtener curvas de reflectancia, utilizando una fuente de luz halógena-tungsteno (mostrado en el cuadrado celeste de la Figura 3.1).

En la segunda etapa para obtener imágenes en campo claro (BF), se utilizó el monocromador Oriel 7240 (Figura 3.2, derecha) con un ancho de banda de 10 nm entre la fuente de luz y el microscopio; en la zona del analizador se acopló la cámara monocromática QHY-9. Para obtener imágenes de campo oscuro (DF) no se utilizó monocromador y se acopló el filtro VariSpec previo a la cámara QHY-9.

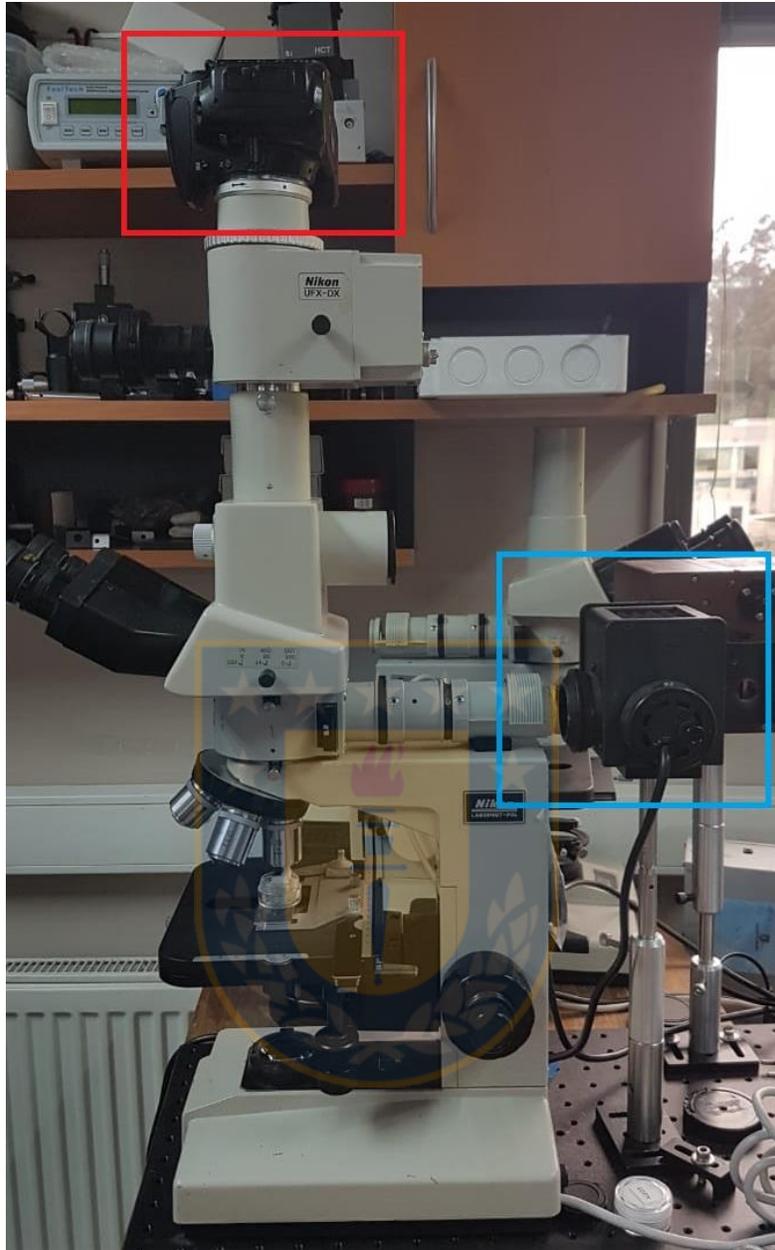


Figura 3.1: montaje experimental; en la parte superior se encuentra una cámara (cuadrado rojo) utilizada para recibir la luz reflejada, al lado derecho (cuadrado celeste) se encuentra una fuente de luz sin monocromador. Las zonas marcadas corresponden a las herramientas que varían dependiendo del tipo de información que pretende obtenerse.



Figura 3.2: izquierda: espectrómetro Ocean Optics HR2000+ES, derecha: monocromador Oriel 7240



3.3. Detalle de muestras

Las muestras de minerales, en formas de pellets, fueron analizadas a través de un análisis QEMSCAN en el instituto GEA de la Universidad de Concepción (Tabla 2). Las muestras de concentrado fueron suministradas por la Fundición Potrerillos y analizadas mediante FRX (Tabla 3).

Tabla 2: composición de las muestras.

	py	cpy	cov	ccs
Pirita	97,2	7,56	0,02	0,01
Calcopirita	0,00	81,4	1,17	0,52
Covelina	0,00	0,00	92,05	1,22
Cuarzo	0,43	0,56	0,07	2,16
Calcosina	0,06	0,00	2,68	86,9
Otros	2,34	10,5	4,01	9,17

Tabla 3: composición química de muestras de concentrado.

Muestra	Cu (%)	Fe (%)	S (%)	Si (%)	Suma (%)
M01	23,3	25,7	35,1	4,23	88,3
M02	26,0	23,8	28,9	5,35	84,1
M03	25,5	22,9	29,2	5,44	83,0
M04	23,2	24,1	28,5	5,57	81,4
M05	23,9	24,1	29,7	5,76	83,5
M06	22,9	23,0	29,0	5,55	80,4
M07	23,3	23,0	32,7	5,45	84,4
M08	23,6	23,7	30,8	5,50	83,6
M09	23,6	23,7	28,9	5,24	81,4
M10	23,9	23,5	30,0	5,52	82,9
M11	23,6	20,3	31,2	4,81	79,9

3.4. Desarrollo experimental

3.4.1. Primera etapa

En la primera etapa, se utilizó el espectrómetro y una fuente de luz halógena-tungsteno sin monocromador. Se utilizaron pellets de calcopirita, calcosina, piritita y covelina (Tabla 2), además de pellets de otro origen de calcopirita, piritita, covelina, bornita y enargita.

Una medición se define como el promedio de 100 curvas de reflectancia en un mismo punto, y en cada pellet se tomaron 125 mediciones en puntos distintos elegidos a discreción del experimentador con el cuidado de no repetir zonas del mineral. En cada medición, se utilizó una configuración de 180 μm de área de integración⁴ usando un lente, 40 μm de área de integración sin lente, 16 milisegundos de tiempo de exposición y un filtrado *boxcar smoothing* de tres puntos, utilizando el software OceanView. Las mediciones son comparadas contra una muestra de óxido de titanio para obtener una medición de reflectancia.

El software mencionado entrega archivos de texto tipo csv (*comma separated values*), los que pueden importarse y analizarse mediante el lenguaje de programación Python, donde se aplica análisis mediante componentes principales (PCA), utilizando una librería de elaboración propia, para determinar longitudes de onda relevantes para la clasificación de cada mineral. Posteriormente se probaron modelos de ML (explicados en el inciso 2.4 Aprendizaje supervisado y modelos de clasificación) utilizando librerías scikit-learn, con el objetivo de determinar la posibilidad de crear un modelo de clasificación y establecer longitudes de onda discriminantes para cada mineral.

En el modelo de árbol de decisiones (inciso 2.4.2) se limitó el número de hojas a seis utilizando entropía de información como métrica, correspondientes a los minerales disponibles; KNN fue utilizado con 40 como número de vecinos a considerar, SVM se utilizó con técnica OVO como clasificación multiclase. Se efectuaron pruebas de *cross validation* con un 50% de los datos.

⁴ También conocido como FoV (*field of view*).

3.4.2. Segunda etapa (BF)

Como analizador se utilizó la cámara monocromática QHY-9 y para la fuente de luz se recurrió a la lámpara halógena-tungsteno acoplada al monocromador Oriel 7240. Las muestras analizadas corresponden a las señaladas en la Tabla 2. Las longitudes de onda utilizadas corresponden a 400, 500, 600, 800 y 900 nm con un ancho de banda de 10 nm. La iluminación es de tipo especular (campo brillante, BF).

Para capturar las fotografías, se utilizó el óxido de titanio como blanco de referencia y se determinó un tiempo de exposición para cada longitud de onda de tal forma que se maximizara la luz de entrada sin saturar el sensor de la cámara⁵; luego la misma configuración de cada longitud de onda se utiliza para obtener imágenes de las muestras de mineral a dicha longitud de onda, todo esto utilizando el software AMIS.

La resolución de la fotografía es 1674 x 1248 píxeles, de tal forma que 5 píxeles equivalen a 16 μm , sin embargo, solo se consideraron las zonas completamente iluminadas de la muestra, zonas oscuras fueron descartadas del análisis. El formato de imagen entregado por AMIS, corresponde a un archivo TIFF (*Tagged Image File Format*) de 16 bits sin compresión con un solo canal de información, o sea, cada píxel solo tiene información de la intensidad de luz en un número desde 0 hasta 65535, donde 0 corresponde a que no existe luz y 65535 corresponde a luz completamente blanca, de esta forma, el valor del píxel es un símil a un conteo de luz en unidades arbitrarias.

Para realizar una conversión entre conteos de luz en unidades arbitrarias a reflectancia, en cada longitud de onda se considera el valor del píxel en el percentil 99 del blanco de referencia. Luego, en las fotografías de minerales se aplica la ecuación (2.1), donde la intensidad de referencia corresponde al valor del píxel del percentil 99, obteniendo un mapa de reflectancias para cada píxel de la fotografía.

Para analizar los datos de la imagen, estos se importan utilizando skimage en Python. Luego con estos valores se aplica PCA y se probaron los modelos de ML descritos en el inciso 2.4. También se estudió el efecto de la resolución (conversión píxeles/longitud) en la precisión del modelo, utilizando métricas de *cross-validation* para el 50% de los

⁵ Si existen píxeles con valores de 65535 se asume que el sensor fue saturado.

datos. La variación de resolución fue realizada con la librería skimage, sin aplicar suavizado en el re escalado de la imagen.

El objetivo de esta etapa fue obtener modelos de ML que puedan aplicarse para analizar concentrados de cobre en condiciones de iluminación especular (*bright field*), considerando el intervalo de longitudes de onda desde 400 nm hasta 900 nm.

3.4.3. Segunda etapa (DF)

Se utilizó la fuente de luz halógena-tungsteno sin monocromador y en el analizador se acopló el filtro líquido variable VariSpec junto con la cámara monocromática QHY-9. La iluminación fue de tipo difusa (campo oscuro, DF). Las muestras utilizadas corresponden a las señaladas en la Tabla 2.

El filtro VariSpec tiene un intervalo de operación dentro de la luz visible (desde 400 nm hasta 720 nm), por lo que se tomaron fotografías cada 10 nm con un ancho de banda de 10 nm. Esto es manejado por la aplicación propietaria vsui.

