

# Aplicación de modelos de aprendizaje automático en la detección de asbesto cemento en imágenes multiespectrales

## Application of machine learning models for asbestos cement detection in multispectral images

Gabriel E. Chanchí-Golondrino<sup>1</sup> Manuel A. Ospina-Alarcón<sup>1</sup> Manuel Saba<sup>1</sup>

<sup>1</sup> Universidad de Cartagena, Facultad de Ingeniería, Cartagena de Indias, Colombia

### Resumen

**Introducción:** La detección de asbesto cemento ha sido predominantemente realizada utilizando datos hiperespectrales. El desafío radica en mejorar la eficiencia de la detección sin comprometer la precisión, especialmente en contextos donde solo están disponibles imágenes multiespectrales.

**Objetivos:** El objetivo principal de este estudio es aplicar modelos de aprendizaje automático para la detección de asbesto cemento en imágenes multiespectrales.

**Metodología:** La investigación se llevó a cabo utilizando una adaptación de cuatro fases de la metodología CRISP-DM, que incluyó los siguientes pasos: Creación del dataset: Se recopiló y preparó un conjunto de datos a partir de una imagen multiespectral de referencia de la ciudad de Cartagena. Evaluación de las bandas relevantes: Se seleccionaron las bandas espectrales más relevantes para la detección de asbesto cemento. Entrenamiento y evaluación de modelos: Se entrenaron y evaluaron cuatro modelos de aprendizaje automático: kNN, árboles de decisión, máquinas de soporte vectorial (SVM) y regresión logística. Despliegue del mejor modelo: Finalmente, el modelo con mejor rendimiento fue desplegado sobre la imagen de referencia para evaluar su desempeño en un escenario real.

**Resultados:** de los cuatro modelos evaluados, el modelo de árboles de decisión demostró la mayor eficiencia y consistencia en los conjuntos de entrenamiento y prueba, logrando un valor de 0.93 tanto en las métricas de Precisión como de Recall. En contraste, el modelo de regresión logística fue el que obtuvo el peor rendimiento, con valores de 0.829 y 0.875 en Precisión y Recall, respectivamente.

**Conclusiones:** el modelo propuesto de árboles de decisión demostró ser una herramienta efectiva y consistente para la detección de asbesto cemento en imágenes multiespectrales, con un buen balance entre precisión y recall. Este modelo ofrece una solución viable para la detección automática de materiales en contextos donde solo se disponen de imágenes multiespectrales y puede servir como referencia tanto en el ámbito académico como industrial para su integración en sistemas de detección de materiales.

**Palabras clave:** Detección de asbesto, aprendizaje automático, imágenes multiespectrales, sensado remoto.

### Abstract

**Introduction:** the detection of asbestos cement has predominantly been carried out using hyperspectral data. The challenge lies in improving detection efficiency without compromising accuracy, especially in contexts where only multispectral images are available.

**Objectives:** the main objective of this study is to apply machine learning models for the detection of asbestos cement in multispectral images.

**Methodology:** the research was conducted using a four-phase adaptation of the CRISP-DM methodology, which included the following steps: Dataset creation: A dataset was compiled and prepared from a reference multispectral image of the city of Cartagena. Evaluation of relevant bands: The most relevant spectral bands for detecting asbestos cement were selected. Model training and evaluation: Four machine learning models were trained and evaluated: kNN, decision trees, support vector machines (SVM), and logistic regression. Deployment of the best model: Finally, the best-performing model was deployed on the reference image to assess its performance in a real-world scenario.

**Results:** of the four models evaluated, the decision tree model demonstrated the highest efficiency and consistency on both the training and testing sets, achieving a value of 0.93 for both Precision and Recall metrics. In contrast, the logistic regression model performed the worst, with values of 0.829 and 0.875 for Precision and Recall, respectively.

**Conclusions:** the proposed decision tree model proved to be an effective and consistent tool for detecting asbestos cement in multispectral images, with a good balance between precision and recall. This model provides a viable solution for automatic material detection in contexts where only multispectral images are available and can serve as a reference in both academic and industrial settings for integration into material detection systems.

**Keywords:** Asbestos detection, machine learning, multispectral imaging, remote sensing.

### ¿Cómo citar?

Chanchí-Golondrino, G.E., Ospina-Alarcón, M.A., Saba, M. Aplicación de modelos de aprendizaje automático en la detección de asbesto cemento en imágenes multiespectrales. Ingeniería y Competitividad, 2025, 27(1)e-21014672

<https://doi.org/10.25100/iyc.v27i1.14672>

Recibido: 30/12/24

Revisado: 10/03/24

Aceptado: 14/03/25

Online: 19/03/25

**Correspondencia**

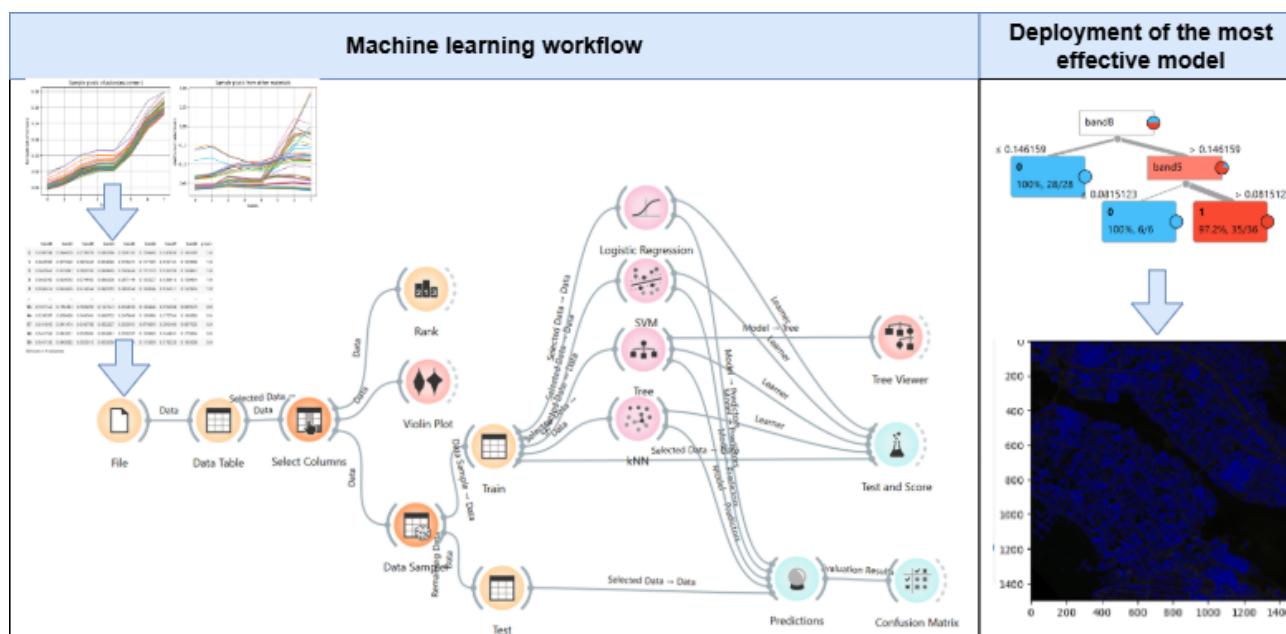
gchanchig@unicartagena.edu.co



## Contribución a la literatura

Considerando la legislación colombiana que prohíbe el uso de asbesto-cemento debido a sus implicaciones para la salud pública, resulta relevante realizar investigaciones centradas en la detección automatizada de asbesto-cemento, aprovechando las ventajas que ofrecen las técnicas de teledetección y visión artificial. En este sentido, dado el gran volumen de información contenida en las imágenes hiperespectrales y la demanda de recursos computacionales asociada, es necesario identificar métodos computacionales que permitan la detección de asbesto-cemento en escenarios de menor dimensión, como las imágenes multiespectrales.

En esta investigación, se implementó un flujo de trabajo de aprendizaje automático sobre un conjunto de datos de firmas espe-ctrales de asbesto-cemento y otros materiales, extraídos de una imagen multiespectral de la ciudad de Cartagena de Indias. El flujo de trabajo se implementó utilizando la herramienta de programación visual Orange, lo que permitió la evaluación de cuatro modelos de aprendizaje automático (kNN, árboles de decisión, máquinas de vectores de soporte y regresión logística). Los resultados indicaron que el modelo de árbol de decisión logró el mejor y más consistente rendimiento, con una puntuación de precisión y recuperación de 0,93. Además, se determinó que las bandas más relevantes para la identificación del asbesto son las bandas 8 y 5, lo que demuestra que la presencia del material puede detectarse utilizando únicamente estas dos bandas de la imagen multiespectral. Estos hallazgos contribuyen significativamente a la eficiencia computacional en el procesamiento de imágenes multiespectrales, ya que no es necesario procesar las ocho bandas de reflectancia, lo que mejora los sistemas de monitoreo ambiental basados en teledetección.



