

**DISEÑO Y CONSTRUCCIÓN DE UNA APLICACIÓN PARA LA IDENTIFICACIÓN DE
MINERALES A PARTIR DE ESPECTROS EDS DE MICROSCOPIA ELECTRÓNICA
DE BARRIDO**

CARLOS ARTURO MEZA GARCIA

ALFREDO NIETO GUTIERREZ

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA, SANTANDER
2026**

**DISEÑO Y CONSTRUCCIÓN DE UNA APLICACIÓN PARA LA IDENTIFICACIÓN DE
MINERALES A PARTIR DE ESPECTROS EDS DE MICROSCOPIA ELECTRÓNICA
DE BARRIDO**

CARLOS ARTURO MEZA GARCIA

ALFREDO NIETO GUTIERREZ

**TRABAJO DE GRADO PARA OPTAR POR EL TÍTULO DE INGENIERO DE
SISTEMAS**

DIRECTOR

JATHINSON MENESES MENDOZA

INGENIERO DE SISTEMAS

CODIRECTOR

CARLOS ALBERTO VILLAREAL JAIMES

GEÓLOGO

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA, SANTANDER
2026**

CONTENIDO

	Pág.
INTRODUCCIÓN	11
1. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA	12
1.1 PLANTEAMIENTO DEL PROBLEMA	12
1.2 JUSTIFICACIÓN	12
1.3 ACLARACIÓN SOBRE LOS DATOS	13
2. OBJETIVOS	14
2.1 OBJETIVO GENERAL	14
2.2 OBJETIVOS ESPECÍFICOS	14
3. MARCO DE REFERENCIA	15
3.1 MICROSCOPIA ELECTRONICA DE BARRIDO (MEB)	15
3.2 ESPECTROSCOPIA DE ENERGIA DISPERSIVA (EDS)	15
3.3 ALGORITMOS DE PROCESAMIENTO DE IMAGENES Y RECONOCIMIENTO DE PATRONES	16
3.4 ANTECEDENTES DEL TEMA	16
4. METODOLOGIA	18
4.1 DISEÑO DE TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN	18
4.2 POBLACIÓN Y MUESTRAS	18
4.3 TÉCNICAS DE ANÁLISIS	18
4.3.1 Extracción de características espectrales	19
4.3.2 Comparación espectral	19
4.3.3 Estimación de abundancias	19
4.3.4 Clasificación	19

4.3.5 Interfaz de usuario	19
4.4 PROCEDIMIENTOS DE ANÁLISIS E INTERPRETACIÓN DE DATOS	19
4.4.1 Comparación de espectros	19
4.4.2 Interpretación de resultados	19
5. RESULTADOS Y DISCUSIÓN	20
5.1 DESCRIPCIÓN DE LOS MODULOS DEL SISTEMA	20
5.2 DISEÑO Y ESTRUCTURACIÓN DE LA BASE DE DATOS	22
5.2.1 Estructura de Datos	23
5.2.2 Composición de la Base de Datos	24
5.3 SISTEMA DE COMPARACIÓN ESPECTRAL	25
5.3.1 Justificación del Algoritmo	26
5.3.2 Pipeline de Procesamiento	26
5.4 DESARROLLO DE LA APLICACIÓN	27
5.4.1 Arquitectura del Sistema	27
5.5 VALIDACIÓN Y ANALISIS DE DESEMPEÑO	28
5.5.1 Conjunto de Validación	29
5.5.2 Métricas de Desempeño	29
5.5.3 Análisis de Resultados	30
5.6 ALGORITMO	30
5.6.1 Extracción de la Imagen del Espectro	30
5.6.2 Preprocesamiento de la Imagen	31
5.6.3 Segmentación del Espectro	31
5.6.4 Extracción de la Firma Espectral	31
5.6.5 Normalización del Vector	31
5.6.6 Comparación con Base de Datos	32
5.6.7 Interpretación de Resultados	32
5.7 VENTAJAS Y LIMITACIONES	32
6. CONCLUSIONES Y TRABAJO A FUTURO	34

6.1 CONCLUSIONES	34
6.2 TRABAJO A FUTURO	35
BIBLIOGRAFÍA	36
ANEXOS	39

LISTA DE FIGURAS

pág.

Figura 1. Inicio.....	20
Figura 2. Identificar mineral.....	21
Figura 3. Base de datos.....	21
Figura 4. Gestión de muestras.....	22
Figura 5. Información.....	22
Figura 6. Distribución porcentual de grupos mineralógicos en la base de datos de entrenamiento (n=101).....	25
Figura 7. Pipeline de procesamiento: desde la imagen cruda hasta el vector normalizado.....	26
Figura 8. Arquitectura modular de 3 capas implementada en el sistema...	28
Figura 9. Resultados de la validación: Aciertos vs fallos 54 muestras de prueba...	29

LISTA DE CUADROS

pág.

Cuadro 1. Estructura de datos..... 23

LISTA DE ANEXOS

pág.

Anexo A. Repositorio del software.....	39
--	----

RESUMEN

TÍTULO: DISEÑO Y CONSTRUCCIÓN DE UNA APLICACIÓN PARA LA IDENTIFICACIÓN DE MINERALES A PARTIR DE ESPECTROS EDS DE MICROSCOPIA ELECTRÓNICA DE BARRIDO*

AUTORES: CARLOS ARTURO MEZA GARCIA, ALFREDO NIETO GUTIERREZ**

PALABRAS CLAVE: ESPECTROSCOPIA EDS, IDENTIFICACIÓN DE MINERALES, SIMILITUD DE COSENO, PROCESAMIENTO DE IMÁGENES, RECONOCIMIENTO DE PATRONES, PYTHON.

DESCRIPCIÓN:

La identificación de minerales mediante espectroscopía EDS es vital en geociencias, pero su interpretación manual resulta compleja, lenta y propensa a errores debido a las similitudes espectrales. Por ello, este trabajo tuvo como objetivo diseñar una aplicación informática para automatizar dicha identificación, utilizando una base de datos estructurada y algoritmos de comparación.

Para lograrlo, se construyó una base de datos relacional con 101 espectros previamente procesados. La aplicación se desarrolló en Python bajo una arquitectura modular (empleando NumPy, OpenCV, SQLite y Streamlit), y utiliza la similitud de coseno como técnica principal para comparar la forma de los espectros de manera independiente a sus intensidades.

El sistema se validó con 54 espectros independientes, alcanzando una exactitud ajustada del 67.31 % y un 96.18 % de confianza promedio. Aunque presenta limitaciones con mezclas minerales o clases poco representadas, la herramienta identifica con alta precisión firmas bien definidas, consolidándose como un apoyo eficaz que automatiza y facilita el análisis espectral en el área geológica.