El *software* AMIS se utilizó de manera similar al caso de iluminación en campo brillante (BF), pero en este caso se utilizó una resolución de 3348 x 2496 píxeles, con una resolución métrica de 5 píxeles equivalentes a 32 micrones. Debido a las características del filtro y la iluminación, la totalidad de la fotografía es útil para su análisis y no fue necesario solo considerar ciertas regiones.

El análisis de datos es similar al señalado en el caso de iluminación BF y el objetivo fue obtener modelos de ML que permitan clasificar concentrados de cobre con iluminación difusa en el intervalo de luz visible.

3.4.4. Análisis de concentrados

Para analizar los concentrados (señalados en la Tabla 3), se obtuvieron fotografías de los concentrados mediante las técnicas (BF y DF) y posteriormente, fueron analizadas por los modelos obtenidos anteriormente en las condiciones que maximizaron las pruebas de precisión (*cross-validation*).

Debido a que el análisis de la Tabla 3 es elemental y no mineralógico se aplicó una transformación usando las fracciones en peso según especificaciones de la Tabla 4. Cuando un píxel de la fotografía fue clasificado como un mineral determinado, este

aportó con su porcentaje correspondiente al conteo elemental final. En la ecuación (3.1) se muestra el cálculo de la fracción elemental (C_x) a partir del número de píxeles detectados de un mineral n_i , su fracción en peso W_{xi} y el número total N de los píxeles de la imagen, esto se aplica para todos los elementos y minerales de interés.

Posteriormente, estos resultados fueron comparados con las composiciones elementales de cada concentrado utilizando un error relativo según la ecuación (3.2).

Tabla 4: fracciones elementales de cada mineral.

Fracción en peso (W_x) [%]	Pirita	Calcopirita	Calcosina	Covelina
Cobre	0	34.62	79.85	66.46
Hierro	46.55	30.43	0	0
Azufre	53.45	34.95	20.15	33.54

$$C_x [\%] = \frac{n_i \times W_{xi}}{N} \times 100\% \quad (3.1)$$

$$E[\%] = \frac{|X_{medido} - X_{referencia}|}{X_{referencia}} \times 100\% \quad (3.2)$$

4 RESULTADOS Y DISCUSIONES

4.1. Determinación de curvas de reflectancia espectral

En primer lugar, se identificó el efecto del tipo de iluminación en las curvas de reflectancia de los minerales. Cuando se utilizó iluminación especular (BF) se obtienen curvas con formas similares (como se puede observar en Figura 4.1) donde si bien existe una cierta tendencia en cada mineral, estas no poseen el mismo perfil que tienen las curvas de Criddle & Stanley [19], por lo que no es posible asegurar que se está midiendo la reflectancia del mineral en vez de otras características de la muestra, lo que sí ocurre a grandes rasgos en el caso de la iluminación difusa (DF) considerando un *offset* vertical, visible en la Figura 4.2.

Otro efecto visible es la diferencia entre las desviaciones estándar cuando se utiliza un área de integración de 40 μm (Figura 4.2) y 180 μm (Figura 4.3), por lo que se sugiere que existe una relación entre el área de integración y el tamaño de partícula que afecta la varianza de las mediciones. Para probar distintas áreas de integración deben utilizarse distintos lentes acoplados a la fibra óptica, de los cuales no se disponía.

En la Figura 4.3 se consideró el análisis de ambos sets de pellets de minerales (ver inciso 3.4.2), analizados en un espectro VISNIR con un área de integración de 180 μm . Es visible que existen regiones donde cada mineral puede identificarse independientemente de otros, además de que la calcopirita posee un gran nivel de varianza, esto debido a un posible efecto del tamaño de partícula sobre la curva de reflectancia formada (Figura 4.4).

De esto se puede determinar que:

1. Es de suma importancia el tipo de iluminación. En la luz reflejada mediante iluminación en campo oscuro, no se consideran caras de cristales del mineral, pero si se entrega información del resto de las superficies que no reflejan especularmente, los que en un sistema particulado corresponden a la mayoría de la superficie estudiada. En contraparte, en iluminación de campo brillante solo son visibles las caras de minerales que reflejan especularmente la luz (ver Figura 2.5), las zonas que no cumplen con esta característica no aportan información importante al análisis (ver Figura 7.2).

2. Existe la posibilidad de que la razón entre el tamaño de partícula y área de integración sean de relevancia en la curva de reflectancia espectral medida, por lo tanto, en una fotografía la resolución podría ser una variable relevante.

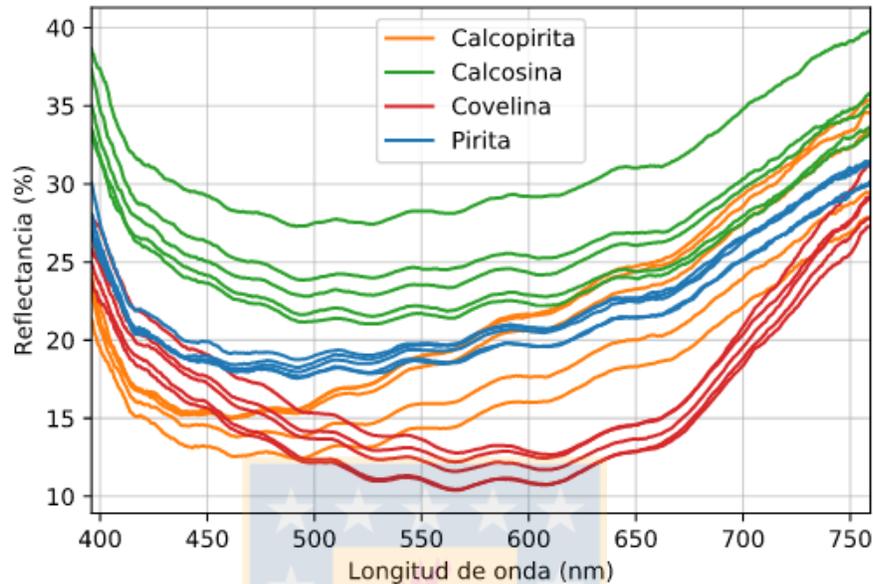


Figura 4.1: reflectancia de pellets en BF intervalo de luz visible con un área de integración de 40 μm.

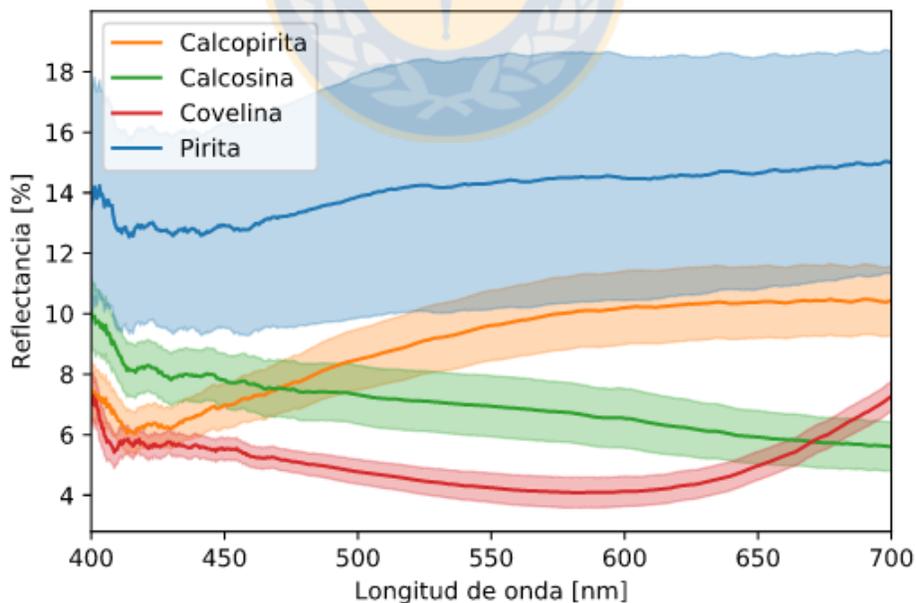


Figura 4.2: reflectancia DF en luz visible, área de integración de 40 μm. Las sombras representan desviaciones estándar.

Algo destacable de la Figura 4.3 es que, a pesar de existir dos minerales con gran varianza (pirita y calcopirita), cada curva es claramente identificable, existiendo zonas donde las mediciones se pueden distinguir de otras. Por ejemplo, a 600 nm es fácilmente distinguir covelina del resto de los minerales.

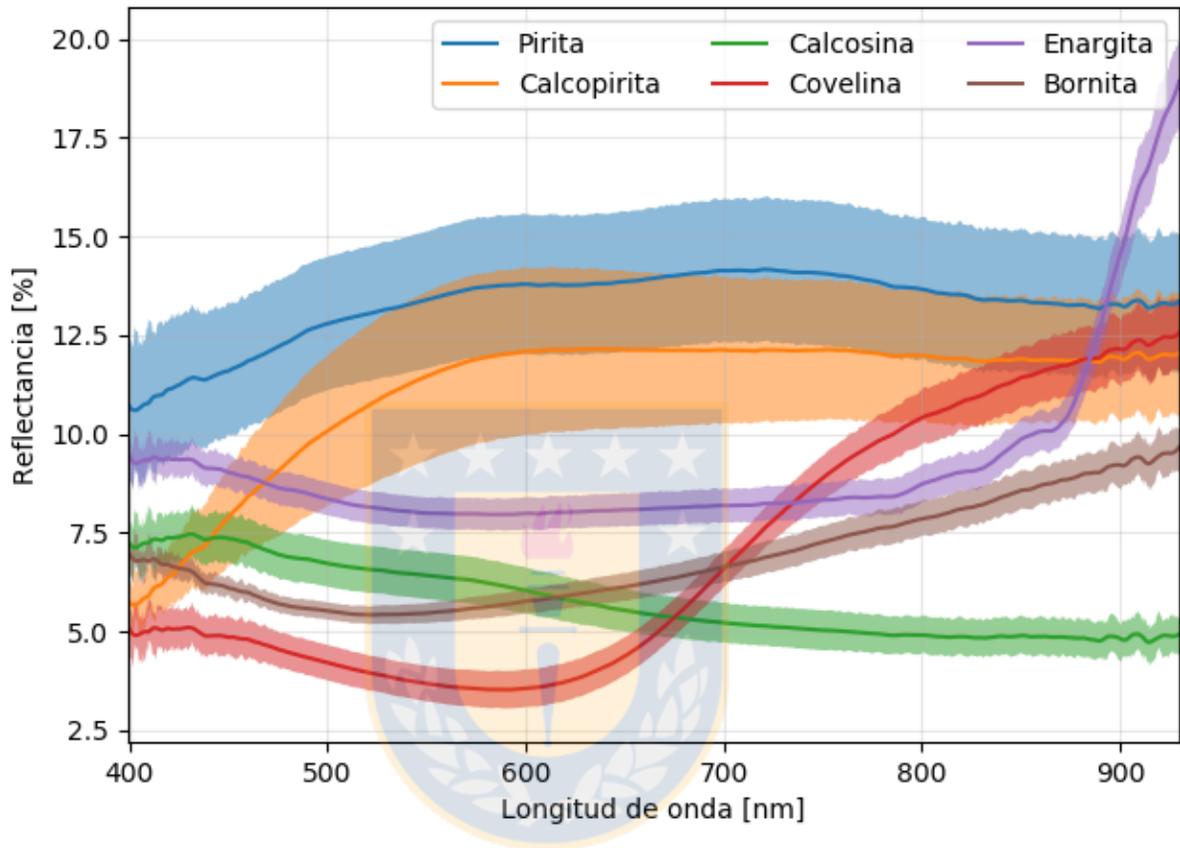


Figura 4.3: reflectancia VISNIR en campo oscuro con un área de integración de 180 μm .