## Introducción

La teledetección se refiere a la ciencia dedicada a adquirir y procesar información sobre la superficie de la Tierra mediante el uso de sensores montados en drones, aviones y satélites. Al analizar la interacción de la energía electromagnética con los objetos de la superficie, es posible obtener datos relacionados con sus propiedades físicas y químicas sin necesidad de contacto físico (1–3). En el contexto de las imágenes multiespectrales, el aprovechamiento de la información de varios espectros, como el visible y el infrarrojo, permite una detección de objetos robusta y fiable en entornos abiertos (4,5). En comparación con la detección de objetos mediante imágenes RGB, las imágenes multiespectrales han demostrado ser más precisas en diferentes entornos, ofreciendo una solución eficaz en condiciones de poca iluminación y condiciones climáticas adversas, donde las imágenes RGB a menudo fallan (4,6,7).

Las imágenes multiespectrales se han aplicado ampliamente en varios campos para la detección y clasificación de materiales. En la agricultura, han sido fundamentales para identificar especies de plantas en función de sus firmas espectrales, facilitando la detección de áreas propensas a plagas y evaluando las condiciones de los cultivos, como en las plantaciones de café, a través de imágenes convencionales y capturadas por drones (8–10). Los avances recientes en el aprendizaje automático han mejorado aún más las aplicaciones de imágenes multiespectrales en el monitoreo agrícola. Por ejemplo, las imágenes multiespectrales de alta resolución recopiladas a través de vehículos aéreos no tripulados, combinadas con modelos de máquinas de vectores de soporte (SVM), se han utilizado para estimar variables agronómicas clave en el cultivo de arroz inundado. Este enfoque demostró fuertes correlaciones para predecir el índice de área foliar (LAI), el contenido de nitrógeno (Narea) y el rendimiento de grano, logrando un coeficiente de correlación de Pearson de hasta 0.89 en entrenamiento y 0.87 en pruebas (11).

De manera similar, en el contexto ambiental, se ha demostrado que el procesamiento digital de imágenes satelitales monitorea eficazmente fenómenos como la degradación de la tierra mediante el análisis de las variaciones de las imágenes a lo largo del tiempo (12). La integración de imágenes multiespectrales con métodos de análisis de imágenes basadas en objetos (OBIA) se ha empleado con éxito en la clasificación de especies de manglares dentro de los ecosistemas de humedales. Un estudio realizado en la comuna de Dong Rui, en el norte de Vietnam, utilizó vehículos aéreos no tripulados Phantom 4 para adquirir datos de reflectancia espectral para la clasificación de especies de árboles, logrando una precisión general del 91,11% y un coeficiente kappa de 0,87. Este método proporcionó una base confiable para la planificación de la conservación de humedales y el monitoreo de especies, demostrando el potencial de las imágenes multiespectrales en la investigación ecológica (13).

Además, la alta resolución espectral de las imágenes multiespectrales ha permitido un análisis detallado de las condiciones de la vegetación, incluida la detección de componentes específicos como nutrientes y pigmentos como la clorofila (14–16). La fusión de imágenes multiespectrales con técnicas de aprendizaje profundo ha mejorado aún más el monitoreo agrícola, como lo demuestra un estudio sobre cultivos de piña MD2 en Colombia. Utilizando datos multiespectrales derivados de UAV, datos de sensores ecológicos in situ y técnicas de aprendizaje automático como los regresores

Extreme Gradient Boosting (XGBoost) y Multilayer Perceptron (MLP), los investigadores lograron la estimación del contenido de nitrógeno con un coeficiente de determinación ( $R^2$ ) de hasta el 86,98%. Este estudio subraya la importancia de integrar imágenes multiespectrales con modelos computacionales avanzados para mejorar el diagnóstico de nutrientes y optimizar la gestión agrícola (17).

Más allá de las aplicaciones ambientales, las imágenes multiespectrales han demostrado ser efectivas en el control y clasificación de la calidad de los alimentos. La combinación de imágenes multiespectrales basadas en reflectancia y autofluorescencia con algoritmos de aprendizaje automático se ha empleado con éxito para diferenciar granos de café verde especiales y tradicionales. En este contexto, los modelos SVM lograron una precisión impresionante del 96% en la clasificación del café verde, y los datos de autofluorescencia desempeñaron un papel crucial en la distinción de diferentes tipos de café en función de sus composiciones químicas (18).

Además, la integración de imágenes multiespectrales con datos de radar mejora la capacidad de monitorear la distribución de especies y las estructuras de los ecosistemas, lo que es particularmente valioso para la conservación de la biodiversidad (19). Las imágenes multiespectrales también han demostrado un potencial significativo en la detección de objetos en condiciones de visibilidad limitada, como las que se encuentran durante las operaciones de búsqueda y rescate cuando se combinan con técnicas de aprendizaje profundo (20). En aplicaciones geológicas, se han utilizado imágenes multiespectrales para el mapeo de minerales, lo que permite la identificación de compuestos como caolinita, illita, alunita, calcita, dolomita, hematita, goethita y cuarzo. Además, estas imágenes facilitan el mapeo simultáneo de zonas de alteración, como las zonas propilíticas y argílicas, que son cruciales para la exploración minera y la industria minera (21,22).

Por lo tanto, las imágenes multiespectrales, particularmente cuando se integran con modelos de aprendizaje automático, han avanzado significativamente la precisión y la eficiencia de las tareas de clasificación y detección en varios dominios. Desde la agricultura de precisión y la conservación de humedales hasta el control de la calidad de los alimentos y la exploración de minerales, la combinación de datos espectrales de alta resolución con técnicas computacionales continúa mejorando las capacidades analíticas, ofreciendo nuevos conocimientos y soluciones prácticas en diversos campos.

En cuanto a la detección de fibrocemento, que es el foco principal de esta investigación, es importante destacar que las imágenes hiperespectrales se han utilizado principalmente para la detección y clasificación efectiva de amianto en productos de cemento. Estos métodos permiten la identificación de diferentes tipos de minerales de asbestos, como amosite, crocidolita y crisotilo, aprovechando sus firmas espectrales únicas en el rango infrarrojo de onda corta (SWIR: 1000-2500 nm) (23). Del mismo modo, se han empleado imágenes hiperespectrales para evaluar el estado de deterioro de los techos de fibrocemento, contribuyendo al mapeo y priorización de la eliminación de estos materiales peligrosos que presentan importantes riesgos para la salud (24,25).

Además, las redes neuronales convolucionales y otras técnicas de clasificación se han aplicado de manera efectiva para identificar techos de fibrocemento utilizando datos derivados de imágenes hiperespectrales, lo que demuestra que una alta resolución espectral es crucial para

una clasificación y discriminación precisas del asbesto (26). Estas investigaciones previas son particularmente relevantes para la salud pública, ya que el asbesto es un carcinógeno conocido asociado con el mesotelioma y los cánceres de pulmón, laringe y ovario, e incluso la exposición a niveles bajos aumenta significativamente el riesgo de desarrollar estas enfermedades (27–29).

Teniendo en cuenta que las imágenes hiperespectrales se han utilizado de manera amplia y efectiva en la detección de asbesto a través de técnicas de teledetección, es importante resaltar que la alta dimensionalidad de estas imágenes requiere un uso eficiente de la memoria y la implementación de arquitecturas de alto rendimiento basadas en computación paralela (30–32). Además, dado que las imágenes hiperespectrales contienen un gran número de bandas, algunas de las cuales pueden ser redundantes, se requieren técnicas de reducción de dimensionalidad para facilitar su interpretación y clasificación (33–35).

Esta alta dimensionalidad puede aumentar los tiempos de procesamiento, por lo que es esencial emplear algoritmos o métodos más eficientes capaces de gestionar la complejidad sin comprometer la precisión (32,36). La complejidad inherente al procesamiento de imágenes hiperespectrales se reduce significativamente cuando se procesan imágenes multiespectrales satelitales, que requieren una menor demanda computacional y consumen menos energía (37–39). Por lo tanto, se hace necesario evaluar y seleccionar los modelos más adecuados para el contexto de las imágenes satelitales para permitir la detección eficiente de fibrocemento sin sacrificar la precisión.