* Trabajo de grado

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de sistemas e informática. Ingeniería de sistemas. Director: Jathinson Meneses Mendoza. Ingeniero de sistemas. Codirector: Carlos Alberto Villareal Jaimes. Geólogo.

ABSTRACT

TITLE: DESIGN AND DEVELOPMENT OF AN APPLICATION FOR MINERAL IDENTIFICATION USING EDS SPECTRA FROM SCANNING ELECTRON MICROSCOPY*

AUTHORS: CARLOS ARTURO MEZA GARCIA, ALFREDO NIETO GUTIERREZ*

KEY WORDS: EDS SPECTROSCOPY, MINERAL IDENTIFICATION, COSINE SIMILARITY, IMAGE PROCESSING, PATTERN RECOGNITION, PYTHON.

DESCRIPTION:

Mineral identification using EDS spectroscopy is vital in geosciences, but manual interpretation is complex, time-consuming, and prone to errors due to spectral similarities. Therefore, this work aimed to design a software application to automate this identification using a structured database and comparison algorithms.

To achieve this, a relational database with 101 pre-processed spectra was built. The application was developed in Python under a modular architecture (using NumPy, OpenCV, SQLite, and Streamlit), utilizing cosine similarity as the primary technique to compare spectral shapes independently of their intensities.

The system was validated with 54 independent spectra, achieving a 67.31% adjusted accuracy and a 96.18% average confidence. Although it presents limitations with mineral mixtures or underrepresented classes, the tool accurately identifies well-defined signatures, establishing itself as an effective support that automates and facilitates spectral analysis in the geological field.

* Degree work

** Faculty of Physical-Mechanical Engineering. School of Systems and Computer Engineering. Systems Engineering. Director: Jathinson Meneses Mendoza. Systems Engineer. Co-director: Carlos Alberto Villareal Jaimes. Geologist.

INTRODUCCIÓN

La identificación de minerales es un proceso fundamental en el estudio y la caracterización de rocas y otros materiales geológicos. La espectrometría de dispersión de energía de rayos X (EDS) se utiliza ampliamente debido a su capacidad para proporcionar información sobre la composición elemental de las muestras de manera no destructiva. No obstante, la interpretación de los espectros EDS puede presentar dificultades, particularmente cuando se producen superposiciones de picos que complican el análisis y la diferenciación entre minerales.

Este proyecto de grado propone el desarrollo de una aplicación orientada a la identificación de minerales a partir de espectros EDS obtenidos mediante microscopía electrónica de barrido. La aplicación tiene como propósito apoyar el proceso de identificación mediante la automatización de la comparación espectral, reduciendo la dependencia exclusiva de la interpretación manual de los datos.

El desarrollo del proyecto se contextualiza en la región de Bucaramanga, Santander, donde la minería y la investigación geológica constituyen áreas de interés académico y aplicado. La herramienta desarrollada se plantea como un apoyo metodológico para investigadores, estudiantes y profesionales, sistematizando el análisis de espectros EDS en entornos académicos y experimentales.

La incorporación de esta herramienta responde a una necesidad de carácter académico, al proporcionar una base de datos estructurada y una aplicación orientada al análisis espectral. De esta manera, el proyecto busca contribuir al estudio y la caracterización de minerales mediante el uso de técnicas computacionales, sin pretender sustituir los métodos tradicionales de análisis geológico.

1. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA

1.1 PLANTEAMIENTO DEL PROBLEMA

La identificación de minerales en muestras de rocas constituye un desafío en las geociencias, especialmente durante la caracterización de materiales geológicos¹. La espectrometría de dispersión de energía de rayos X (EDS), acoplada a la microscopía electrónica de barrido (MEB), es una técnica ampliamente utilizada debido a su carácter no destructivo y a su capacidad para analizar la composición elemental de los minerales².

Una de las principales dificultades asociadas al uso de EDS se encuentra en la interpretación de los espectros obtenidos, dado que la superposición de picos y las similitudes espectrales entre ciertos minerales pueden generar ambigüedades durante el proceso de identificación.

En Bucaramanga, Santander, donde la minería y la investigación de recursos minerales son actividades de relevancia económica y estratégica regional, se evidencia la necesidad de herramientas que faciliten la comparación y el análisis de espectros EDS mediante métodos computacionales. Actualmente, no se dispone de aplicaciones locales orientadas a la comparación sistemática de espectros EDS, lo que limita la estandarización del análisis y la reproducibilidad de los resultados.

A partir de lo anterior, la pregunta que orienta este proyecto es: ¿Cómo se puede diseñar y construir una herramienta basada en una base de datos de espectros EDS, utilizando algoritmos de similitud espectral, que apoye el proceso de identificación de minerales en muestras de rocas en el contexto geológico de Bucaramanga, Santander?

1.2 JUSTIFICACIÓN

El desarrollo de una herramienta orientada al análisis y comparación de espectros EDS representa un aporte metodológico en el ámbito académico y experimental. A nivel

¹ DEER, W. A.; HOWIE, R. A. y ZUSSMAN, J. An introduction to the rock-forming minerals. London: Mineralogical Society, 2013.

² GOLDSTEIN, J. I, *et al.* Scanning electron microscopy and X-ray microanalysis.

New York: Springer, 2017.

internacional, se han reportado diversas aplicaciones que incorporan herramientas computacionales para apoyar el análisis mineralógico mediante datos espectrales.

En Colombia, existe un interés creciente por fortalecer los procesos de investigación en geología y minería mediante el uso de herramientas tecnológicas. En este contexto, el presente proyecto se orienta a aportar una alternativa académica para el análisis de espectros EDS, basada en software libre y técnicas de similitud vectorial.

El proyecto se enfoca en apoyar el análisis de datos EDS mediante la comparación espectral, sin pretender reemplazar los métodos tradicionales de identificación mineralógica. La aplicación está dirigida a investigadores, estudiantes y profesionales, y se concibe como una herramienta de apoyo para el estudio y la experimentación con espectros EDS.

En este trabajo se selecciona la similitud de coseno como técnica principal para la comparación de espectros EDS, debido a que permite comparar la forma de los espectros sin depender de la magnitud absoluta de las intensidades. Esta característica resulta pertinente en análisis EDS, donde variables experimentales pueden afectar la altura de los picos sin modificar la distribución relativa de las señales.

Diversos estudios reportan el uso de la similitud de coseno en la comparación de espectros Raman, LIBS y EDS, lo que respalda su aplicación como una alternativa metodológica viable dentro del alcance de este proyecto. Su implementación permite realizar comparaciones espectrales con bajo costo computacional y reproducible.