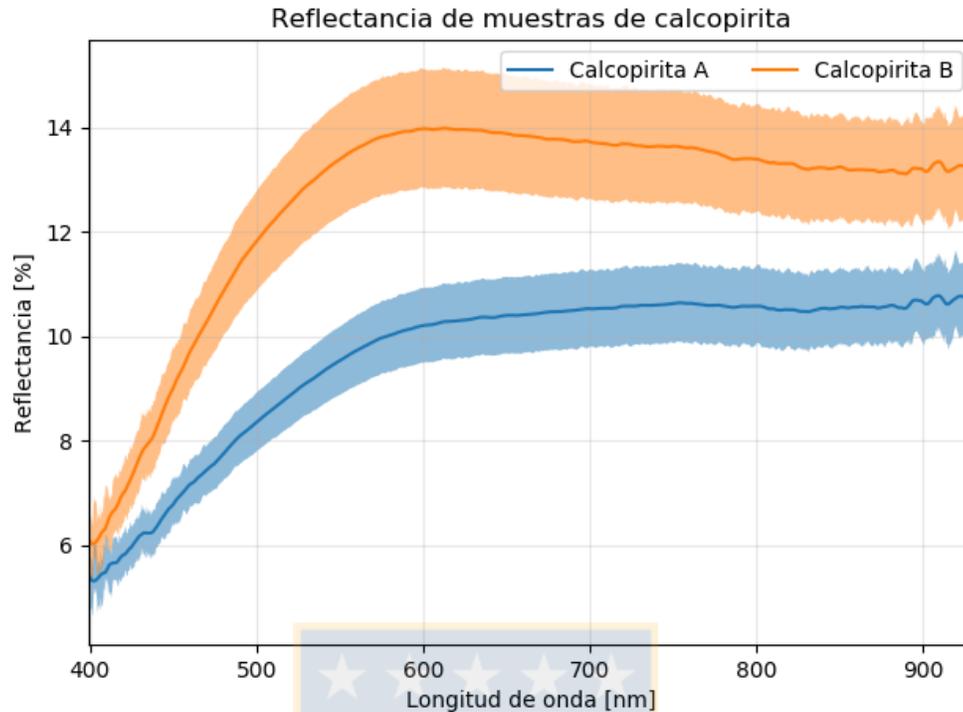


Figura 4.4: *efecto* de la granulometría en pellets de calcopirita, si bien los tamaños no están cuantificados, en el microscopio era claramente visible el mayor tamaño de partícula de la muestra A.

4.2. Análisis de curvas de reflectancia espectral y preparación de modelos

Para analizar los datos (mostrados en la Figura 4.3) se utilizó PCA en forma exploratoria, y mediante los *scores*, si es que es posible separar estadísticamente cada medición, pues mostrarían clústers de datos de mediciones similares, lo cual es mostrado en la Figura 4.5. Los *loadings* muestran zonas de alta varianza en cada componente principal, lo que se traduce en regiones donde los datos se separan y son potenciales candidatos para una longitud de onda discriminante en los datos disponibles, lo cual se analiza en la Figura 4.6, y las longitudes de onda se muestran en la Tabla 5.

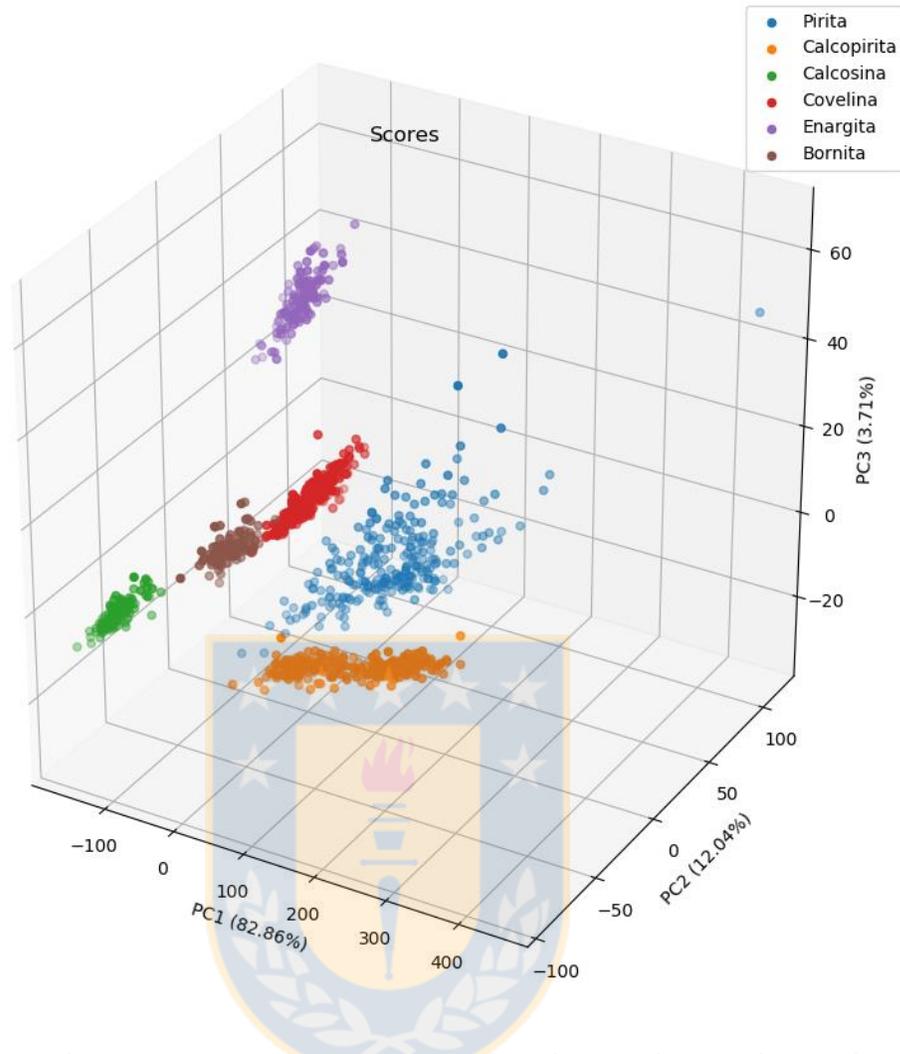


Figura 4.5: scores de tres componentes principales, los cuales explican el 98,61% de la varianza de los datos medidos.

Tabla 5: Máximos locales de cada componente principal, donde cada uno de estos corresponde a una potencial longitud de onda discriminante para clasificación.

Componente Principal	Máximo(s) en loadings [nm]	Varianza explicada del PC [%]
PC1	600	82,9
PC2	920	12,0
PC3	400, 920	3,71
PC4	790	1,25
PC5	490	0,09
PC6	875	0,02

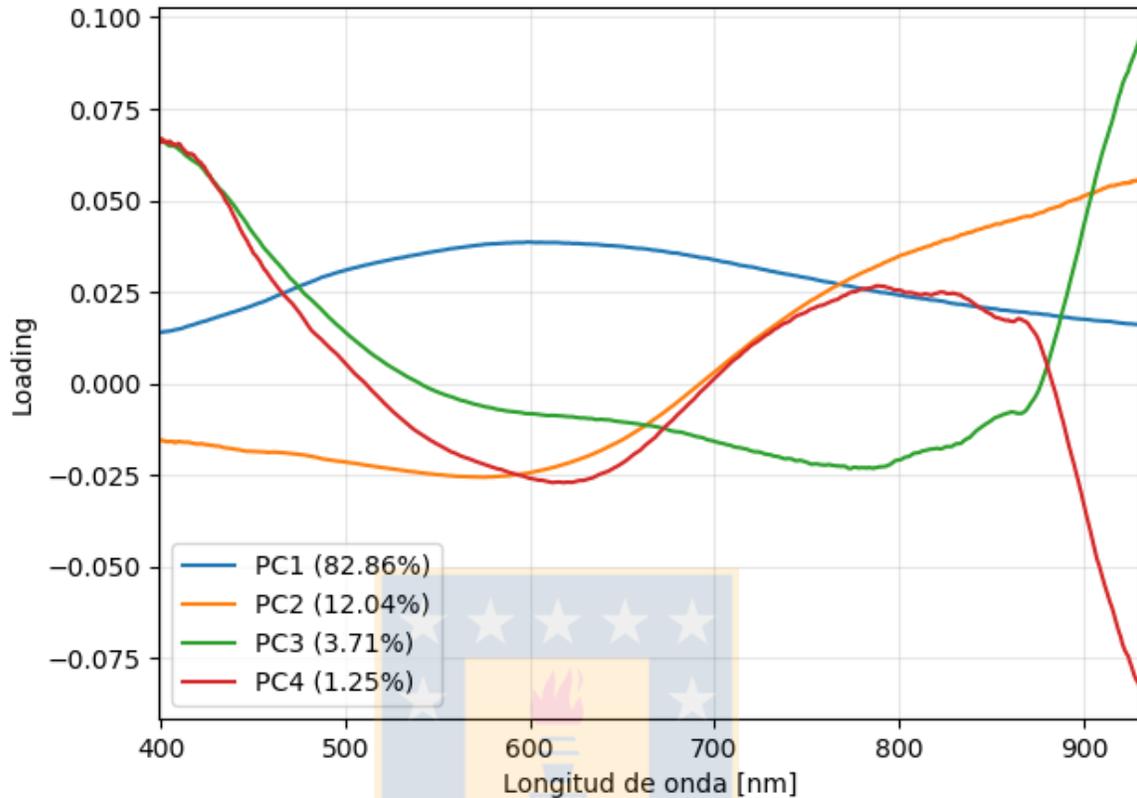


Figura 4.6: *loadings* en el sentido de varianza explicada. Se consideran los primeros cuatro componentes principales con una varianza explicada total del 99,86%.