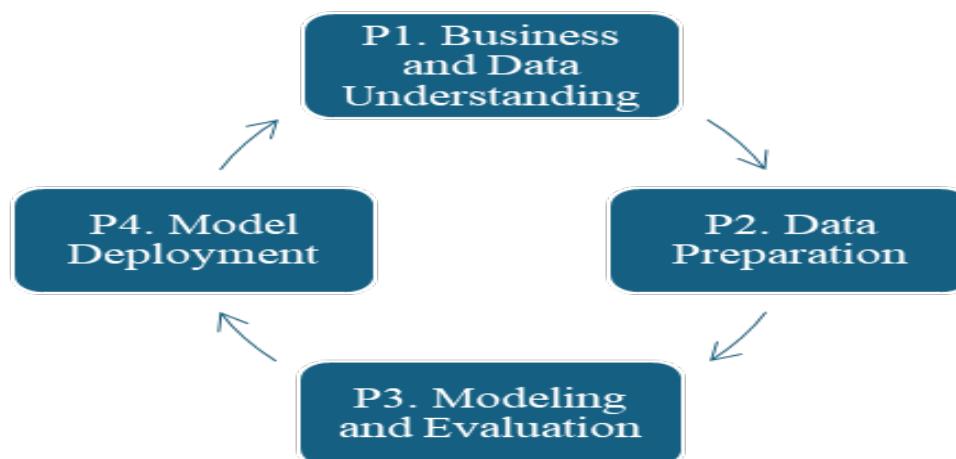
Este artículo propone la aplicación de modelos de aprendizaje automático para la detección de fibrocemento en imágenes multiespectrales satelitales como una contribución, con el objetivo de identificar el modelo con la mayor precisión de clasificación para el amianto basado en la caracterización adecuada de su firma espectral. Para lograr esto, inicialmente se construyó un conjunto de datos, que comprende 100 observaciones de 8 atributos correspondientes a la información espectral de 8 bandas de 50 píxeles de fibrocemento y 50 píxeles de otros materiales. Vale la pena señalar que las observaciones del conjunto de datos se muestraron a partir de una imagen multiespectral VNIR (Visible e Infrarrojo Cercano) de referencia obtenida de la ciudad de Cartagena, Colombia, con 8 bandas espectrales que cubren tanto el espectro visible (425,0 nm a 660,0 nm) como el espectro del infrarrojo cercano (725,0 nm a 950,0 nm).

Este conjunto de datos se utilizó para entrenar y validar cuatro modelos de aprendizaje automático utilizando la herramienta de programación visual Orange. De manera similar, para implementar el modelo con el mejor rendimiento y consistencia tanto en los conjuntos de entrenamiento como en los de prueba, se aprovecharon las ventajas proporcionadas por bibliotecas como Spectral, NumPy, pandas y Scikit-learn para implementar y evaluar el modelo en la imagen de referencia completa, que tiene dimensiones de 1500x1500 píxeles, cada una con ocho bandas espectrales. Se seleccionó la herramienta Visual Orange, considerando que proporciona una interfaz gráfica que facilita la construcción visual de los flujos de trabajo de aprendizaje automático, permitiendo tanto a principiantes como a expertos realizar análisis complejos, como la clasificación de imágenes y la agrupación, sin necesidad de programación (40).

Los resultados de esta investigación pretenden servir de referencia para aplicaciones académicas e industriales en la integración de sistemas ambientales para la detección de materiales en imágenes multiespectrales. El resto del artículo está organizado de la siguiente manera: la Sección 2 describe el enfoque metodológico, la Sección 3 presenta los resultados y la Sección 4 describe las conclusiones y las direcciones futuras de la investigación.

## Metodología

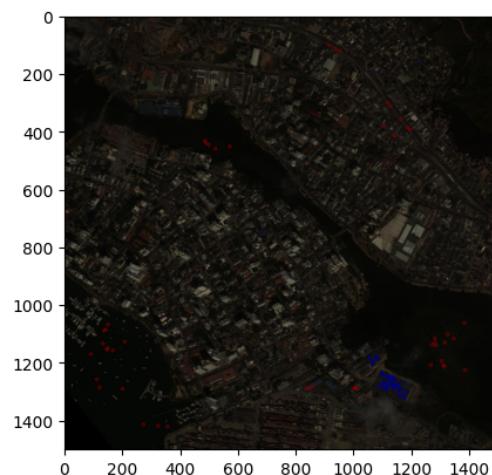
Para el desarrollo de esta investigación, la metodología CRISP-DM se adaptó en cuatro fases: P1. Negocios y comprensión de datos, P2. Preparación de datos, P3. Modelado y evaluación, y P4. Implementación del modelo (consulte la Figura 1). Aunque la Figura 1 presenta la metodología de manera secuencial, es importante tener en cuenta que CRISP-DM es un proceso iterativo. En la práctica, si el rendimiento del modelo no es óptimo, puede ser necesario revisar fases anteriores, como la preparación de datos o la selección de características, para refinar los resultados. Esta naturaleza cíclica garantiza la mejora continua y la adaptación a lo largo del proceso de modelado. Esta metodología se adaptó considerando que es un modelo de proceso estándar ampliamente utilizado para el desarrollo de proyectos de minería de datos y aprendizaje automático, con la ventaja de ser independiente del sector industrial y del tipo de tecnología empleada, lo que la hace aplicable a diversos contextos (41-44).



**Figura 1.** Metodología considerada. Fuente: elaboración propia.

En la Fase 1 de la metodología, se exploraron inicialmente herramientas para acceder a la información de banda espectral de imágenes satelitales, identificando que las bibliotecas de Python Spectral y NumPy permiten el procesamiento de datos de banda espectral de dichas imágenes. Utilizando estas librerías seleccionadas y basándose en una imagen multiespectral VNIR de 8 bandas preprocesada de 1500x1500 píxeles de la ciudad de Cartagena, que se sometió a corrección atmosférica, se extrajeron datos correspondientes a 50 píxeles de fibrocemento y 50 píxeles de otros materiales. Cabe mencionar que los píxeles de la muestra se seleccionaron mediante inspección visual de una representación RGB de la imagen multiespectral, eligiendo techos ubicados en áreas donde previamente se habían realizado visitas de campo y pruebas de verificación por parte de los profesionales involucrados en el proyecto dentro del cual se enmarca

esta investigación. En consecuencia, la Figura 2 presenta la representación RGB de la imagen satelital de referencia, donde los 50 píxeles muestreados de asbestos están resaltados en azul y los 50 píxeles de otros materiales están marcados en rojo.

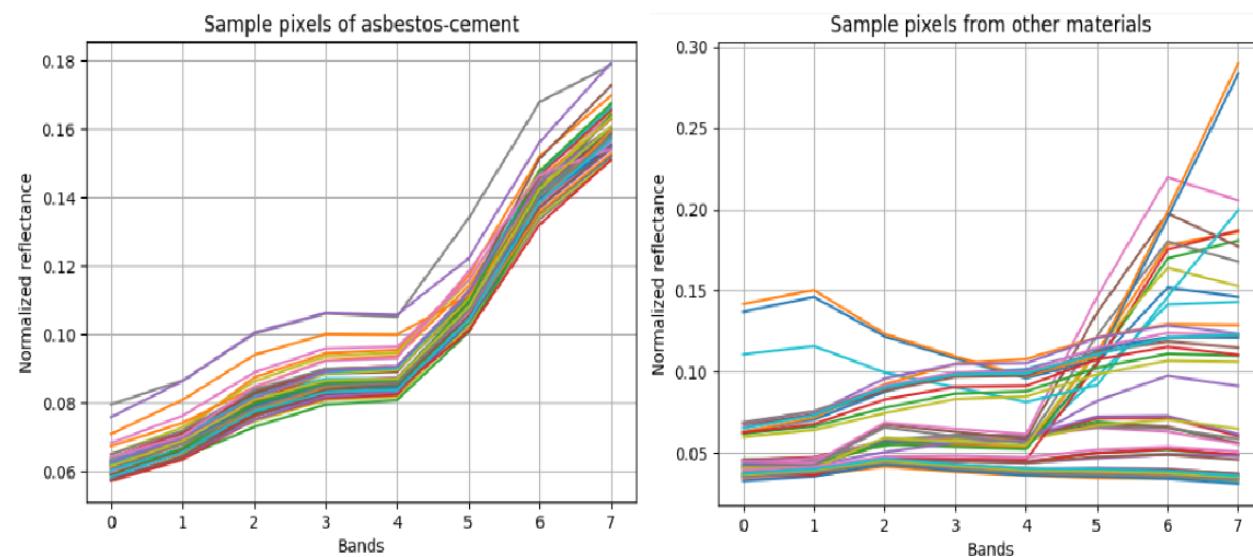


**Figura 2.** Píxeles de muestra seleccionados de fibrocemento y otros materiales. Fuente: elaboración propia.

En la Fase 2 de la metodología, los datos de las 8 bandas de los 50 píxeles de asbestos se normalizaron y etiquetaron con un valor de 1, mientras que los datos correspondientes a las 8 bandas de los 50 píxeles de otros materiales se etiquetaron con un valor de 0, utilizando las bibliotecas Pandas y NumPy. La normalización se realizó identificando los valores de reflectancia máxima y mínima para escalar los datos en consecuencia. Estos datos se integraron y organizaron en una estructura de marco de datos para crear el conjunto de datos de trabajo.