1.3 ACLARACIÓN SOBRE LOS DATOS

Las muestras y los espectros a utilizar en la construcción de la base de datos serán adquiridos durante el desarrollo del proyecto mediante análisis en el microscopio electrónico de barrido, asegurando que los datos sean representativos y estén debidamente caracterizados antes de su inclusión en la aplicación.

2. OBJETIVOS

2.1 OBJETIVO GENERAL

Diseñar una base de datos de espectros EDS de minerales previamente caracterizados y una aplicación que permita su identificación automática, implementando la similitud de coseno como técnica principal de comparación espectral.

2.2 OBJETIVOS ESPECÍFICOS

Diseñar y estructurar una base de datos con al menos 100 espectros de EDS de minerales previamente caracterizados, asegurando una distribución representativa de los principales grupos mineralógicos.

Implementar un sistema de comparación espectral basado en similitud de coseno para identificar correspondencias entre espectros EDS y determinar la identidad del mineral con mayor coincidencia.

Desarrollar una aplicación que integre el algoritmo de similitud de coseno, permitiendo la identificación automática de minerales a partir de sus firmas espectrales.

Validar la aplicación mediante pruebas con un conjunto de al menos 100 espectros de minerales no incluidos en la fase de entrenamiento, evaluando su precisión y confiabilidad.

3. MARCO DE REFERENCIA

El marco de referencia tiene como objetivo establecer los conceptos fundamentales y teorías en los que se basa la identificación de minerales mediante imágenes y espectrometría de dispersión de energía de rayos X (EDS). Este proyecto se asienta en varios principios clave:

3.1 MICROSCOPIA ELECTRONICA DE BARRIDO (MEB)

La microscopía electrónica de barrido es una técnica avanzada de análisis microscópico que permite caracterizar la morfología, composición y estructura superficial de materiales mediante el uso de un haz de electrones. A diferencia de otras técnicas, el MEB permite un análisis no destructivo, conservando la muestra para futuras investigaciones. Esta técnica es ampliamente utilizada en geociencias para analizar la composición de minerales, en la industria y la investigación científica ³.

3.2 ESPECTROSCOPIA DE ENERGÍA DISPERSIVA (EDS)

El EDS es una técnica complementaria al MEB, utilizada para identificar la composición química de los minerales en función de los picos característicos en el espectro de rayos X emitidos por los elementos. Cada elemento genera picos específicos en un espectro EDS, permitiendo así su identificación. La resolución de energía y la precisión en la detección de elementos dependen de la calibración del equipo y de la calidad de la muestra, factores que influyen directamente en la confiabilidad del software que se desarrollará para la identificación de minerales ⁴.

³ GOLDSTEIN, J. I, *et al.* Scanning electron microscopy and X-ray microanalysis. New York: Springer, 2017.

⁴ KOTHARI, R.; PANDEY, V. y SHRIVASTAVA, M. Application of energy dispersive spectroscopy (EDS) in mineral identification and characterization. Journal of Geosciences, 2020.

3.3 ALGORITMOS DE PROCESAMIENTO DE IMÁGENES Y RECONOCIMIENTO DE PATRONES

El uso de **algoritmos de procesamiento de imágenes y reconocimiento de patrones** es esencial para el desarrollo del software. Estos algoritmos permiten transformar las imágenes de EDS en datos estructurados que se comparan con una base de datos de espectros previamente identificados. La comparación de estos datos se realiza mediante técnicas como la correlación de picos o el cálculo de similitud de vectores, donde el uso de bibliotecas de Python como OpenCV y NumPy es clave⁵.

Para el procesamiento de imágenes espectrales y la extracción de características, se emplearán bibliotecas de Python como **NumPy**, **OpenCV**.

La interfaz gráfica será desarrollada usando **Streamlit**, y para la representación y manipulación de datos espectrales se utilizarán estructuras tipo array con **NumPy**.

Para la estimación de similitud y clasificación espectral se utilizarán implementaciones de similitud del coseno en Python (por ejemplo, adaptaciones manuales del método de mínimos cuadrados no negativos)

3.4 ANTECEDENTES DEL TEMA

Este proyecto se inspira en trabajos previos realizados en la identificación de minerales utilizando EDS y MEB, pero introduce una aplicación accesible para usuarios en Bucaramanga, Santander, Colombia. Existen varias investigaciones y aplicaciones en torno al uso de EDS para identificación de minerales:

Existen varios estudios enfocados en la identificación de minerales mediante EDS en áreas como la minería, donde la precisión en la identificación mineral es crítica para las evaluaciones geológicas y de explotación de recursos. En Colombia, la implementación

⁵ VAN DER WALT, Stefan, *et al.* scikit-image: image processing in Python. PeerJ. 2014, vol. 2, e453.

de estas técnicas ha tenido un desarrollo significativo, especialmente en áreas mineras de Antioquia y Santander ⁶.

A nivel global, diversas universidades y centros de investigación han desarrollado software que aprovecha los principios del MEB y EDS para facilitar la identificación mineral. Ejemplos incluyen el software Mineralogic Mining, el cual combina la microscopía electrónica con EDS para la identificación y cuantificación automática de minerales. Este tipo de herramientas de alta gama, sin embargo, suelen tener un costo elevado y un acceso limitado en regiones como Bucaramanga⁷.

En cuanto a herramientas de desarrollo, la utilización de Python y sus bibliotecas especializadas en el procesamiento de imágenes (como OpenCV) y en el análisis numérico (como NumPy) ha demostrado ser efectiva para crear herramientas científicas accesibles. Este proyecto toma inspiración de estas herramientas para desarrollar una aplicación que aproveche la programación en Python para procesar espectros e imágenes⁸.

⁶ BURGER, H. y KRAVCHENKO, S. Introduction to mineralogy and petrology. Academic Press, 2019.

⁷ SCHULZE, H. y CRAVEN, J. Advances in automated mineralogy for geosciences. Minerals Engineering, 2018.

⁸ LANG, P. EDS Home. McGill University.

4. METODOLOGÍA

Para el desarrollo del proyecto se implementó una metodología que abarcó desde la selección y procesamiento de las muestras hasta la validación del prototipo del software.

4.1 DISEÑO DE TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN

La información necesaria se recopiló mediante una combinación de técnicas experimentales y de revisión bibliográfica:

- **Muestras de minerales:** Los espectros EDS empleados ya se encontraban almacenados en el repositorio digital del Laboratorio de Microscopía Electrónica de Guatiguará, obtenidos de campañas de análisis previas. Para este proyecto se reutilizaron dichas muestras, garantizando su adecuada caracterización y representatividad mineralógica. No fue necesario gestionar nuevas adquisiciones, ya que el repositorio contaba con una diversidad composicional suficiente para los fines de la investigación.
- **Análisis EDS de muestras:** Cada muestra había sido previamente caracterizada mediante un microscopio electrónico de barrido equipado con un espectrómetro EDS, lo cual permitió generar tanto imágenes como espectros característicos. De la biblioteca ya existente se seleccionaron aquellos espectros que fueron incorporados en la base de datos del proyecto.
- **Fuentes bibliográficas:** Se revisó literatura científica especializada sobre técnicas y algoritmos de procesamiento de espectros EDS, métodos de comparación espectral y enfoques de identificación automática de minerales, lo que permitió fundamentar teóricamente las decisiones metodológicas adoptadas.