Se determinó que con los tres primeros máximos locales (600, 920 y 400 nm) se pueden obtener modelos de clasificación con precisión aceptable, el árbol de decisiones es visible en la Figura 4.7, mientras que el resto de las pruebas de validación son visibles en la Tabla 6.

La lectura de un árbol de decisiones empieza desde el primer cuadro, donde la primera línea representa una condición, si esta condición se cumple se sigue la flecha izquierda, caso contrario se sigue la flecha derecha. La segunda línea, correspondiente a la entropía, es una medición de la clasificación de los datos, una entropía alta, mayor a 1, representa que no existe una separación de los datos, mientras una entropía cercana a 0 indica que solo existen datos de una única clase.

Tabla 6: resultados de modelos ML.

Método	Condiciones	Cross-validation [%]
Árbol de decisiones	Seis hojas como máximo, entropía como métrica	98,76
KNN	K = 20 hasta K = 70, distancia euclidiana.	99,41
SVM	Sin kernel (lineal), C = 1, One- vs-One	100 ⁶
SGD	No utilizado	No utilizado

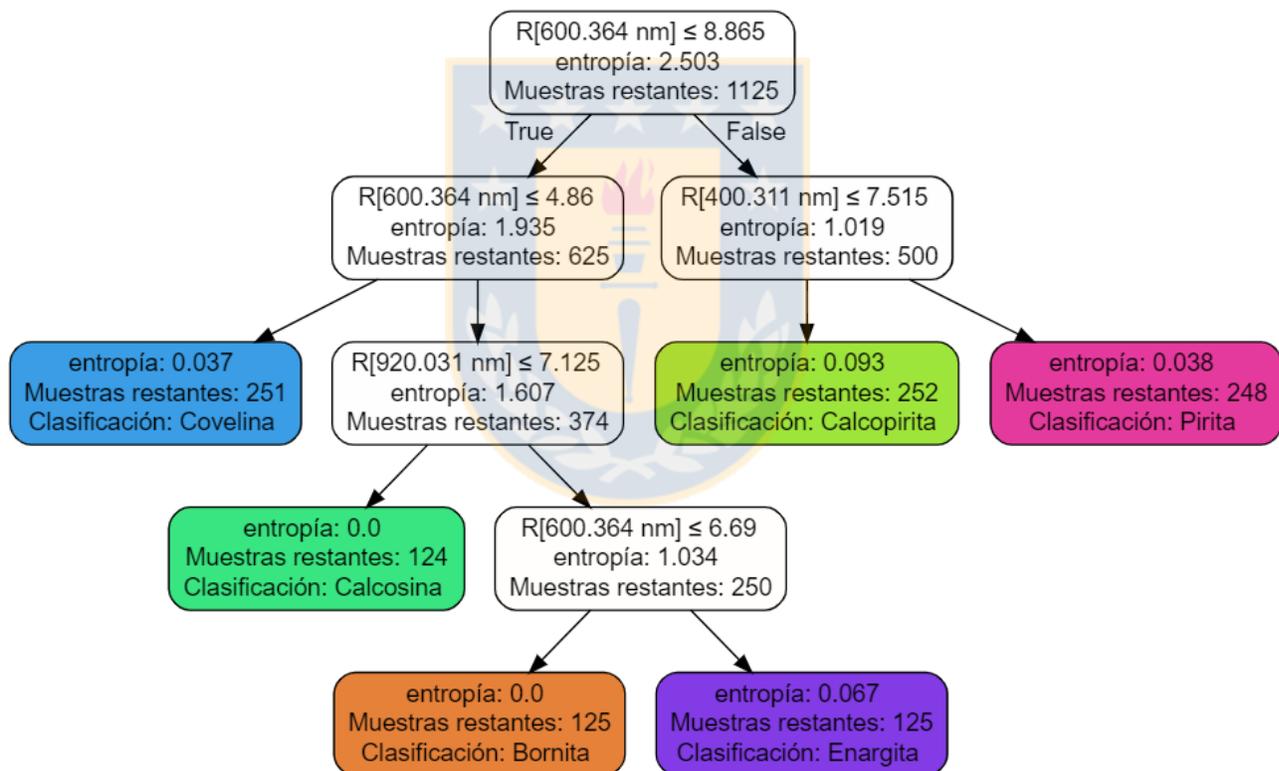


Figura 4.7: árbol de decisiones para las 1125 muestras de minerales con una validación del 98.76%. Un árbol más detallado es incluido en los anexos.

⁶ Normalmente precisiones del 100% pueden significar *overfitting*, pero considerando que el resto de los modelos tienen altos porcentajes de *cross-validation* la precisión puede considerarse válida.

4.3. Toma y análisis de imágenes de minerales (BF)

Tal como se mencionó en el inciso 3.4.2, se tomaron imágenes de las muestras de minerales a resolución 1674 x 1248 con una equivalencia en la cual 16 píxeles equivalen a 5 μm , en luz monocromática a longitudes de onda de 400, 500, 600, 800 y 900 nm. Debido al ancho de banda del monocromador, las micrografías son válidas solo entre los píxeles verticales desde 400 hasta 900, por lo que la resolución efectiva de la imagen corresponde a 500 x 1248 a la misma equivalencia métrica, para un total de 624 mil datos por cada micrografía.

Como observación, debido a que las muestras de mineral se encuentran distribuidas azarosamente en un sistema particulado, existen zonas donde el lente de la cámara se satura (entregando reflectancias mayores al 100%) debido a las caras de los cristales que reflejan especularmente la luz incidente, por lo que estos datos aportan ruido al ajuste y validación de los modelos. Esto es típico para el tipo de microscopía (BF) por lo que no es algo evitable en esta técnica.

4.3.1. Ajuste de modelos

El mejor resultado en los modelos se obtuvo utilizando luz monocromática en longitudes de onda de 400, 600 y 900 nm, similar a lo predicho en el inciso 4.2. De este análisis se puede resumir los siguientes resultados:

- El máximo puntaje de validación obtenido en el árbol de decisiones fue 25.71%, lo que, dado a que existen cuatro clasificaciones posibles (pirita, calcopirita, calcosina y covelina) no es un modelo útil y sólo entrega resultados al azar.
- Debido a limitaciones de memoria y poder de cálculo no se pudo utilizar SVM, en específico, pues este modelo escala en orden cúbico $O(3n^3)$, lo que para $n = 1.87 \cdot 10^6$ valores de entrada a punto flotante de 64 bits no se disponían de la memoria suficiente.
- SGD entregó una validación del 51,37%, por lo que los datos no son correctamente clasificados.
- KNN entregó la mayor validación (71,45%) a resolución óptima (ver inciso siguiente) con $K = 200$ y distancia Manhattan, lo que, si bien separa las mediciones al azar, esta validación no es suficiente para un modelo de ingeniería. Es posible

que este modelo generalice mejor que SGD y el árbol de decisiones debido al valor de K considerado, lo que puede suprimir en parte el efecto de los valores de ruido en la entrada.

4.3.2. Efecto resolución

La imagen fue reescalada sin técnicas de suavizado (*aliasing*) desde un 10% del tamaño original hasta el 90% en pasos de 10%. El promedio de todas las precisiones fue 70,68% con una desviación estándar de 0,47. A una escala del 50% del tamaño original se obtuvo el mayor puntaje de validación, 71.45% (p-value < 0.01) con una resolución métrica de 8 píxeles que equivalen a 5 μm (Figura 4.8). Si se aplica suavizado en el reescalado de la imagen esta es estrictamente creciente a menores resoluciones, donde al 10% del tamaño original (16 píxeles = 1 μm) la validación se determina en un 81.17%.

Debido a que el efecto del suavizado en el reescalado no es conocido en detalle, este modelo no fue considerado.

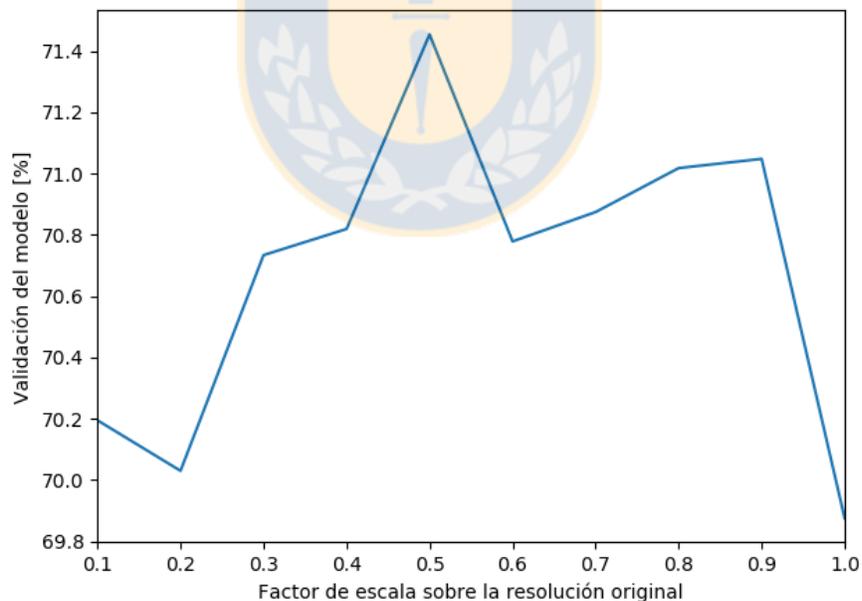


Figura 4.8: efecto de la escala de la resolución sobre la validación del modelo en BF. Es visible que a una escala del 50% el modelo tiene un máximo de validación.

4.4. Toma y análisis de imágenes de minerales (DF)

Para este caso, debido a que el uso del filtro VariSpec permitió un mayor ingreso de luz, fue posible utilizar una resolución mayor en cada fotografía, pero debido a especificaciones del equipo solo es posible trabajar en un intervalo espectral desde 400 hasta 720 nm. La resolución de cada imagen es de 3348 x 2496 píxeles, con una transformación métrica de 32 píxeles equivalentes a 5 μm . Para este caso, todos los píxeles son considerados datos válidos, teniendo un total de 8356608 datos por cada longitud de onda.

Se tomaron imágenes cada 10 nm con un ancho de banda de 10 nm, para un total de 33 imágenes por muestra, o sea, más de 275 millones de datos por cada muestra. Debido a la gran cantidad de datos fue de suma importancia determinar las longitudes de onda discriminantes dentro del intervalo espectral de operación.

En comparación a las imágenes obtenidas por microscopía BF, en este caso la cantidad de regiones donde la cámara se saturó (lo que se traduce reflectancias superiores al 100%) fue considerablemente menor. Esto se asocia al tipo de iluminación aplicada (ver Figura 2.6), donde se analizan rayos de luz dispersos.