El conjunto de datos contiene un número igual de instancias etiquetadas como asbestos y otros materiales, lo que indica una distribución de clases equilibrada, lo cual es esencial para evitar que los modelos desarrollen sesgos o favorezcan inherentemente una de las dos etiquetas definidas, lo que garantiza un proceso de entrenamiento más equitativo y representativo.

Para facilitar el procesamiento posterior, este conjunto de datos se exportó a un archivo de Excel para su posterior carga en el flujo de trabajo de aprendizaje automático de la herramienta Orange. En consecuencia, la Figura 3 presenta tanto las 50 firmas espectrales de los píxeles de asbestos como las firmas espectrales de los 50 píxeles que no son de asbestos incluidos en el conjunto de datos. Es evidente que para cada material, las firmas espectrales informan los valores de reflectancia normalizados en las 8 bandas espectrales.



**Figura 3.** Píxeles de muestra seleccionados de fibrocemento y otros materiales. Fuente: elaboración propia.

De manera similar, la Tabla 1 muestra el conjunto de datos construido a partir de las firmas espectrales presentadas en la Figura 3, donde cada firma espectral se etiquetó con un 1 o un 0, dependiendo de si corresponde a fibrocemento u otro material. El conjunto de datos consta de 100 instancias, cada una de las cuales contiene 9 columnas: las primeras 8 columnas representan los valores de reflectancia para las respectivas bandas espectrales, y la 9<sup>a</sup> columna contiene la etiqueta asignada según el tipo de material.

**Tabla 1.** Píxeles de muestra seleccionados de fibrocemento y otros materiales

Instancia	Banda 1	Banda 2	Banda 3	Banda 4	Banda 5	Banda 6	Banda 7	Banda 8	Píxel
0	0.05979	0.06693	0.07807	0.08426	0.08514	0.10567	0.14085	0.16016	1
1	0.06298	0.07184	0.08747	0.09453	0.09542	0.11713	0.15215	0.16987	1
2	0.06296	0.07049	0.08320	0.08989	0.09066	0.11107	0.14334	0.15987	1
3	0.06235	0.06959	0.07991	0.08633	0.08715	0.10551	0.13862	0.15846	1
4	0.05951	0.06591	0.07627	0.08221	0.08296	0.10309	0.13552	0.15231	1
...	...	...	...	...	...	...	...	...	...
95	0.13716	0.14646	0.12032	0.10737	0.09584	0.10583	0.19554	0.28928	0
96	0.03956	0.03842	0.04956	0.04873	0.04787	0.10939	0.17577	0.18053	0
97	0.04169	0.04147	0.04678	0.05234	0.05395	0.07482	0.09049	0.08792	0
98	0.04176	0.04333	0.05508	0.05436	0.05339	0.10384	0.16483	0.17557	0
99	0.04113	0.04058	0.05251	0.05329	0.05305	0.11339	0.17822	0.18324	0

100 filas x 9 columnas

En la Fase 3 de la metodología, los atributos del conjunto de datos se analizaron inicialmente utilizando métodos de ganancia de información (Gain Ratio y Chi-Square), que son bien conocidos por su efectividad para identificar la relevancia de los atributos en relación con la variable objetivo, y gráficos de violina, que permiten la visualización simultánea de la distribución y densidad de datos dentro de cada categoría del atributo predictor. Esto facilitó una estimación precisa de los atributos con mayor impacto en el atributo predictor, que en este caso corresponde al tipo de material (columna de píxeles).

Dentro de esta fase, los datos se dividieron en conjuntos de entrenamiento y prueba utilizando el componente Data Sampler en Orange. A continuación, se entrenaron y ajustaron cuatro modelos convencionales de aprendizaje automático: KNN, máquinas de vectores de soporte, regresión logística y árbol de decisión, utilizando el conjunto de entrenamiento. Cabe mencionar que se eligieron modelos convencionales de aprendizaje automático, considerando que aunque los modelos basados en aprendizaje profundo exhiben una alta efectividad, a menudo requieren estructuras de red complejas y grandes volúmenes de datos, que pueden ser intensivos en términos de tiempo de procesamiento y recursos computacionales, lo que los hace inadecuados para aplicaciones de monitoreo ambiental (45). Finalmente, los componentes de evaluación se integraron en el flujo de trabajo de Orange para evaluar los modelos tanto en el conjunto de entrenamiento como en el de prueba, obteniendo métricas de evaluación clave (exactitud, precisión, recuperación y puntuación F1), así como la matriz de confusión para el conjunto de prueba. Con base en los resultados de estas métricas, se seleccionó el modelo con la mejor consistencia y rendimiento en el entrenamiento y las pruebas para su aplicación a la imagen de referencia completa. Cabe mencionar que no se consideró la métrica AUC-ROC, ya que es más útil y proporciona información significativa en el contexto de conjuntos de datos desequilibrados (46).

En la Fase 4 de la metodología, se implementó el modelo con el mejor rendimiento y consistencia y se aplicó a toda la imagen multiespectral, aprovechando las ventajas proporcionadas por las bibliotecas de Python Spectral, Pandas, Scikit-learn y NumPy. Además, en base a la detección realizada por el método en toda la imagen, se determinó el porcentaje de píxeles correspondientes al asbesto. El porcentaje resultante es de particular interés para las autoridades ambientales para identificar techos de asbesto en los barrios de Cartagena.

De acuerdo con lo anterior, el presente estudio tiene como objetivo abordar las siguientes preguntas: ¿Cuál es el rendimiento de los métodos de aprendizaje automático para la detección de amianto en imágenes multiespectrales? ¿Cuál es la eficacia y eficiencia del modelo de mejor rendimiento en comparación con los métodos convencionales? ¿En qué campos de aplicación se puede extrapolar el enfoque propuesto?

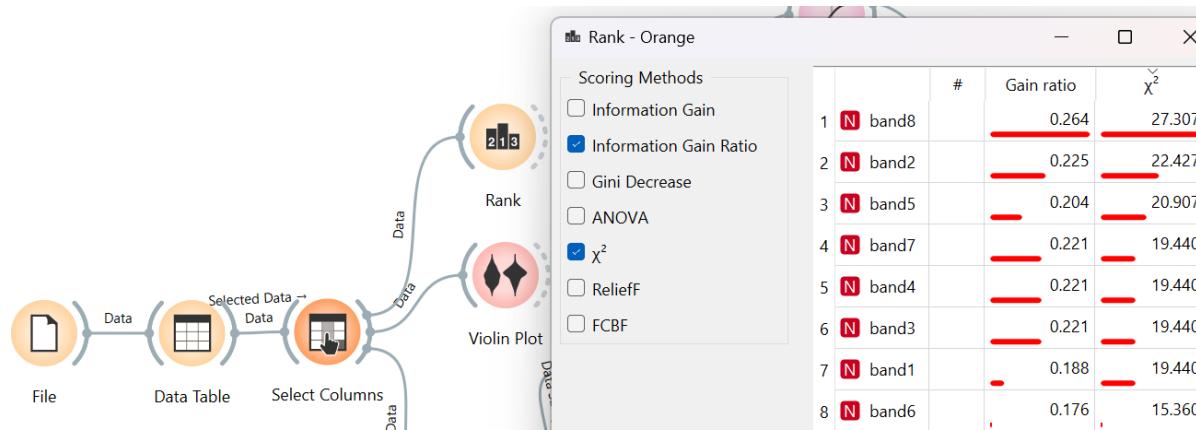
## Resultados

En cuanto a los resultados, el conjunto de datos de la Tabla 1 se cargó primero en la herramienta de programación visual Orange, y se aplicó el módulo “Rank” para determinar los atributos con mayor impacto en el atributo predictor. Esto se logró utilizando los métodos Gain Ratio y Chi-Square, con los resultados presentados en la Figura 4.