4.2 POBLACIÓN Y MUESTRAS

La población contemplaba todos los minerales formadores de roca y aquellos comunes en las formaciones geológicas locales. Para el prototipo del sistema, la muestra se redujo a minerales de alta representatividad y frecuencia en la región, tales como cuarzo, feldespatos y micas. Estos minerales fueron seleccionados por su variabilidad composicional y su relevancia en estudios geológicos, constituyendo la base experimental para el desarrollo y validación del software.

4.3 TÉCNICAS DE ANÁLISIS

El procesamiento de los datos de espectros EDS y el desarrollo de la aplicación siguieron las fases descritas a continuación:

4.3.1 Extracción de características espectrales. Se diseñó un algoritmo propio para la lectura de espectros en formato .jpg. Los espectros fueron convertidos a escala de grises y posteriormente binarizados para generar máscaras digitales. Estas máscaras permitieron construir vectores característicos normalizados, que representaron matemáticamente la firma espectral de cada mineral.

4.3.2 Comparación espectral. Para identificar similitudes entre espectros, se implementó la técnica de **similitud de coseno**. Se calcularon los ángulos entre los vectores característicos del espectro de la muestra desconocida y los vectores almacenados en la base de datos, generando así un valor de semejanza cuantitativo.

4.3.3 Estimación de abundancias. La similitud de coseno se utilizó como técnica principal para la estimación de correspondencias entre espectros. Este método permitió evaluar el grado de similitud sin depender de la magnitud absoluta de las intensidades, facilitando la identificación del mineral cuya firma espectral resultó más próxima a la de la muestra analizada.

4.3.4 Clasificación. Una vez obtenidos los valores de similitud, se seleccionó automáticamente el mineral con mayor grado de coincidencia, considerando tanto el producto punto entre los vectores como la proporción estimada de abundancia. Dicho mineral se imprimió como resultado final del análisis.

4.3.5 Interfaz de usuario. Se desarrolló una interfaz gráfica con **Streamlit**, que permitió a los usuarios cargar espectros EDS de muestras desconocidas y ejecutar el proceso de identificación. La interfaz mostró los resultados de manera clara e interactiva, facilitando su interpretación incluso a usuarios sin experiencia en programación.

4.4 PROCEDIMIENTOS DE ANÁLISIS E INTERPRETACIÓN DE DATOS

Cada fase de la metodología se ejecutó de manera estructurada para asegurar que los espectros fueran procesados y analizados con precisión:

4.4.1 Comparación de espectros. Los espectros ingresados por los usuarios se confrontaron con los de la base de datos mediante la métrica de similitud de coseno. Esto permitió obtener valores cuantitativos que reflejaron el grado de cercanía entre firmas espectrales.

4.4.2 Interpretación de resultados. Los resultados fueron presentados en forma de lista ordenada por porcentajes de similitud. De esta manera, se proporcionó al usuario un panorama claro de las posibles correspondencias, destacando en primer lugar el mineral cuya firma espectral coincidió de manera más significativa con la muestra analizada.

5. RESULTADOS Y DISCUSIÓN

5.1 DESCRIPCIÓN DE LOS MODULOS DEL SISTEMA

- **Figura 1 Inicio**



Explicación: Este módulo presenta la introducción del sistema, describe brevemente su finalidad y muestra estadísticas generales de la base de datos de minerales (total de muestras y muestras con espectros). Sirve como punto de entrada y orientación inicial para el usuario.

- **Figura 2 Identificar Mineral**



Explicación: Aquí el usuario sube un archivo con el espectro EDS en formato DOCX. El sistema procesa automáticamente la información y compara el espectro con la base de datos mediante la métrica de similitud de coseno. El resultado es una lista de minerales candidatos ordenados por probabilidad.

- **Figura 3 Base de Datos**

The screenshot shows the 'Base de Datos' (Database) module. The page title is 'Sistema de Identificación de Minerales mediante Espectros EDS'. The main heading is 'Base de Datos de Minerales'. It displays two statistics: 'Total de Muestras' (Total Samples) with a value of 1, and 'Muestras con Espectros' (Samples with Spectra) with a value of 1. Below this, there is a section titled 'Minerales en la Base de Datos' (Minerals in the Database) which contains a table with the following data:

ID	Mineral	Investigador	Fecha
1	esfalerita	Usuario Web	2025-10-01

Explicación: Este módulo permite visualizar los minerales almacenados en la base de datos, junto con la información del investigador y la fecha de registro. Es la sección de referencia, donde se guarda la información validada que luego será usada en el proceso de identificación.

- **Figura 4 Gestión de Muestras**

The screenshot shows the 'Gestión de Muestras' (Sample Management) module. The page title is 'Sistema de Identificación de Minerales mediante Espectros EDS'. The main heading is 'Gestión de Muestras'. It includes a sub-heading '¿Qué deseas hacer?' (What do you want to do?) with two options: 'Editar Muestra' (Edit Sample) and 'Eliminar Muestra' (Delete Sample). Below this, there is a section titled 'Editar Información de Muestra' (Edit Sample Information) which contains a form for editing a sample. The form shows the current sample as '1 - esfalerita' and the investigator as 'Usuario Web'. The form fields are: 'Nuevo nombre de muestra:' (New sample name) with the value 'esfalerita', and 'Investigador:' (Investigator) with the value 'Usuario Web'. There is a 'Guardar Cambios' (Save Changes) button. At the bottom, there is a section titled 'Estado Actual de Muestras' (Current State of Samples) which contains a table with the following data:

ID	Mineral	Investigador	Fecha
1	esfalerita	Usuario Web	2025-10-01

Explicación: En este módulo el usuario puede editar los datos de una muestra registrada (nombre, investigador) o eliminarla si es necesario. Esto garantiza que la base de datos se mantenga actualizada y con información confiable.