En la Figura 4.9 se observa que los scores de cada componente principal no están clusterizados de tal forma que se pueda obtener una clasificación, sumado a que los *loadings* de cada componente principal (Figura 4.10), especialmente el primero (que tiene la mayor varianza de datos explicada), muestran que existe varianza a lo largo de todas las longitudes de onda, por lo que ninguna sería de interés estadístico para discriminar, sin embargo, se utilizó un método de "fuerza bruta", o sea, se probaron todas las combinaciones posibles, usando SGD para determinar la combinación de longitudes de onda que maximizaron el puntaje de validación de los modelos.

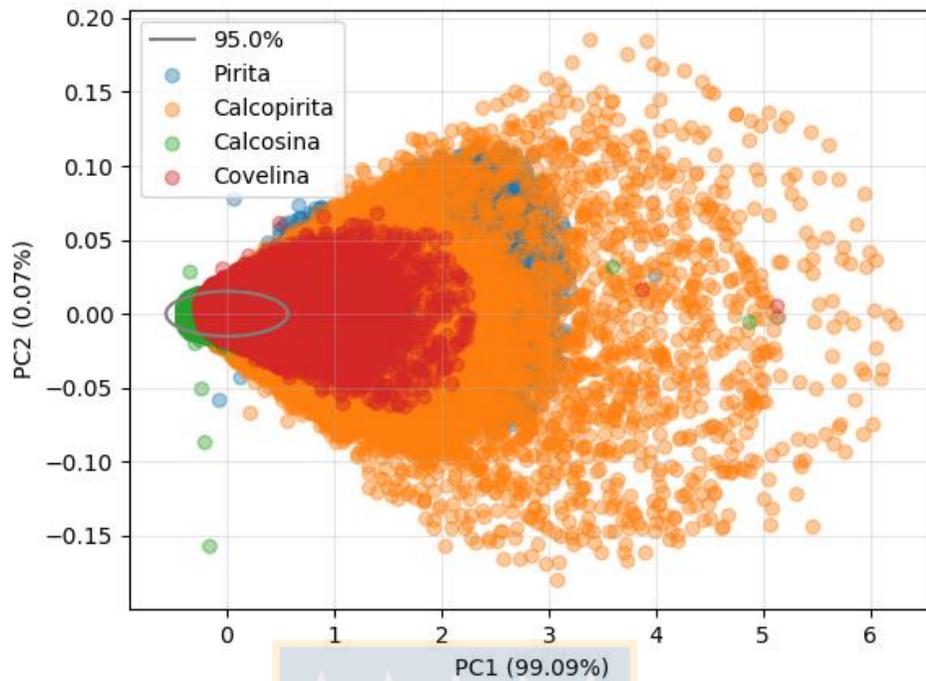


Figura 4.9: *scores* de los dos primeros componentes principales para el caso de iluminación difusa. No existen clústers claramente diferenciables sobre otros.

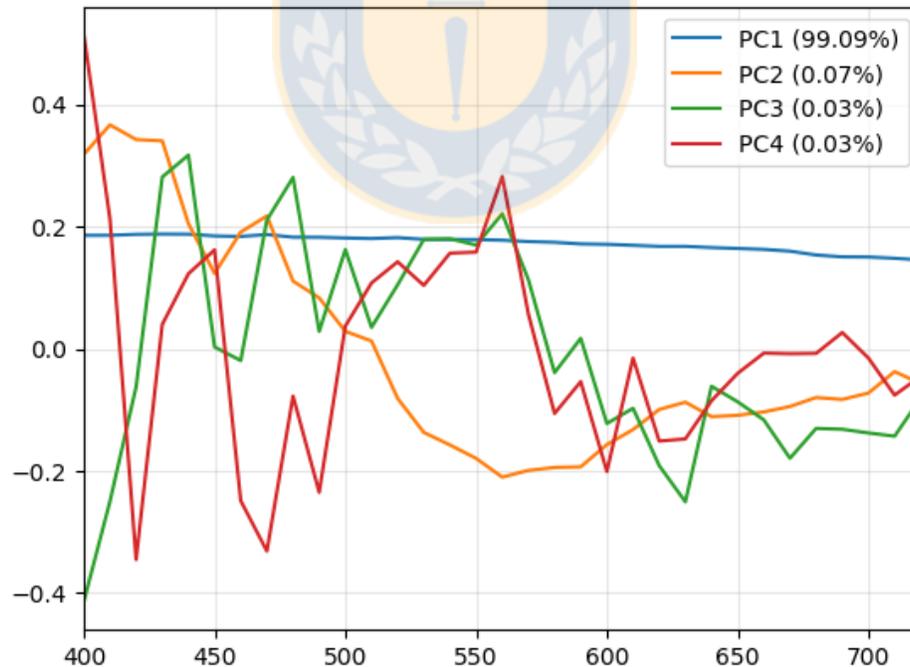


Figura 4.10: *loadings* de los primeros cuatro componentes principales para iluminación difusa. Si bien existen máximos locales visibles, estos no son explicados por componentes principales que sean de relevancia.

4.4.1. Ajuste de modelos

Anteriormente, se determinó que la segunda longitud de onda de mayor importancia en la clasificación de minerales corresponde a la banda de 920 nm (ver Tabla 5), sin embargo, debido a las especificaciones del filtro VariSpec, sólo es posible trabajar hasta 720 nm, además, como se observa en la Figura 4.10, no es posible determinar longitudes de onda discriminantes.

Debido a esto, aprovechando el bajo orden de complejidad del modelo SGD junto con su alta velocidad de ajuste del modelo⁷, se ajustaron y validaron todas las combinaciones posibles de longitudes de onda correspondientes a los *peaks* de la Figura 4.10 sumado a los *peaks* de la Tabla 5, alcanzando un total de 127 combinaciones posibles. De esta forma, la combinación de longitudes de onda discriminantes que maximizó la precisión del modelo corresponde a 400, 430 y 560 nm con un 50.13% en SGD a la resolución inicial. A partir de esto, puede considerarse lo siguiente:

- El árbol de decisiones entregó una validación del 27,31%, lo que no es lejano al azar estadístico, por lo que este modelo no se considera válido.
- Similar al caso de microscopía BF, no fue posible utilizar el modelo SVM debido a limitaciones de memoria y poder de cálculo computacional.
- SGD entregó una validación del 50,13%.
- KNN validó con un 55,47%, considerando $K = 200$ y distancia Manhattan.

4.4.2. Efecto resolución

Al igual que en el caso de BF, se alcanzó un máximo a una escala del 25% con una resolución de 8 píxeles equivalentes a 5 μm ($p\text{-value} < 0.01$), obteniéndose una validación de 58,13% en KNN con $K = 200$ y distancia Manhattan. Tener en cuenta que la imagen tomada con técnica DF tiene más resolución que la tomada en técnica BF, por lo que las escalas (en razón $\text{pixel}/\mu\text{m}$) son equivalentes para ambos casos.

⁷ Asociada a la posibilidad de realizar cálculos aprovechando computación paralela (procesadores multinúcleo).

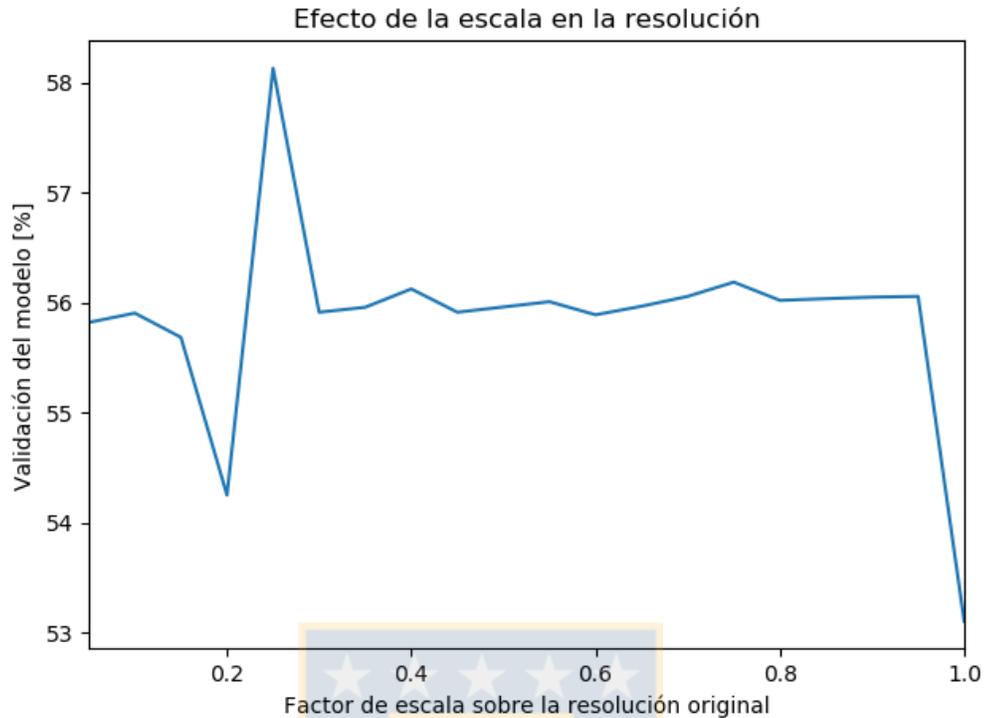


Figura 4.11: efecto de la escala de la resolución sobre la validación del modelo en DF.

4.5. Resumen de ajuste de modelos

En la Tabla 7, se presenta un resumen de la validación de cada modelo utilizado en ambas técnicas de microscopía. Para ambos casos se considera la resolución óptima determinada anteriormente con sus debidas condiciones.

Tabla 7: resumen de puntajes de validación para cada técnica con distintos modelos.

Técnica	Árbol de decisiones	KNN	SGD
BF	25.71%	71.45%	51.37%
DF	27.31%	58.13%	55.21%

4.6. Resultados de análisis de concentrados

En la Tabla 8 se encuentran los resultados de la aplicación del modelo basado en KNN a los concentrados con composición elemental conocida (ver Tabla 3). Los errores elementales varían desde el 8.05% hasta el 77.62%, pues es evidente que la cantidad de hierro (principal componente de la pirita) es muy elevada en comparación al cobre.

De la Figura 4.3 es visible que tanto la pirita como calcopirita a partir de aproximadamente los 550 nm de longitud de onda no son diferenciables entre sí, por lo que entre esos dos minerales la longitud de onda discriminante corresponde a los 400 nm. Esto representa un problema por dos motivos:

1. Debido a que la fuente de luz es halógena, la emisión de luz en el intervalo 390 hasta 410 nm es reducida, para compensar esto, las imágenes capturadas por la cámara tuvieron tiempos de exposición prolongados, lo que aumentó significativamente el ruido en las mediciones.
2. Una alta varianza de datos en la banda de 400 nm, tanto por el tipo de iluminación, las diferencias de tamaño de partícula y el ruido asociado al tiempo de exposición puede hacer que la pirita y calcopirita sean indistinguibles.