De acuerdo con los resultados presentados en la Figura 4, los métodos Gain Ratio y Chi-Square identificaron que las bandas que proporcionan la contribución más significativa al atributo predictor (columna de píxeles del conjunto de datos) son las bandas 8, 2 y 5, cada una con un porcentaje de ganancia de información superior al 20,9%. La banda 8 fue la más relevante, con un valor de Gain Ratio de 0,264 y una contribución de Chi-cuadrado de 27,307.

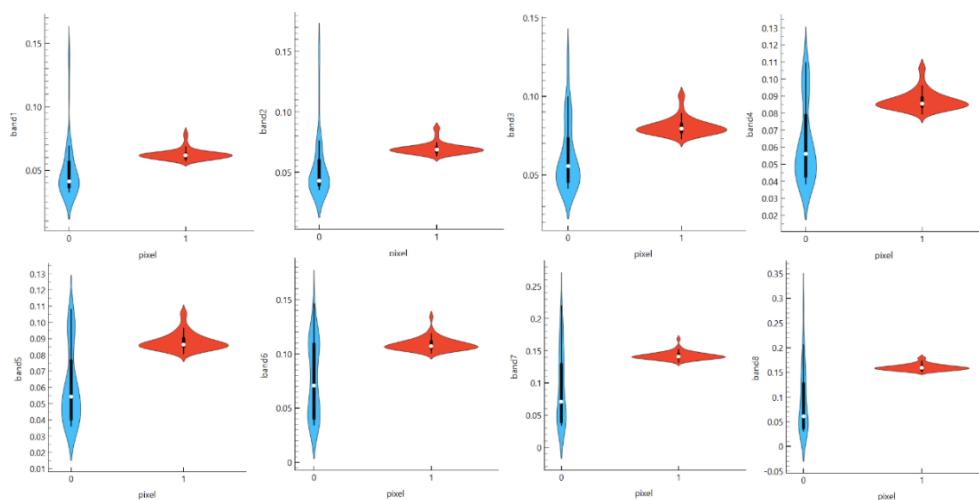
Es posible observar cómo las bandas 2 (480 nm, azul), 5 (660 nm, rojo) y 8 (950 nm, infrarrojo cercano) pueden ser clave para la detección de amianto-cemento debido a su interacción con la

composición espectral del material. La Banda 2 destaca su reflectancia en el rango visible, la Banda 5 captura la absorción de minerales característicos y la Banda 8 permite la diferenciación a través de su respuesta en el infrarrojo. Estas bandas, colectivamente, parecen optimizar la identificación del asbesto-cemento al capturar sus propiedades espectrales distintivas en diferentes rangos del espectro electromagnético.



**Figura 4.** Aplicación de métodos de obtención de información

De igual forma, se observó que la banda 6 proporciona la menor información, con un valor de Razón de Ganancia de 0,176 y una contribución porcentual de 15,360% en Chi-Cuadrado. Para explorar más a fondo la posible relación entre la distribución de los valores de reflectancia en las 8 bandas de la imagen multiespectral y las categorías de la variable predictora (0 o 1), se utilizaron gráficos de violín a través del módulo “Gráfico de violín” en naranja (ver Figura 5).

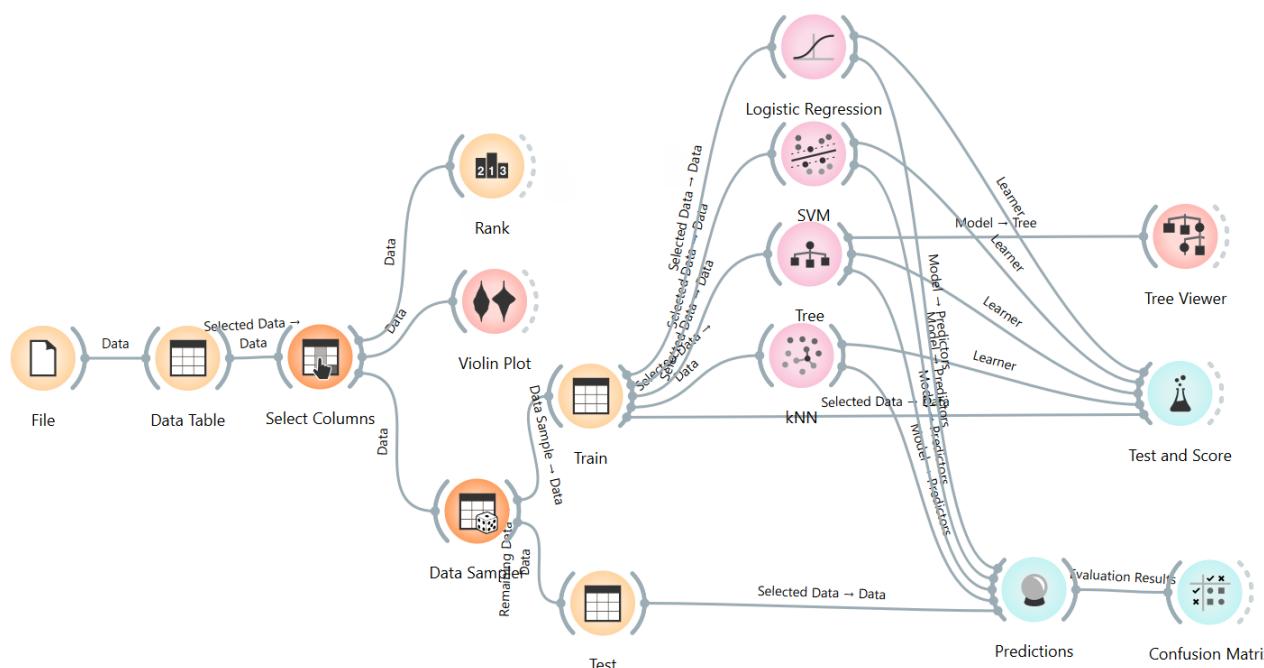


**Figura 5.** Gráficos de violín de las bandas de reflectancia en la imagen multiespectral

A partir de los gráficos de violín presentados en la Figura 5, es evidente que la distribución de los valores de reflectancia en las bandas 8, 2 y 5 permite una diferenciación más clara entre los tipos de píxeles 0 (sin asbesto) y 1 (asbesto), ya que no se observa una superposición significativa en sus densidades. Esto se alinea con los resultados obtenidos a través del método Chi-Cuadrado, donde se destacan estas tres bandas con un porcentaje de ganancia de información superior al 20,9%.

En comparación, las bandas restantes exhiben una mayor similitud en sus distribuciones, lo que las hace menos útiles para clasificar los tipos de píxeles. Además, se observa que, en general, los valores de reflectancia para el tipo de píxel 0 muestran una mayor dispersión y variabilidad, mientras que los del tipo de píxel 1 exhiben distribuciones más compactas concentradas dentro de rangos más estrechos. Estos hallazgos sugieren que las bandas 8, 2 y 5 son particularmente efectivas para distinguir entre asbesto y otros materiales, de acuerdo con los resultados obtenidos utilizando métodos de obtención de información.

Una vez analizadas las bandas de reflectancia más relevantes, se desplegó el flujo de trabajo completo para los cuatro modelos de aprendizaje automático utilizando la herramienta de programación visual Orange, como se ilustra en la Figura 6.



**Figura 6.** Flujo de trabajo de los modelos de aprendizaje automático en Orange

La Figura 6 ilustra claramente los diversos procesos llevados a cabo en el flujo de trabajo desarrollado, que incluyen: cargar el conjunto de datos en formato Excel utilizando el módulo "Archivo"; seleccionando los 8 atributos (bandas de reflectancia) y el atributo predictor (tipo de píxel) de las 9 columnas del conjunto de datos utilizando el módulo "Seleccionar columnas".

Posteriormente, se aplicaron métodos de ganancia de información y gráficos de violín a través de los módulos "Rango" y "Gráfico de violín". Luego, el conjunto de datos se dividió en conjuntos de entrenamiento (70%) y prueba (30%) utilizando el módulo "Muestreador de datos". Los cuatro modelos de aprendizaje automático considerados (KNN, árboles de decisión, máquinas de vectores de soporte y regresión logística) se entrenaron a través de los módulos "kNN", "Tree", "SVM" y "Logistic Regression".

A continuación, se visualizó el diagrama de árbol de decisión con el módulo "Visor de árboles". Los modelos se validaron en el conjunto de entrenamiento utilizando el módulo "Test and Score" y en

el conjunto de pruebas utilizando los módulos "Predictions" y "Confusion Matrix". Una vez que se implementó el flujo de trabajo en la herramienta Orange, se realizó el análisis de las métricas del modelo para los conjuntos de entrenamiento y prueba, como se presenta en las Tablas 2 y 3.