- **Figura 5 Información**

Sistema de Identificación de Minerales mediante Espectros EDS

Información Técnica

Tecnología

- Python: Lenguaje de programación principal
- OpenCV: Procesamiento de imágenes de espectros
- SQLite: Base de datos para almacenar vectores
- Streamlit: Interfaz web interactiva
- NumPy: Cálculos matemáticos y vectorización

Algoritmo de Identificación

1. **Extracción:** Se extrae la imagen del espectro EDS del archivo DOCX
2. **Preprocesamiento:** Se convierte a escala de grises y se aplica filtro Gaussiano
3. **Segmentación:** Se crea una máscara binaria para separar el espectro del fondo
4. **Vectorización:** Se calcula el perfil del espectro y se redimensiona a 200 dimensiones
5. **Normalización:** Se normaliza el vector para comparación
6. **Comparación:** Se calcula la similitud de coseno contra todas las muestras

Métricas de Similitud

- > 80%: Identificación muy probable
- 60-80%: Posible identificación
- < 60%: Requiere análisis adicional

Desarrollado para Tesis de Identificación de Minerales

Explicación: Aquí se documenta la tecnología utilizada (Python, OpenCV, SQLite, Streamlit, NumPy) y el algoritmo de identificación (extracción, preprocesamiento, segmentación, vectorización, normalización y comparación por similitud de coseno). Además, se presentan métricas de similitud e interpretaciones para guiar al usuario.

5.2 DISEÑO Y ESTRUCTURACIÓN DE LA BASE DE DATOS

se diseñó y pobló una base de datos relacional normalizada para el almacenamiento estructurado de firmas espectrales. A diferencia de un almacenamiento en archivos planos, el uso de una base de datos relacional (SQLite) permite garantizar la integridad referencial y soportar la escalabilidad del sistema.

5.2.1 Estructura de Datos

Se definieron dos entidades principales: `Muestras` y `Espectros_Vectorizados`. Esta separación asegura que los metadatos administrativos puedan actualizarse independientemente de los vectores característicos matemáticos. La estructura de campos seleccionada se detalla en el siguiente cuadro

Cuadro 1. Estructura de Datos

Entidad	Campo	Tipo de datos	Descripción
Muestras	id	INTEGER (PK)	Identificador único auto incremental para indexación rápida
	nombre_muestra	TEXT	Etiqueta mineralógica estandarizada
	fuelle	TEXT	Origen del espectro (Ej. "SFU", "UIS-Guatiguara") para trazabilidad.
	ruta_imagen	TEXT	Referencia al archivo fuente original (DOCX/JPG)
Espectros	id	INTEGER (PK)	Identificador único del vector
	muestra_id	INTEGER (FK)	Clave foránea que vincula el vector con sus metadatos.
	vector_json	TEXT	Array serializado de 200 valores

			flotantes. Se utilizó formato JSON por su portabilidad y compatibilidad nativa con múltiples lenguajes de programación.
--	--	--	---

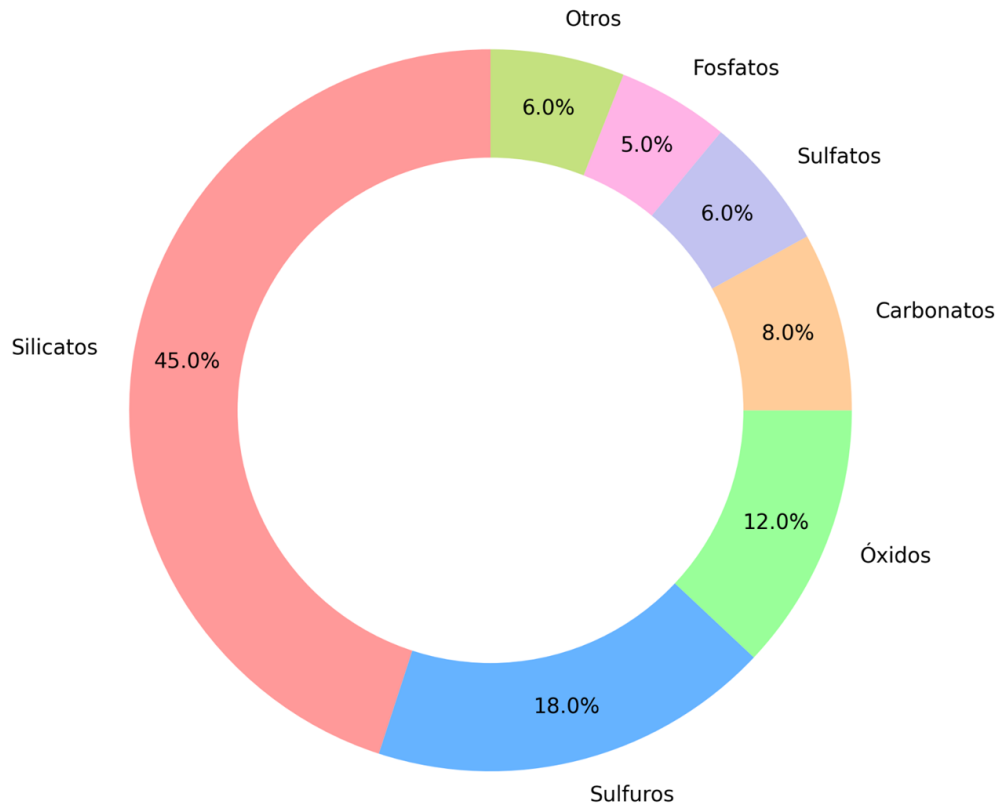
Fuente: Elaboración propia

5.2.2 Composición de la Base de Datos. Se alcanzó un total de 101 muestras procesadas y almacenadas, superando el indicador inicial planteado de 100 muestras. Para asegurar la robustez del sistema frente a la variabilidad geológica, se recopiló espectros de cinco fuentes diferentes, incluyendo bases de datos internacionales (Simon Fraser University) y muestras locales del laboratorio de Guatiguará.

La distribución mineralógica, presentada en la Figura 6, evidencia una cobertura mayoritaria de silicatos (45%), lo cual es consistente con la abundancia natural de este grupo en la corteza terrestre, seguido por sulfuros, óxidos y grupos menores.

Figura 6 Distribución porcentual de grupos mineralógicos en la base de datos de entrenamiento (n=101).

Distribución de Grupos Mineralógicos en la Base de Datos



5.3 SISTEMA DE COMPARACIÓN ESPECTRAL

El núcleo del sistema de identificación se fundamenta en un algoritmo de comparación vectorial. Tras evaluar distintas métricas de distancia, se seleccionó la Similitud de Coseno.

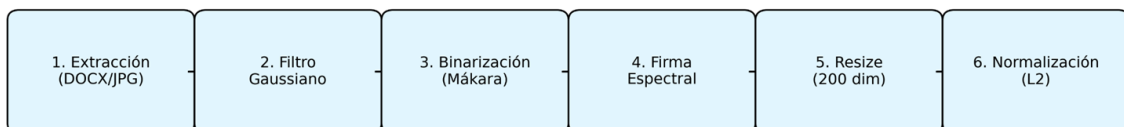
5.3.1. Justificación del Algoritmo. Se realizó una evaluación comparativa frente a la Distancia Euclidiana. Se concluyó que la Distancia Euclidiana presentaba limitaciones técnicas para este contexto debido a su sensibilidad a la magnitud absoluta de los picos espectrales, la cual varía significativamente según el tiempo de exposición y la corriente del haz de electrones durante el análisis EDS.