Un punto relevante al momento de calcular los errores de medición es que, solo se consideraron los elementos cobre, hierro y azufre, sin embargo, de la composición elemental obtenida en la Tabla 3, se puede inferir que existe cuarzo (SiO_2) en los concentrados. El cuarzo es un mineral que permite la transmisión de luz, incluso para tamaños de partícula elevados, lo que agrega imprecisiones a la aplicación del modelo.

Si bien en general el error en las mediciones elementales del cobre es menor al del hierro, incluso alcanzando errores del 3%, el promedio del error de todas las mediciones alcanza el 37.73% con una desviación estándar de 6.16%.

Aparte de las ya mencionadas, una fuente de error adicional del modelo es la consideración de minerales. Es posible que existan minerales adicionales que no fueron considerados.

Tabla 8: Resultados de modelo KNN en iluminación BF para longitudes de onda de 400, 600 y 900 nm.

Muestra	Composición mineralógica calculada [%]		Composición elemental calculada [%]		Composición elemental medida [%]		Error relativo [%]
M-01	CuFeS ₂	51,9	Cu	18,9	Cu	23,3	22,0
	FeS ₂	47,8	Fe	38,1	Fe	25,7	48,0
	Cu ₂ S	0,16	S	43,8	S	35,1	24,7
	CuS	0,14					
M-02	CuFeS ₂	51,7	Cu	22,5	Cu	26,0	13,3
	FeS ₂	42,1	Fe	35,3	Fe	23,8	48,4
	Cu ₂ S	3,87	S	42,1	S	28,9	45,8
	CuS	2,34					
M-03	CuFeS ₂	64,4	Cu	22,6	Cu	25,5	11,3
	FeS ₂	35,2	Fe	35,9	Fe	22,9	57,0
	Cu ₂ S	0,31	S	41,4	S	29,2	41,8
	CuS	0,15					
M-04	CuFeS ₂	49,2	Cu	18,3	Cu	23,2	51,1
	FeS ₂	49,0	Fe	37,8	Fe	24,1	56,8
	Cu ₂ S	0,64	S	43,9	S	29,2	54,1
	CuS	1,13					
M-05	CuFeS ₂	65,3	Cu	23,1	Cu	23,9	3,25
	FeS ₂	34,0	Fe	35,7	Fe	24,1	48,2
	Cu ₂ S	0,46	S	41,2	S	29,7	38,6
	CuS	0,22					
M-06	CuFeS ₂	59,0	Cu	21,1	Cu	22,9	8,05
	FeS ₂	40,1	Fe	36,6	Fe	23,0	59,3
	Cu ₂ S	0,37	S	42,3	S	29,0	45,9
	CuS	0,50					

Muestra	Composición mineralógica calculada [%]		Composición elemental calculada [%]		Composición elemental medida [%]		Error [%]
M-07	CuFeS ₂	37.0	Cu	14.1	Cu	23.3	39.7
	FeS ₂	61.3	Fe	39.8	Fe	23.0	72.9
	Cu ₂ S	0.62	S	46.2	S	32.7	41.2
	CuS	1.13					
M-08	CuFeS ₂	49.0	Cu	17.9	Cu	23.6	24.0
	FeS ₂	49.7	Fe	38.1	Fe	23.7	60.6
	Cu ₂ S	0.94	S	44.0	S	30.8	42.9
	CuS	0.33					
M-09	CuFeS ₂	56.6	Cu	20.9	Cu	23.6	11.3
	FeS ₂	41.7	Fe	36.6	Fe	23.7	54.5
	Cu ₂ S	1.37	S	42.5	S	28.9	46.9
	CuS	0.37					
M-10	CuFeS ₂	80.2	Cu	28.5	Cu	23.9	19.1
	FeS ₂	18.8	Fe	33.2	Fe	23.5	41.1
	Cu ₂ S	0.31	S	37.4	S	30.0	27.9
	CuS	0.67					
M-11	CuFeS ₂	64.3	Cu	22.5	Cu	23.6	4.8
	FeS ₂	35.4	Fe	30.1	Fe	20.3	77.6
	Cu ₂ S	0.10	S	41.5	S	31.2	33.0
	CuS	0.17					

5 CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones y recomendaciones generales

Si bien en estudios anteriores se ha demostrado que es posible hacer una caracterización mineralógica utilizando reflectancia espectral en el intervalo VISNIR, en esos estudios se detallaba la importancia de la preparación de muestras (Ver inciso 2.5, donde se destaca el estudio de Chopard [34]) para obtener una reflectancia de tipo especular a través de un pulido completo.

Sin embargo, el objetivo del presente informe fue determinar si existe la posibilidad de clasificación en pellets sin pulido, donde para minerales aislados (pellets de muestras) esto es completamente posible, con altos nivel de precisión (>99%) siempre que el área de integración del espectrómetro sea lo suficientemente grande en comparación al tamaño de partícula. A menores áreas de integración del espectrómetro aumenta la varianza en las mediciones.

Si bien el árbol de decisiones no entregó buenos resultados en las imágenes de minerales, este obtuvo resultados aceptables para el análisis con el espectrómetro. La principal ventaja del árbol de decisiones es que puede graficarse (Figura 4.7), por lo que pueden identificarse fácilmente las longitudes de onda que distinguen a un mineral de otro y a qué reflectancia.

Los promedios de errores en ley de cobre, hierro y azufre son 18,9%, 60,7% y 40,1% respectivamente, esto puede deberse a un sesgo en el modelo que sobrerrepresenta la pirita. Esto se debe a que la longitud de onda que diferencia la pirita de la calcopirita es la banda de 400 nm, donde existe un considerable nivel de ruido, gracias a los altos tiempos de exposición. Sin embargo, de esto se puede ver que el error en la ley de cobre no es, necesariamente, una medición inválida.

Existen concentrados con características similares: M05 y M06. De esto se esperaría que los resultados del modelo no tengan una gran varianza, lo que es visible en los errores de la ley de cobre, correspondientes a 3,25% y 8,05% respectivamente. En contraparte, el concentrado M04, que también posee características similares, tiene un error en la ley de cobre superior al 50%, donde es probable que la pirita haya sido sobrerrepresentada.

Con respecto al ajuste de modelos para aplicar en imágenes monocromáticas de concentrados, se determina que:

1. El costo computacional para imágenes de alta resolución es elevado, en el caso de SGD y SVM el ajuste del modelo puede realizarse en un equipo con especificaciones -procesador, memoria disponible- adecuadas, puesto que la aplicación de estos modelos en datos es trivial y no requiere de gran poder de cálculo. Sin embargo, como KNN no genera un modelo y sólo hace comparaciones, el equipo encargado de aplicar el algoritmo debe tener las especificaciones necesarias.
2. Es de suma relevancia determinar el tipo de iluminación y tiempo de exposición para no agregar datos saturados (reflectancia superior al 100%) a los modelos, puestos que estos datos contribuyen al ruido y la imprecisión de los mismos. Utilizar microscopía DF reduce considerablemente los píxeles saturados.
3. Determinar una resolución óptima ayuda tanto en la precisión del modelo como en el uso de memoria, a menores resoluciones se requiere menor memoria y tiempo de procesamiento de datos.

En cuanto a la aplicación de los modelos se deben considerar las limitaciones del ordenador que realiza los cálculos, si bien en SGD esto no es un problema, KNN no escala bien a situaciones con un elevado número de datos (ver inciso 2.4.3), ya que se debe mantener en la memoria los datos para comparar, una lista de todas las distancias medidas, los datos por medir y los resultados para cada valor de entrada. Si se utiliza una gran cantidad de datos es fácil utilizar la memoria disponible del sistema. SVM no pudo utilizarse en este caso debido al orden de complejidad del modelo.

5.2. Recomendaciones para la continuidad del estudio

1. Se mostró que podría existir un efecto producido por la razón entre el tamaño de la partícula, por lo que es de interés investigar cómo los tamaños de partícula afectan las mediciones de reflectancia, tanto en minerales como en concentrados (ver Figura 4.4 y Figura 7.1 para caso de mineral a distinto tamaño de partícula). Esto puede lograrse mediante el control de la distribución de tamaños de partículas de la muestra y alterando el área de integración usando diversos lentes acoplados a la fibra óptica en el caso del espectrómetro y distintas magnificaciones en la toma de fotografías.

2. Si para el caso de microscopía BF se limita al intervalo de luz en el espectro visible (400 a 720 nm), la validación no supera el 40%, mientras que usando DF esta alcanzó 58,13%, sin embargo, al extender el intervalo de luz hasta 900 nm en el caso de microscopía BF la validación alcanza el 71,45%. De esto se puede teorizar que utilizar iluminación difusa llegando al intervalo infrarrojo debería generar un modelo que generalice de mejor forma los datos. Para lograr esto se requiere utilizar filtros que operen en el intervalo de luz infrarroja.
3. Además de los modelos y algoritmos de clasificación utilizados (DT, KNN, SGD y SVM) se puede considerar el uso de redes neuronales o diseñar un modelo en particular para este caso que optimice el uso de recursos computacionales.
4. Es de gran importancia conocer la composición mineralógica, además de la elemental para que la comparación entre los resultados del modelo y la composición real sea más cercana a la realidad.

5.3. Línea de investigación complementaria

Un método que no fue estudiado en el presente informe y que podría ser digno de investigación es realizar una iluminación en un área mayor, cercana a un centímetro de diámetro y estudiar el espectro reflectivo de la muestra en general y no a nivel partícula. En la Figura 5.1 se muestran los espectros de reflectancia de las muestras de la Tabla 3 bajo dichas condiciones. En la Figura 5.2 se muestran los espectros de reflectancia de los pellets de minerales. Un enfoque que podría aplicarse es la posibilidad de descomponer el espectro de un concentrado de cobre en combinaciones lineales de los espectros de minerales que lo componen, esto debido a que el espectro obtenido en un área mayor promedia los espectros individuales de cada partícula.