**Tabla 2.** Resultados de los modelos sobre el rendimiento del conjunto de trenes

Modelo	AUC	CA	F1	Precisión	Recordar
<b>kNN</b>	0.986	0.971	0.971	0.973	0.971
<b>Árbol</b>	0.906	0.929	0.928	0.932	0.929
<b>SVM</b>	1.000	0.943	0.943	0.949	0.943
<b>Regresión logística</b>	0.860	0.771	0.759	0.843	0.771

**Tabla 3.** Resultados de los modelos Rendimiento del conjunto de pruebas

Modelo	AUC	CA	F1	Precisión	Recordar
<b>kNN</b>	1.000	1.000	1.000	1.000	1.000
<b>Árbol</b>	0.933	0.933	0.933	0.933	0.933
<b>SVM</b>	1.000	1.000	1.000	1.000	1.000
<b>Regresión logística</b>	0.800	0.833	0.829	0.875	0.833

Teniendo en cuenta los resultados obtenidos por los modelos en las métricas de Precisión y Retiro, cabe destacar que el modelo kNN con tres vecinos, árboles de decisión y el modelo de máquina de vectores de soporte (SVM) con un kernel RBF alcanzaron los valores más altos en el conjunto de entrenamiento. Entre estos, el modelo kNN mostró el mejor desempeño, con métricas de Precisión y Recuperación de 0.973 y 0.971, respectivamente, seguido por el modelo SVM con valores de 0.949 y 0.943, y el modelo de árbol de decisión con valores de 0.932 y 0.939, todos demostrando un desempeño adecuado.

Por el contrario, el modelo de regresión logística alcanzó valores de precisión y recuperación de 0,843 y 0,771, respectivamente, lo que indica un rendimiento más limitado en comparación con los otros tres modelos. En el conjunto de pruebas, tanto el modelo kNN como el SVM mantuvieron un rendimiento perfecto (1.0) en ambas métricas, mientras que el modelo de árbol de decisión registró valores ligeramente más bajos de 0.933 tanto en Precisión como en Recuperación.

Esta ligera disminución en el rendimiento del modelo de árbol de decisión es un resultado positivo, ya que indica una mejor capacidad de generalización y un menor riesgo de sobreajuste, lo que enfatiza la importancia de lograr resultados más equilibrados entre los conjuntos de entrenamiento y prueba para garantizar la solidez del modelo. Por lo tanto, aunque se pudieron utilizar los tres modelos con el mejor rendimiento, el modelo de árbol de decisión demostró la mayor consistencia y se seleccionó para su despliegue en toda la imagen multiespectral de referencia. En consecuencia, la Figura 7 presenta la matriz de confusión obtenida para el modelo seleccionado.

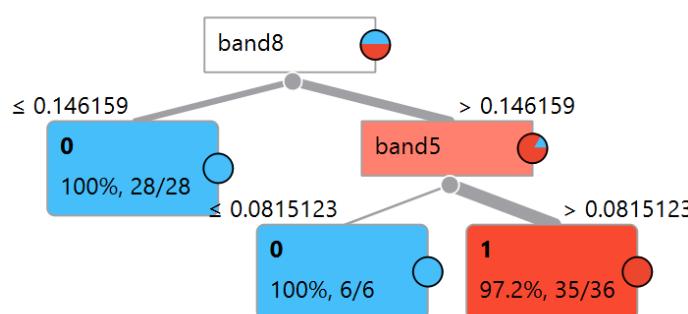
		Predicted		$\Sigma$
		0	1	
Actual	0	14	1	15
	1	1	14	15
$\Sigma$		15	15	30

**Figura 7.** Matriz de confusión obtenida para el modelo de árbol de decisión seleccionado

La Figura 7 muestra que, de un total de 30 casos en el conjunto de prueba (15 de fibrocemento y 15 de otros materiales), el modelo de árbol de decisión clasificó correctamente 14 casos como pertenecientes a otros materiales y clasificó erróneamente 1 como fibrocemento. Para los casos de fibrocemento, el modelo identificó correctamente 14 de los 15, con uno clasificado erróneamente como perteneciente a otros materiales. Esto corresponde a una tasa de precisión del 93,3% en la clasificación de muestras de fibrocemento, lo que demuestra un rendimiento adecuado con una ligera tasa de error en esta categoría.

Además, la Figura 8 ilustra el árbol de decisión generado por el modelo, destacando los atributos identificados como más relevantes, siendo las bandas 8 y 5 las más significativas en el conjunto de datos. Esto se alinea con los resultados obtenidos a través de métodos de obtención de información y gráficos de violín, donde estas bandas se encuentran entre los tres atributos más relevantes.

En concreto, la regla que habilita la clasificación de los píxeles de fibrocemento (etiqueta 1) establece que si el valor de reflectancia de la banda 8 supera 0,146159 y el de la banda 5 es superior a 0,0815123, el modelo clasifica estos píxeles como pertenecientes a esta categoría, consiguiendo una precisión del 97,2% (35/36). Este resultado subraya la capacidad del modelo para discriminar eficazmente entre clases, particularmente en el caso de píxeles de fibrocemento.



**Figura 8.** Árbol de decisión obtenido por el modelo

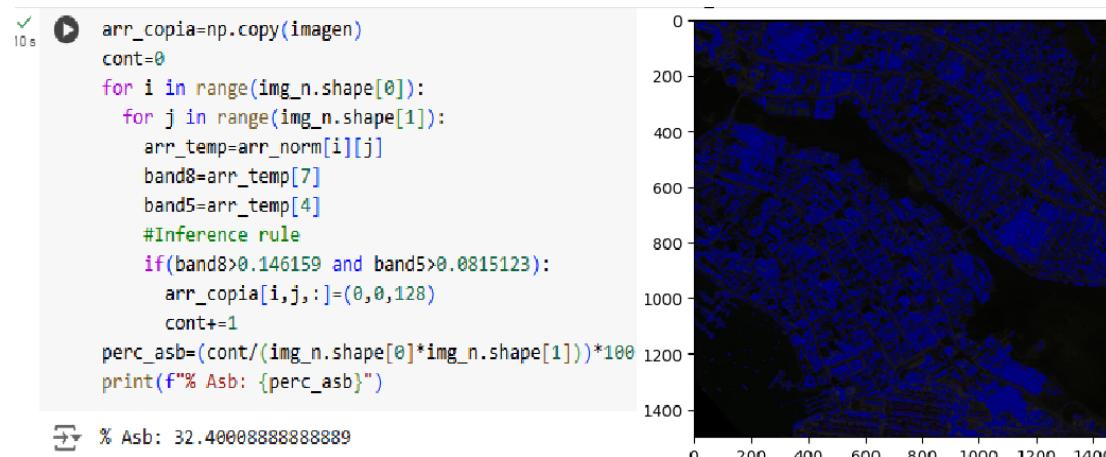
En consonancia con lo anterior, en la tabla 4 se presentan las tres reglas de inferencia derivadas del árbol de decisión, que se refieren a las bandas 8 y 5. Estas tres reglas son muy útiles para clasificar

los píxeles de asbesto en toda la imagen, particularmente la Regla 3, que se refiere a la detección de píxeles de fibrocemento.

**Tabla 4.** Reglas de inferencia obtenidas del modelo de árbol de decisión

Identificación	Regla de inferencia
1	Si la banda 8 $\leq 0.146159$ , entonces la clase es 0 (otros materiales), con una precisión del 100% (28/28).
2	Si la banda 8 $> 0.146159$ y la banda 5 $\leq 0.0815123$ , entonces la clase es 0 (otros materiales), con una precisión del 100% (6/6).
3	Si la banda 8 $> 0.146159$ y la banda 5 $> 0.0815123$ , entonces la clase es 1 (fibrocemento), con una precisión del 97,2% (35/36).

Con base en las reglas de inferencia obtenidas y presentadas en la Tabla 4, se aplicó la Regla 3 para la identificación de píxeles de fibrocemento en el total de píxeles de la imagen multiespectral de referencia que se muestra en la Figura 2, utilizando las bibliotecas Spectral, NumPy, Pandas y Matplotlib. La Figura 9 ilustra la implementación del método de detección de asbesto en Python, donde el algoritmo itera a través de cada píxel de la imagen, extrayendo sus 8 bandas espectrales y aplicando la Regla de Inferencia 3, que involucra las bandas 8 y 5. Como resultado de este método, la Figura 9 resalta en azul las áreas detectadas que contienen fibrocemento en la imagen presentada en la Figura 2, revelando que el 32,4% de la imagen corresponde a fibrocemento.