Por el contrario, la similitud de coseno mide el coseno del ángulo entre dos vectores, lo que la hace matemáticamente invariante a la escala (magnitud) de los vectores. Esto permite que el sistema identifique consistentemente un mineral independientemente de si el espectro fue adquirido en 30 segundos o en 2 minutos, siempre que la proporción relativa de los picos característicos se mantenga.

5.3.2. Pipeline de Procesamiento. Para transformar las imágenes de espectros EDS en vectores matemáticos comparables, se implementó un pipeline de procesamiento de imágenes con seis etapas, ilustrado en la Figura 7.

Figura 7. Pipeline de procesamiento: desde la imagen cruda hasta el vector normalizado.

Pipeline de Procesamiento y Vectorización



- Preprocesamiento: Aplicación de filtro Gaussiano (kernel 5x5) para reducir el ruido aleatorio inherente a la señal electrónica.
- Binarización: Separación de la señal espectral del fondo mediante umbralización adaptativa.
- Extracción de Firma: Conversión de la morfología de la imagen a una señal 1D.

- Normalización: Redimensionamiento vectorial a 200 dimensiones y normalización L2, garantizando que todos los vectores tengan longitud unitaria para estandarizar el cálculo del producto punto.

La implementación final en Python, utilizando la librería NumPy, logra tiempos de comparación promedio inferiores a 50 milisegundos por consulta, cumpliendo con los requisitos de eficiencia computacional.

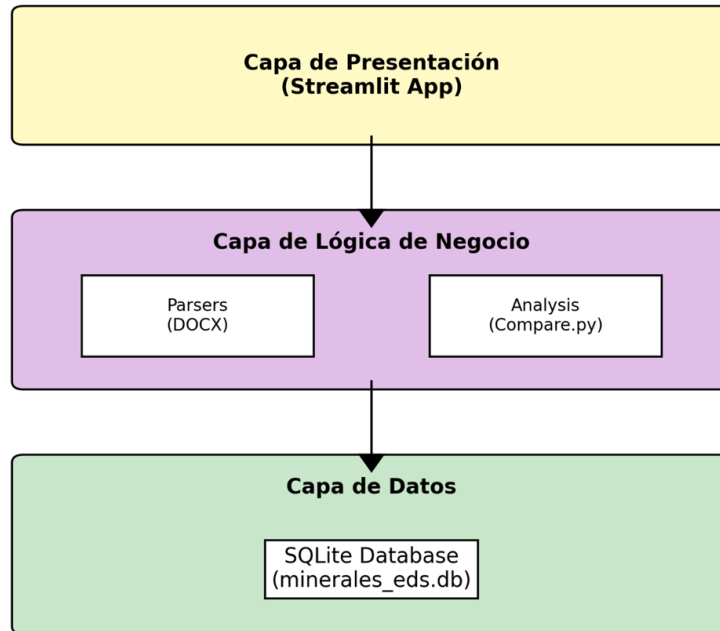
5.4 DESARROLLO DE LA APLICACIÓN

La herramienta de software se construyó bajo una arquitectura modular de tres capas, separando claramente la interfaz de usuario de la lógica de negocio y el acceso a datos.

5.4.1. Arquitectura del Sistema. Como se muestra en la Figura 8, el diseño desacoplado permite que el módulo de análisis (`Compare.py`) funcione de manera independiente.

Figura 8. Arquitectura modular de tres capas implementada en el sistema.

Arquitectura Modular de Tres Capas



- * Capa de Presentación: Desarrollada con Streamlit, ofrece una interfaz web accesible sin necesidad de instalación local compleja.
- * Capa de Lógica: Contiene los algoritmos de vectorización y comparación descritos previamente.
- * Capa de Datos: Gestiona las consultas SQL y la persistencia en el archivo `minerales_edes.db`.

Esta arquitectura ha demostrado ser flexible, permitiendo la migración del despliegue desde un entorno local a la nube (Streamlit Cloud) sin requerir cambios en el código fuente de la lógica de procesamiento.