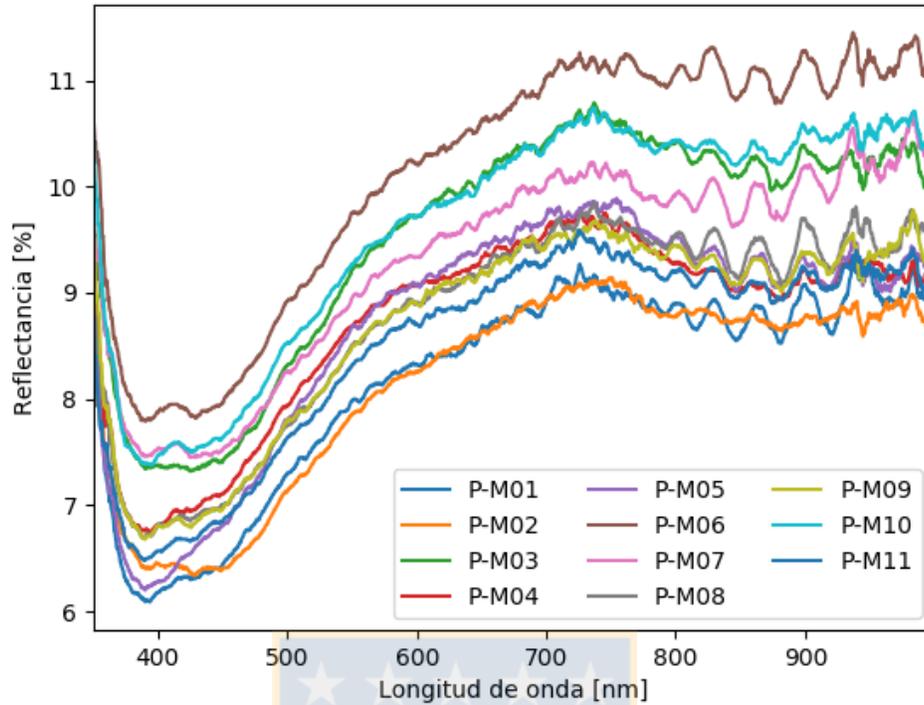


Figura 5.1: espectros de reflectancia de concentrados de cobre, tomados bajo el área completa de una superficie del pellet.

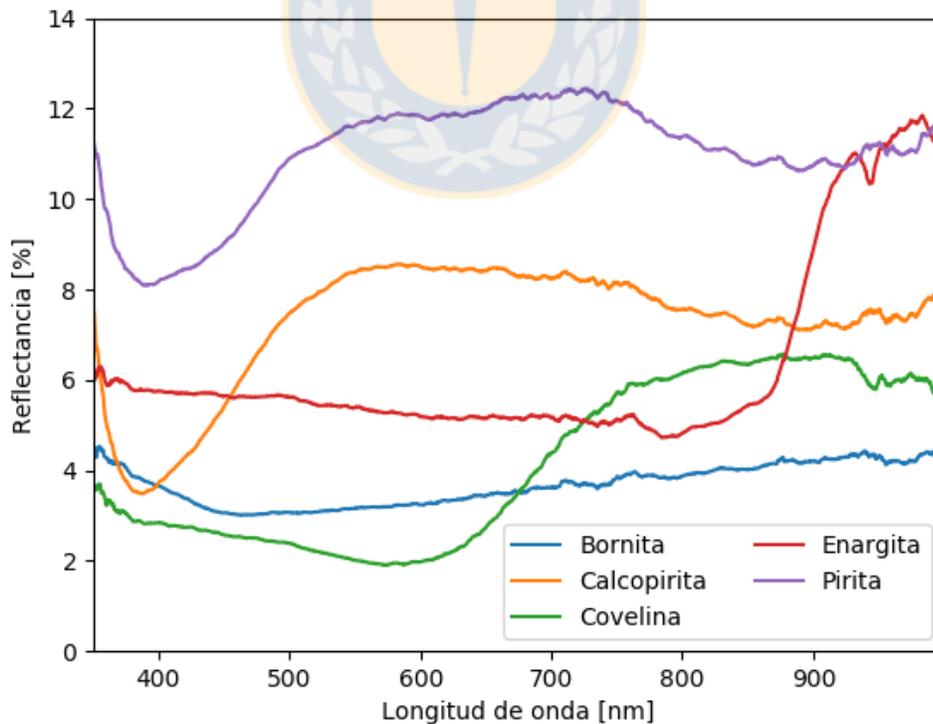


Figura 5.2: Espectros de reflectancia de pellets de minerales, tomadas bajo el área completa de la superficie del pellet.

6 BIBLIOGRAFÍA Y REFERENCIAS

- [1] Boric, R., Diaz, J., *et al.* (2009). *Geology of the Ministro Hales Mine (MMH), Chuquicamata District, Chile*. Paper presented at the Actas XII Congreso Geológico Chileno, Santiago (electronic version).
- [2] Kortüm, G. (1969). *Reflectance spectroscopy: principles, methods, applications*: Springer Science & Business Media.
- [3] Avantes. (2019). AvaLight-HAL-S-Mini Compact Halogen Light Source. Retrieved 2019-11-10, from <https://www.avantes.com/products/light-sources/item/901-avalight-hal-s-mini>
- [4] ThorLabs. (2019). Stabilized Deuterium UV Light Source. Retrieved 2019-11-10, from https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=11783
- [5] Lu, V., & Park, J. (2016). Tunable Light Sources: A Popular Choice for Measurement Applications. Retrieved 2019-11-10, from https://www.photonics.com/Articles/Tunable_Light_Sources_A_Popular_Choice_for/a58302
- [6] Dunn, W. J., & Bush, A. C. (2002). A comparison of polymerization by light-emitting diode and halogen-based light-curing units. *J Am Dent Assoc*, 133(3), 335-341. doi: 10.14219/jada.archive.2002.0173
- [7] Zeiss. (2011). Tungsten-halogen lamp information. Retrieved 2019-11-10, from <https://web.archive.org/web/20110303211343/http://zeiss-campus.magnet.fsu.edu/articles/lightsources/tungstenhalogen.html>
- [8] Craig, J. R., Vaughan, D. J., *et al.* (1981). *Ore microscopy and ore petrography* (Vol. 406): Wiley New York.
- [9] Piller, H. (1977). *Microscope Photometry*: Springer, Berlin, Heidelberg.
- [10] Chuang, Y.-C. (2015). Monochromator *Encyclopedia of Color Science and Technology* (pp. 1-4).
- [11] WikimediaCommons. (2018). File: Diagram of a monochromator.png. Retrieved 2019-11-10, from https://commons.wikimedia.org/w/index.php?title=File:Diagram_of_a_monochromator.png&oldid=280978563
- [12] Beeckman, J. (2011). Liquid-crystal photonic applications. *Optical Engineering*, 50(8). doi: 10.1117/1.3565046

- [13] Beeckman, J., Hui, T., *et al.* (2009). Polarization Selective Wavelength Tunable Filter. *Molecular Crystals and Liquid Crystals*, 502(1), 19-28. doi: 10.1080/15421400902813626
- [14] Fellers, T. J., & Davidson, M. W. (n.d.). Acousto-Optic Tunable Filters (AOTFs). Retrieved 2019-11-11, from <https://www.olympus-lifescience.com/en/microscope-resource/primer/techniques/confocal/aotfintro/>
- [15] Harris, S. E., & Wallace, R. W. (1969). Acousto-Optic Tunable Filter. *Journal of the Optical Society of America*, 59(6). doi: 10.1364/josa.59.000744
- [16] Qin, J., Chao, K., *et al.* (2013). Hyperspectral and multispectral imaging for evaluating food safety and quality. *Journal of Food Engineering*, 118(2), 157-171. doi: 10.1016/j.jfoodeng.2013.04.001
- [17] Butler, L. R. P., & Laqua, K. (1995). Nomenclature, symbols, units and their usage in spectrochemical analysis-IX. Instrumentation for the spectral dispersion and isolation of optical radiation (IUPAC Recommendations 1995). *Pure and Applied Chemistry*, 67(10), 1725-1744. doi: /10.1351/pac199567101725.
- [18] Ocean Optics, I. (2019). WS-1 Reflectance Standards. Retrieved 2019-11-04, from <https://oceanoptics.com/product/ws-1-reflectance-standards/>
- [19] Criddle, A. J., & Stanley, C. J. (1993). *Quantitative Data File for Ore Minerals* (3rd ed.): Springer Netherlands.
- [20] Cormen, T. H., Leiserson, C. E., *et al.* (2009). *Introduction to algorithms*: MIT press.
- [21] Brereton, R. G. (2003). *Chemometrics: data analysis for the laboratory and chemical plant*: John Wiley & Sons.
- [22] Madsen, R. E., Hansen, L. K., *et al.* (2004). Singular value decomposition and principal component analysis. *Neural Networks*, 1, 1-5.
- [23] Wang, Z. (2019). PCA and SVD explained with numpy. Retrieved 2019-10-14, from <https://towardsdatascience.com/pca-and-svd-explained-with-numpy-5d13b0d2a4d8>
- [24] Mohri, M., Rostamizadeh, A., *et al.* (2018). *Foundations of machine learning*: MIT press.
- [25] Pedregosa, F., Varoquaux, G., *et al.* (2019). Decision Trees. Retrieved 2019-10-13, from <https://scikit-learn.org/stable/modules/tree.html>

- [26] Pedregosa, F., Varoquaux, G., *et al.* (2019). Nearest Neighbors. Retrieved 2019-10-15, from <https://scikit-learn.org/stable/modules/neighbors.html#classification>
- [27] Vashist, A., & Gupta, S. (2018). Pros and Cons of K-Nearest Neighbors. Retrieved 2019-10-15, from <https://www.fromthegenesis.com/pros-and-cons-of-k-nearest-neighbors/>
- [28] Pedregosa, F., Varoquaux, G., *et al.* (2019). Support Vector Machines. Retrieved 2019-10-15, from <https://scikit-learn.org/stable/modules/svm.html>
- [29] StanfordUniversity. (2013). Optimization: Stochastic Gradient Descent. Retrieved 2015-10-15, from <http://deeplearning.stanford.edu/tutorial/supervised/OptimizationStochasticGradientDescent/>
- [30] Pedregosa, F., Varoquaux, G., *et al.* (2019). Stochastic Gradient Descent. Retrieved 2019-10-15, from <https://scikit-learn.org/stable/modules/sgd.html>
- [31] Bishop, C. M. (2006). *Pattern recognition and machine learning*: Springer.
- [32] Aly, M. (2005). Survey on multiclass classification methods. *Neural Networks*, 19, 1-9.
- [33] Allwein, E. L., Schapire, R. E., *et al.* (2000). Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of machine learning research*, 1(Dec), 113-141.
- [34] Chopard, A., Marion, P., *et al.* (2019). Automated sulfides quantification by multispectral optical microscopy. *Minerals Engineering*, 131, 38-50. doi: 10.1016/j.mineng.2018.11.005
- [35] Castroviejo, R., & Berrezueta, E. (2010). Reconocimiento automatizado de menas metálicas mediante análisis digital de imagen: un apoyo al proceso mineralúrgico. II: criterios metalogenéticos discriminantes. *Revista de Metalurgia*, 45(6), 439-456. doi: 10.3989/revmetalm.0923
- [36] Catalina, J. C., & Castroviejo, R. (2017). Microscopía de reflectancia multiespectral: Aplicación al reconocimiento automatizado de menas metálicas. *Revista de Metalurgia*, 53(4). doi: 10.3989/revmetalm.107
- [37] Castroviejo, R., Espi, J. A., *et al.* (2011). España Patent No. ES2368321B2. OEPM.