**Figura 9.** Despliegue del método en toda la imagen

## Discusión

A nivel de discusión de resultados y con el fin de abordar la primera pregunta orientadora sobre el desempeño de los modelos de aprendizaje automático evaluados para la detección de fibrocemento, cabe mencionar que este estudio implementó un flujo de trabajo de aprendizaje automático utilizando la herramienta Orange, empleando cuatro modelos: k-Vecinos más cercanos (kNN), árboles de decisión, máquinas de vectores de soporte y regresión logística, para la

detección de fibrocemento en imágenes multiespectrales. En este sentido, es importante destacar que, en el conjunto de pruebas, los modelos de máquina kNN, árbol de decisión y vector de soporte lograron un rendimiento superior a 0,933 en las métricas Precision, Recall y F1-Score, mientras que el modelo de regresión logística exhibió el menor rendimiento en estas métricas, con valores que oscilaron entre 0,829 y 0,875. Este hallazgo lleva a la conclusión de que los tres modelos de mejor rendimiento demuestran una alta efectividad en la detección de fibrocemento en imágenes multiespectrales; sin embargo, el modelo de árbol de decisión muestra la mayor consistencia en los resultados en los conjuntos de entrenamiento y prueba.

Ahora, con respecto a la segunda pregunta orientadora relacionada con la comparación del modelo de mejor rendimiento con los métodos convencionales en términos de eficacia y eficiencia, es importante resaltar que el modelo que logró el mejor equilibrio entre rendimiento y consistencia en los conjuntos de entrenamiento y prueba fue el modelo de árbol de decisión. Este modelo obtuvo valores respectivos de 0,932 y 0,929 para Precisión y Recuperación en el conjunto de entrenamiento, mientras que en el conjunto de prueba, logró un valor de 0,933 para ambas métricas. Además, este modelo identificó que las bandas relevantes para la detección de fibrocemento son las bandas de reflectancia 8 y 5. La identificación de estas bandas es crucial, ya que sugiere que la eficiencia computacional se optimiza a través de este método al considerar solo dos bandas en la regla de inferencia para la detección. Por el contrario, los métodos convencionales, como el enfoque basado en la correlación, utilizan el número total de bandas espectrales de la imagen para sus cálculos, como se presenta en (47), donde se calcula la similitud entre el píxel medio o característico y los píxeles restantes de la imagen.

Del mismo modo, en cuanto a la tercera pregunta orientadora relacionada con los campos en los que se puede extrapolar el enfoque propuesto, cabe mencionar que, dado que el método desarrollado reduce la complejidad de la identificación del amianto-cemento al utilizar solo un par de bandas para la detección, puede aplicarse en investigaciones centradas en el monitoreo ambiental (14–16), donde el área cubierta por las imágenes es extensa y la eficiencia es un requisito fundamental.

En este mismo sentido, se trata de investigaciones centradas en la detección del amianto mediante imágenes hiperespectrales (48), este estudio proporciona un hallazgo relevante, ya que demuestra que con solo unas pocas bandas de reflectancia, los métodos de aprendizaje automático, como los árboles de decisión, permiten la detección eficiente y efectiva de este material. Esto no solo optimiza los costes computacionales, sino que también reduce los gastos en el proceso de recogida de muestras, dado que la infraestructura necesaria para la captura y procesamiento de imágenes hiperespectrales conlleva un mayor coste económico.

## Conclusiones

Teniendo en cuenta que la mayoría de las investigaciones de vanguardia sobre la detección de fibrocemento en imágenes espectrales se han centrado en el uso de imágenes hiperespectrales debido a su precisión y la amplia información espectral que proporcionan, este artículo propone un método basado en el aprendizaje automático para la detección de fibrocemento dentro del dominio de las imágenes satelitales multiespectrales. Este método demostró no solo resultados

efectivos sino también eficiencia computacional, lograda utilizando un número menor de bandas y centrándose en un conjunto reducido de bandas relevantes para clasificar píxeles como fibrocemento. Con base en la eficiencia alcanzada por el modelo de árbol de decisión, el enfoque propuesto se puede extrapolar a dominios donde la eficiencia es primordial, como el monitoreo ambiental mediante imágenes espectrales, donde las imágenes cubren grandes áreas y la optimización de los tiempos de procesamiento es esencial.

En este estudio se evaluaron cuatro modelos de aprendizaje supervisado: kNN con 3 vecinos, árboles de decisión, máquinas de vectores de soporte con kernel RBF y regresión logística. Los modelos de máquina kNN, árbol de decisión y vector de soporte lograron métricas de precisión y recuperación superiores a 0,93, y el modelo de árbol de decisión demostró la mejor consistencia entre los conjuntos de entrenamiento y prueba. Además de su excelente efectividad en ambos conjuntos, el modelo de árbol de decisión demostró ser eficiente al utilizar una regla de inferencia que involucra solo la banda 8 y la banda 5 para la detección de fibrocemento, lo que lo hace competitivo con métodos como los enfoques basados en correlación. En este contexto, los resultados obtenidos representan una contribución significativa en comparación con los estudios de última generación, que se centran predominantemente en el uso de diversos métodos para la detección de amianto en imágenes hiperespectrales. El enfoque propuesto permite la detección de amianto a través de una regla de inferencia basada en la evaluación de dos bandas de reflectancia, lo que no solo optimiza los costes computacionales sino que también reduce los gastos asociados a la adquisición y muestreo de imágenes hiperespectrales.

Para el desarrollo de esta investigación, se utilizó una adaptación de la metodología de minería de datos CRISP-DM. Para implementar las diferentes fases de la metodología, se empleó tanto la herramienta de programación visual Orange como las bibliotecas de Python Spectral, NumPy, Pandas, Scikit-learn y Matplotlib. Las bibliotecas de Python mencionadas se utilizaron para construir el conjunto de datos a partir de una imagen multiespectral de referencia, así como para implementar el modelo de aprendizaje automático en la imagen de referencia completa. Por su parte, se utilizó la herramienta Orange para aplicar métodos de ganancia de información y gráficos de violín, así como para entrenar y evaluar los cuatro modelos de aprendizaje automático considerados, siguiendo la división del conjunto de datos en conjuntos de entrenamiento y prueba. Las herramientas de código abierto consideradas para la implementación del flujo de trabajo proporcionan una alternativa competitiva al software propietario para la detección de materiales en imágenes espectrales, lo que permite a las universidades y centros de investigación replicar y extrapolar estos métodos en diversas aplicaciones.

Como resultado de desplegar el modelo de mejor ajuste sobre la imagen de referencia completa correspondiente a un área de la ciudad de Cartagena, se determinó que el 32,4% del total de píxeles de la imagen se identificaron como fibrocemento. Este resultado tiene importancia no solo en términos computacionales sino también en el contexto de la salud pública, dadas las implicaciones negativas de la exposición humana a las fibras de este material. Por lo tanto, los hallazgos presentados en este artículo tienen como objetivo contribuir a la toma de decisiones estratégicas con respecto a los esfuerzos de mitigación del asbestos en Cartagena por parte de las autoridades gubernamentales.

## Declaración de contribución de autoría CRediT

Conceptualización - Ideas: Manuel Saba, Manuel Alejandro Ospina-Alarcón. Curación de datos: Gabriel Elías Chanchí-Golondrino. Análisis formal: Gabriel Elías Chanchí-Golondrino. Adquisición de financiación: Manuel Saba. Investigación: Manuel Alejandro Ospina-Alarcón, Manuel Saba. Metodología: Manuel Saba, Gabriel Elías Chanchí-Golondrino. Dirección de Proyecto: Manuel Saba. Recursos: Manuel Saba. Software: Gabriel Elías Chanchí-Golondrino. Supervisión: Manuel Alejandro Ospina-Alarcón. Validación: Manuel Alejandro Ospina-Alarcón. Visualización - Preparación: Gabriel Elías Chanchí-Golondrino. Redacción - borrador original - Elaboración: Gabriel Elías Chanchí-Golondrino, Manuel Alejandro Ospina-Alarcón. Redacción - revisión y edición - Elaboración: Manuel Alejandro Ospina-Alarcón, Manuel Saba.

Conflicto de intereses: no declara. Implicaciones éticas: no declara. Agradecimientos y declaración de financiación.