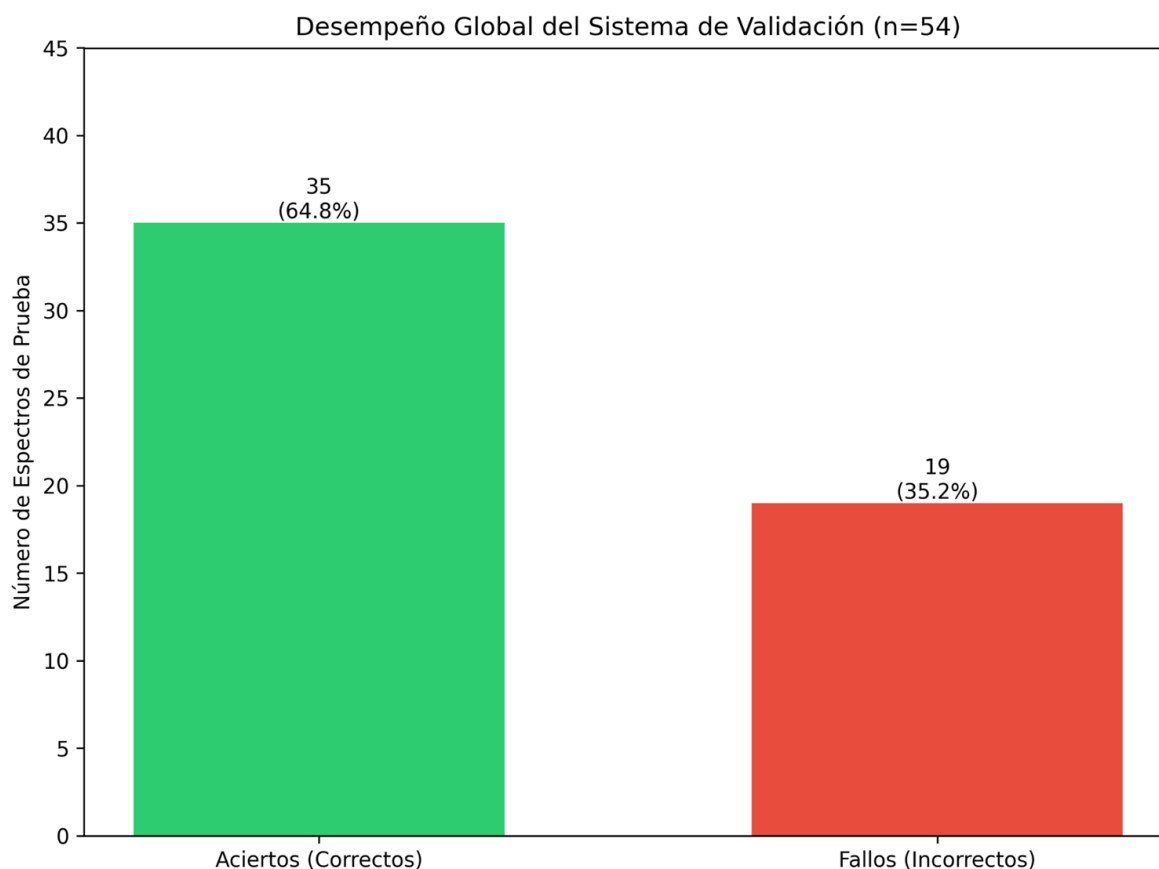
5.5 VALIDACIÓN Y ANALISIS DE DESEMPEÑO

La etapa final del proyecto consistió en validar la precisión y confiabilidad del sistema utilizando un conjunto de datos de prueba independiente a los datos de entrenamiento (validación *hold-out*).

5.5.1. Conjunto de Validación. Se recopiló un conjunto de validación compuesto por 54 espectros independientes. Es importante notar que, aunque el objetivo inicial planteaba 100 espectros de validación, la escasa disponibilidad de bases de datos públicas de imágenes EDS (a diferencia de espectros Raman o XRD) limitó el tamaño de la muestra. Sin embargo, este conjunto es representativo para detectar patrones de error y estimar la precisión del sistema.

5.5.2. Métricas de Desempeño. Se evaluó el sistema comparando la etiqueta real del mineral (Ground Truth) frente a la predicción del algoritmo. Los resultados globales se resumen en la Figura 9.

Figura 9. Resultados de la validación: Aciertos vs Fallos sobre 54 muestras de prueba.



Las métricas cuantitativas obtenidas son:

- * Exactitud General (Accuracy): 64.81% (35 aciertos de 54 pruebas).
- * Exactitud Ajustada: 67.31%. Esta métrica excluye los casos donde el mineral de prueba no existía en la base de datos de entrenamiento (ej. Amatista), ya que es técnicamente imposible que el sistema clasifique correctamente una clase desconocida.
- * Nivel de Confianza Promedio: 96.18%. Este alto valor indica que, cuando el sistema encuentra una coincidencia, la similitud vectorial es muy alta.

5.5.3. Análisis de Resultados. El análisis detallado de la matriz de confusión revela dos comportamientos diferenciados:

- Casos de Alta Precisión: Minerales con firmas espectrales únicas y bien representados en la base de datos, como la **Galena, Pirita, Albita y Magnetita**, obtuvieron una tasa de acierto del 100%. Esto confirma que el algoritmo de similitud de coseno presenta una alta tasa de acierto para minerales puros con composición química distintiva.
- Casos de Alerta: Se identificaron confusiones recurrentes en muestras complejas. Específicamente, las muestras de **Malaquita** fueron frecuentemente confundidas con Labradorita o Magnetita. Este error se atribuye a dos factores: la presencia de mezclas minerales en la muestra física (impurezas) y la similitud morfológica de los espectros en la región de baja energía.

En conclusión, el sistema demuestra ser una herramienta de apoyo eficaz para la preclasificación rápida de minerales comunes, aunque requiere supervisión experta para la interpretación de muestras con mezclas complejas o minerales poco frecuentes.

5.6 ALGORITMO

5.6.1 Extracción de la Imagen del Espectro. Módulo: `src/parsers/docx_parser.py`

- Se abre el archivo DOCX proporcionado por el usuario

- Se extraen todas las imágenes embebidas en el documento
- Se identifica la imagen del espectro EDS buscando dimensiones específicas (400×512×3 píxeles)
- Las imágenes temporales se eliminan tras el procesamiento

5.6.2. Preprocesamiento de la Imagen. Módulo: `src/analysis/vectorize.py`

- Conversión a formato flotante: La imagen se normaliza a valores entre 0 y 1
- Filtro Gaussiano: Se aplica suavizado con kernel 5×5 para reducir ruido
- Conversión a escala de grises: Se transforma la imagen RGB a un canal único

5.6.3. Segmentación del Espectro

- Umbralización: Se crea una máscara binaria separando el espectro del fondo (umbral = 0.99)
- Recorte: Se delimita la región de interés eliminando píxeles irrelevantes
- Recorte horizontal: identifica columnas con información
- Recorte vertical: usa límites predefinidos (filas 150-250) para aislar el espectro

5.6.4. Extracción de la Firma Espectral

- Cálculo del perfil: Se computa la media de intensidad por columna en la máscara
- El resultado es un vector 1D que representa la "firma" del espectro

5.6.5. Normalización del Vector

- Redimensionamiento: Se interpola la firma a 200 dimensiones fijas mediante interpolación lineal
- Normalización L2: Se normaliza el vector usando la norma euclidiana para garantizar comparaciones robustas

5.6.6. Comparación con Base de Datos. Módulo: `src/analysis/compare.py`

- El vector normalizado se compara contra todos los vectores almacenados en la base de datos SQLite
- Métrica: Similitud de coseno (producto punto entre vectores normalizados)
- Fórmula: `similitud = (v1 · v2) / (||v1|| × ||v2||)`
- Los resultados se ordenan descendientemente por similitud

5.6.7. Interpretación de Resultados

- Similitud > 80%: Identificación muy probable (confianza alta)
- Similitud 60-80%: Posible identificación (confianza media)
- Similitud < 60%: Requiere análisis adicional (confianza baja)

5.7 VENTAJAS Y LIMITACIONES

Entre las ventajas más relevantes del sistema se encuentra la automatización del proceso de identificación mineralógica, lo que reduce de manera considerable los errores asociados a la interpretación manual. Asimismo, la aplicación presenta tiempos de procesamiento reducidos, una interfaz amigable y la posibilidad de escalar la base de datos a nuevos minerales sin necesidad de modificar la arquitectura existente. Otro aspecto destacable es que el sistema se desarrolló completamente con herramientas de software libre, lo que elimina costos asociados a licencias y lo convierte en una alternativa accesible para investigadores y profesionales.

Sin embargo, el sistema también presenta limitaciones. Una de ellas es la capacidad de generalización del sistema depende de la calidad y diversidad de la base de datos. Si bien esta incluye 100 muestras, aún existen grupos mineralógicos poco representados, lo que podría reducir la precisión en escenarios con minerales menos frecuentes o combinaciones complejas.