7 ANEXOS

7.1. Abreviaciones de esquemas de luz en microscopios

Abbreviations, symbols and signs: — marginal rays; — principal rays; -·-·-·-·-·-· optical axis; marginal rays forming maximum aperture angles.

The lenses or lens systems including field lenses are indicated by *thick bars* normal to the optical axis.

▬ diaphragms; ▭ real images of diaphragms;  luminous surface of the lamp;  real images of luminous surface; ← stage object; ← real images of stage object.

Dotted straight lines: a virtual image or a virtual ray. *Oblique hatching below or behind a line:* a mirror.

Symbols: u effective aperture angle, object side; u_{\max} potential aperture angle, object side, w field angle; v obliquity of illumination.

Note: For u , v , and w see also Figure 2.2.

Lenses: Aux auxiliary lens; Bl Bertrand lens; Col collector; Con condenser; Obj objective; Ocl ocular; Pl photosensor lens including the lens of the eye; Tl(\pm) tube lens, + positive, - negative.

- L: Luz
- CAD: Diafragma condensador-apertura
- OAD: Diafragma objetivo-apertura
- LFD: Diafragma de campo luminoso
- Obj: Objetivo
- Col: Colector
- Aux: Lente auxiliar
- Con: Condensador
- St: Imagen primaria

La notación prima (L') corresponde a imágenes formadas. [9]

7.2. Figuras

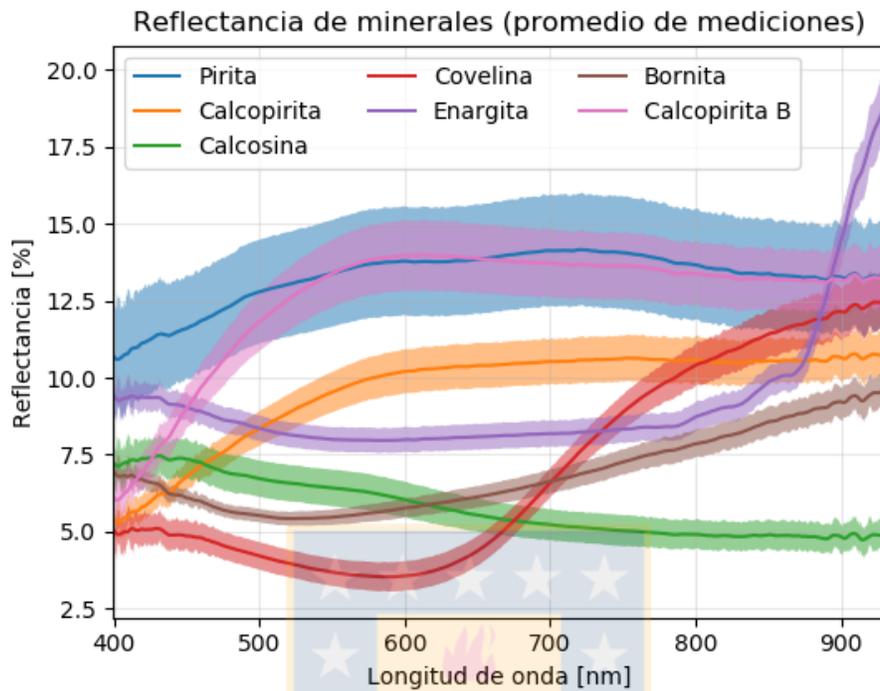


Figura 7.1: Separación de muestras de calcopirita, donde existe una muestra con mayor tamaño de partícula (naranja) que la otra (rosado).

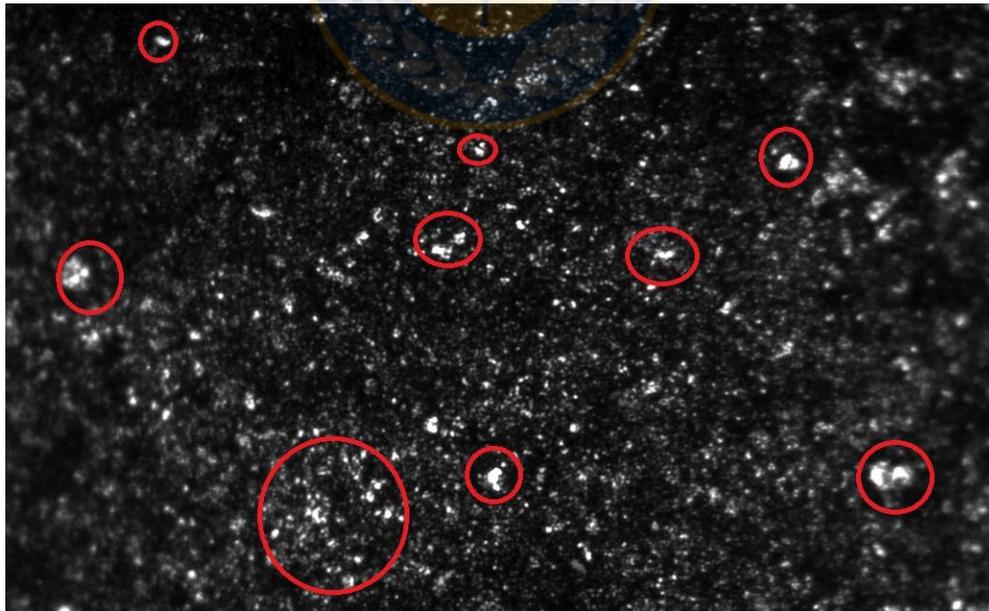


Figura 7.2: Micrografía de calcopirita bajo luz monocromática a 600 nm en iluminación especular (microscopia en campo claro), marcadas zonas donde el lente de la cámara se saturó debido a la luz reflejada por caras de mineral.

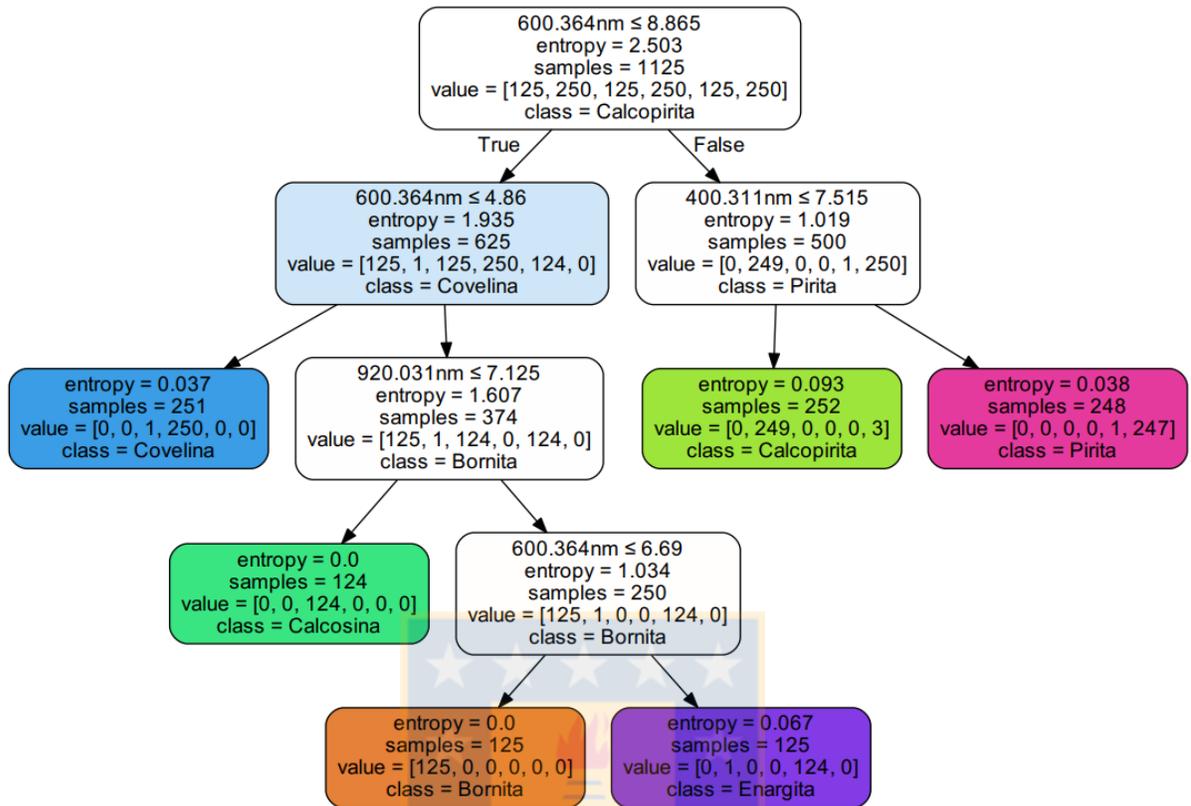


Figura 7.3: Árbol de decisiones detallado, en la primera línea se muestra la condición de separación (no se muestra en hojas), segunda línea la entropía del nodo, tercera línea el número de muestras restantes, cuarta línea el número de muestras por clase (en orden: bornita, calcopirita, calcosina, covelina, enargita y pirita), en la última línea se muestra la clasificación.

Hoja Resumen Memoria de Título

Título: Análisis químico cuantitativo de concentrados de cobre utilizando microscopía óptica multiespectral
Nombre Memorista: Germán Andrés Henríquez Véjar

Modalidad		Profesor(es) Patrocinante(s)	
Concepto			
Calificación			
Fecha		Prof. Daniel Sbárbaro Hofer	Prof. Eduardo Balladares Varela
Prof. Eugenia Araneda H.		Institución: Universidad de Concepción	

Comisión
Prof. Víctor Parra S.

Resumen
<p>El estudio tuvo por objetivo determinar la composición mineralógica de pellets de concentrados de cobre a partir del uso de reflectancia multiespectral. Para esto se estudiaron curvas de reflectancia espectral de minerales de pirita, calcopirita, calcosina y covelina con el fin de obtener longitudes de onda que permitan clasificar los minerales.</p> <p>Se determinó que las longitudes de onda que permiten una clasificación corresponden a 400, 600 y 900 nm, donde el mejor resultado se obtuvo utilizando el modelo SVM en análisis con espectrómetro, mientras que para el estudio con fotografías el mejor resultado lo entregó el algoritmo <i>K Nearest Neighbors</i>.</p> <p>En el estudio de concentrados se determinó experimentalmente que el error de las leyes de cobre, hierro y azufre son 18,9%, 60,7% y 40,1% respectivamente, por lo que, si bien, no es un modelo certero, puede entregar un primer acercamiento a la ley de cobre del concentrado.</p> <p>Se recomienda continuar el estudio considerando el uso de redes neuronales como modelo de clasificación, teniendo un mayor control sobre la granulometría de las muestras y conociendo la caracterización mineralógica de los concentrados.</p>