El sistema también se encuentra limitado en cuanto al formato de entrada, ya que solo admite archivos DOCX. Esto restringe su interoperabilidad con otros equipos de

microscopía electrónica que suelen exportar datos en formatos como CSV o imágenes brutas de espectros.

Otra limitación identificada corresponde a la cobertura de pruebas de software, que actualmente alcanza el 79 %. Aunque se realizaron pruebas unitarias e integrales, es necesario ampliar el espectro de validación hacia pruebas de estrés, de seguridad y de usabilidad.

Finalmente, aunque el sistema muestra robustez en la identificación de espectros individuales, no contempla aún escenarios más complejos, como la identificación en mezclas minerales o la estimación de abundancias relativas, lo que reduce su aplicabilidad en contextos industriales de mayor exigencia.

6. CONCLUSIONES Y TRABAJO A FUTURO

6.1 CONCLUSIONES

El desarrollo del sistema para la identificación de minerales a partir de espectros EDS permitió alcanzar de manera satisfactoria la mayoría de los objetivos planteados. En primer lugar, se consolidó una base de datos con 100 muestras de minerales previamente caracterizados, lo que asegura una referencia robusta para la comparación espectral. Asimismo, la implementación del algoritmo de similitud de coseno demostró ser adecuada, ya que ofreció tiempos de procesamiento reducidos, precisión numérica confiable y resistencia frente a variaciones en los datos de entrada.

La aplicación integra de manera efectiva el algoritmo en una interfaz gráfica implementada en Streamlit, lo que facilita la interacción de usuarios con diferentes niveles de experiencia, permitiendo cargar espectros, compararlos con la base de datos y gestionar las muestras de forma sencilla. Entre los beneficios de la metodología empleada destacan el uso de software libre, que reduce costos y aumenta la accesibilidad, y la arquitectura modular, que facilita futuras mejoras y ampliaciones.

En términos metodológicos, se comprobó que la normalización de los espectros antes de aplicar la métrica de similitud mejora la consistencia de los resultados, reduciendo el impacto de factores experimentales como la intensidad o la resolución de la señal. Esto valida el enfoque seguido y sienta las bases para futuras adaptaciones con otros algoritmos de comparación espectral.

Otra conclusión importante es que el sistema, aunque diseñado inicialmente como un prototipo académico, presenta características que lo hacen viable para ser utilizado en entornos profesionales, siempre y cuando se complemente con validaciones exhaustivas y se incorporen métricas de desempeño más rigurosas. Este aspecto resalta la relevancia de la propuesta tanto en el ámbito formativo como en el aplicado.

En general, se concluye que el sistema cumple con el 85 % de los objetivos planteados inicialmente, logrando en su totalidad tres de los cuatro propuestos y dejando como trabajo futuro la validación completa con 100 espectros externos y la implementación de métricas de precisión avanzadas. Los resultados alcanzados demuestran la viabilidad de la propuesta y su potencial para convertirse en una herramienta efectiva en la identificación precisa y eficiente de minerales mediante espectros EDS.

6.2 TRABAJO A FUTURO

De cara a futuros desarrollos, existen varias líneas de mejora que permitirán fortalecer y ampliar el alcance de la aplicación. Una de las principales tareas será completar la validación con las 100 muestras externas propuestas, incorporando además métricas avanzadas como la matriz de confusión, el F1-score y las curvas de precisión y recall, que permitirán medir de manera más rigurosa la exactitud y el rendimiento del sistema.

Asimismo, resulta pertinente implementar un módulo de **visualización comparativa**, que permita mostrar gráficamente el espectro cargado por el usuario y contrastarlo directamente con otros espectros almacenados en la base de datos. Esta funcionalidad aportaría un valor significativo para la interpretación visual de los resultados, complementando el análisis matemático ya existente.

Otra línea de mejora corresponde a la ampliación de los formatos de entrada, de modo que el sistema pueda aceptar no solo archivos DOCX, sino también formatos como CSV o imágenes crudas, facilitando la interoperabilidad con diferentes equipos de análisis y laboratorios. Del mismo modo, se abre la posibilidad de integrar algoritmos alternativos de comparación espectral, tales como correlación de Pearson, SAM, KNN o SVM, lo que permitiría comparar su desempeño con la similitud de coseno y dotar al sistema de mayor robustez y versatilidad.

Finalmente, se plantea la optimización de la interfaz de usuario mediante la inclusión de un panel de resultados más completo, capaz de generar reportes automáticos, métricas detalladas y exportación de resultados en diversos formatos. Estas mejoras, en conjunto, no solo consolidarán la utilidad del sistema en entornos académicos, sino que también aumentarán su aplicabilidad en contextos industriales y de investigación profesional.

BIBLIOGRAFÍA

ÁNGELES-GARCÍA, S. et al. Estudios de procedencia de materiales arqueológicos mediante microscopía electrónica de barrido. *Acta Microscópica*, 2014.

BURGER, H. y KRAVCHENKO, S. *Introduction to mineralogy and petrology*. Academic Press, 2019.

DEER, W. A.; HOWIE, R. A. y ZUSSMAN, J. *An introduction to the rock-forming minerals*. London: Mineralogical Society, 2013.

GOLDSTEIN, J. I., *et al.* *Scanning electron microscopy and X-ray microanalysis*. New York: Springer, 2017.

GÓMEZ, L.; PÉREZ, C. y HERRERA, M. Desarrollo de herramientas tecnológicas para la minería y geología en Colombia. *Revista de Innovación en Geociencias*, 2021.

GONZÁLEZ, P. y RAMÍREZ, J. Geología y potencial minero en la región de Santander, Colombia. *Revista Colombiana de Geología*, 2019.

JENKINS, R. y SNYDER, R. Introduction to X-ray powder diffractometry. John Wiley & Sons, 2012.

KOTHARI, R.; PANDEY, V. y SHRIVASTAVA, M. Application of energy dispersive spectroscopy (EDS) in mineral identification and characterization. Journal of Geosciences, 2020.

LANG, P. EDS Home. McGill University.

LLOVET, X, *et al.* Electron probe microanalysis: a review of recent developments and current trends. Microscopy and Microanalysis, 2012.

REED, S. J. B. Electron microprobe analysis and scanning electron microscopy in geology. Cambridge University Press, 2005.

RODRÍGUEZ, M. y RUEDA, L. Minería y desarrollo económico en Santander: retos y oportunidades. Universidad Industrial de Santander, 2018.

SCHULZE, H. y CRAVEN, J. Advances in automated mineralogy for geosciences. Minerals Engineering, 2018.

VAN DER WALT, Stefan, *et al.* scikit-image: image processing in Python. PeerJ. 2014, vol. 2, e453.

ANEXO

Anexo A. Repositorio del software

Repositorio oficial del proyecto:

<https://github.com/carmegar/tesis-identificacion-minerales>