### Agradecimientos

Este artículo se considera un producto del proyecto «desarrollo de una estrategia integral para reducir el impacto en la salud pública y el medio ambiente debido a la presencia de asbesto en el territorio del departamento de Bolívar», financiado por el Sistema General de Regalías de Colombia (SGR) e identificado con el código BPIN 2020000100366. Este proyecto fue llevado a cabo por la Universidad de Cartagena, Colombia, y la Fundación Colombia Libre de Asbesto. Finalmente, los autores desean expresar su agradecimiento a Federico Frassy por su apoyo en la gestión y clasificación de los datos hiperespectrales, a Aiken Hernando Ortega Heredia, María Angélica Narváez Cuadro, Carlos Andrés Castrillón Ortiz, Michelle Cecilia Montero Acosta, Margareth Peña Castro, Carlos David Arroyo Angulo y al resto del equipo de investigación por el apoyo logístico y la recolección de muestras en campo. Asimismo, los autores extienden su agradecimiento a Juan Manuel González de la compañía BlackSquare por su ayuda en la adquisición de datos hiperespectrales y a Sean Fitzgerald por el análisis PLM.

Por otro lado, cabe mencionar que, aunque en el caso específico de este estudio, se obtuvieron excelentes resultados en la detección de asbesto-cemento sobre una imagen multiespectral que fue preprocesada aplicando las correcciones atmosféricas necesarias, este aspecto debe ser considerado cuidadosamente, ya que omitir correcciones atmosféricas puede afectar el rendimiento de los modelos de aprendizaje automático al introducir ruido y comprometer la discriminación de la firma espectral del asbesto-cemento. Del mismo modo, es importante aclarar que el muestreo de píxeles es una fase crítica del proceso, por lo que se recomienda recolectar muestras de firmas espectrales de áreas donde no solo se asegura la confirmación visual, sino que también se han realizado visitas de campo y pruebas de verificación para confirmar la presencia del material.

Como trabajo futuro, se debe considerar la aplicación del modelo a otras áreas de Cartagena y diferentes ciudades de Colombia y del mundo. Esto permitiría validar su efectividad en diversos contextos, fortalecer los hallazgos y ampliar el alcance de este estudio. Además, la evaluación de los métodos de conjunto podría mejorar la precisión al aprovechar múltiples modelos para mejorar la capacidad predictiva y la robustez.

También se planea una comparación de la eficiencia del método propuesto con enfoques que no son de aprendizaje automático. Esto ayudaría a evaluar el costo computacional y el tiempo de procesamiento, asegurando que el modelo siga siendo práctico para implementaciones a gran escala.

## References

1. Guzman-Alvarez JA, González-Zuñiga M, Sandoval Fernandez JA, Calvo-Alvarado JC. Uso de sensores remotos en la agricultura: aplicaciones en el cultivo del banano. Agron Mesoam. 2022 Aug 22;48279. <https://doi.org/10.15517/am.v33i3.48279>
2. Jiménez-López AF, Jiménez-López M, Jiménez-López FR. Multispectral analysis of vegetation for remote sensing applications. ITECKNE. 2015 Nov;12(2). <https://doi.org/10.15332/iteckne.v12i2.1242>
3. Awange J, Kiema J. Fundamentals of Remote Sensing. In 2019. p. 115-23. [https://doi.org/10.1007/978-3-030-03017-9\\_7](https://doi.org/10.1007/978-3-030-03017-9_7)
4. Fang Q, Han D, Wang Z. Cross-Modality Fusion Transformer for Multispectral Object Detection. SSRN Electron J. 2022; <https://doi.org/10.2139/ssrn.4227745>
5. Zhang H, Fromont E, Lefevre S, Avignon B. Multispectral Fusion for Object Detection with Cyclic Fuse-and-Refine Blocks. In: 2020 IEEE International Conference on Image Processing (ICIP). IEEE; 2020. p. 276-80. <https://doi.org/10.1109/ICIP40778.2020.9191080>
6. Hu S, Bonardi F, Bouchafa S, Prendinger H, Sidibé D. Rethinking Self-Attention for Multispectral Object Detection. IEEE Trans Intell Transp Syst. 2024 Nov;25(11):16300-11. <https://doi.org/10.1109/TITS.2024.3412417>
7. Takumi K, Watanabe K, Ha Q, Tejero-De-Pablos A, Ushiku Y, Harada T. Multispectral Object Detection for Autonomous Vehicles. In: Proceedings of the on Thematic Workshops of ACM Multimedia 2017. New York, NY, USA: ACM; 2017. p. 35-43. <https://doi.org/10.1145/3126686.3126727>
8. Sánchez-Méndez AG, Argujo-Hernández SP. Análisis de imágenes multiespectrales para la detección de cultivos y detección de plagas y enfermedades en la producción de café. Res Comput Sci. 2018 Dec;147(7):309-17. <https://doi.org/10.13053/rcs-147-7-24>
9. Pelaez Carrillo DA, Gualdrón Guerrero OE, Torres Chavez I. Optimización de la evaluación de pastos mediante la implementación de imágenes multiespectrales y vehículo aéreo no tripulado. Rev Colomb Tecnol Av. 2024 May;1(43):155-62. <https://doi.org/10.24054/rcta.v1i43.2850>
10. Meneses VAB, Téllez JM, Velasquez DFA. Uso de drones para el análisis de imágenes multiespectrales en agricultura de precisión. limentech, Cienc y Tecnol Aliment. 2017 Sep;13(1). <https://doi.org/10.24054/01204211.v1.n1.2015.1647>
11. Eugenio FC, Grohs M, Schuh MS, Venancio LP, Schons C, Badin TL, et al. Flooded rice variables from high-resolution multispectral images and machine learning algorithms. Remote Sens Appl Soc Environ. 2023 Aug;31:100998. <https://doi.org/10.1016/j.rsase.2023.100998>
12. Ormazábal Y, Ávila C, Mena C, Morales Y, Bustos Ó. Caracterización y cuantificación de fragmentos de bosque nativo, en un sector del secano interior de la Región del Maule, Chile. Ciéncia Florest. 2013 Aug;23(3):449-60. <https://doi.org/10.5902/1980509810556>

13. Ngo DT. Mapping tree species of wetlands using multispectral images of UAVs and machine learning: A case study of the Dong Rui Commune. *Helion*. 2024 Aug;10(15):e35159. <https://doi.org/10.1016/j.heliyon.2024.e35159>
14. Peña MA. Propiedades del primer sistema de teledetección hiperespectral chileno. *Rev Geogr Norte Gd.* 2010 Sep;(46). <https://doi.org/10.4067/S0718-34022010000200011>
15. Lu B, He Y, Dao PD. Comparing the Performance of Multispectral and Hyperspectral Images for Estimating Vegetation Properties. *IEEE J Sel Top Appl Earth Obs Remote Sens.* 2019 Jun;12(6):1784-97. <https://doi.org/10.1109/JSTARS.2019.2910558>
16. He J, Yuan Q, Li J, Zhang L. A Knowledge Optimization-Driven Network With Normalizer-Free Group ResNet Prior for Remote Sensing Image Pan-Sharpening. *IEEE Trans Geosci Remote Sens.* 2022;60:1-16. <https://doi.org/10.1109/TGRS.2022.3230846>
17. Chaparro JE, Aedo JE, Lumbreiras Ruiz F. Machine Learning for the estimation of foliar nitrogen content in pineapple crops using multispectral images and Internet of Things (IoT) platforms. *J Agric Food Res.* 2024 Dec;18:101208. <https://doi.org/10.1016/j.jafr.2024.101208>
18. Pinheiro Claro Gomes W, Gonçalves L, Barboza da Silva C, Melchert WR. Application of multispectral imaging combined with machine learning models to discriminate special and traditional green coffee. *Comput Electron Agric.* 2022 Jul;198:107097. <https://doi.org/10.1016/j.compag.2022.107097>
19. Schulte to Bühne H, Pettorelli N. Better together: Integrating and fusing multispectral and radar satellite imagery to inform biodiversity monitoring, ecological research and conservation science. Lecomte N, editor. *Methods Ecol Evol.* 2018 Apr;9(4):849-65. <https://doi.org/10.1111/2041-210X.12942>
20. Galán-Cuenca Á, Vázquez-Martín R, Mandow A, Morales J, García-Cerezo A. Análisis de técnicas de aumento de datos y entrenamiento en YOLOv3 para detección de objetos en imágenes RGB y TIR del UMA-SAR Dataset. In: XLII Jornadas De Automática: Libro De Actas. Servizo de Publicacións da UDC; 2021. p. 686-94. <https://doi.org/10.17979/spudc.9788497498043.686>
21. van der Meer FD, van der Werff HMA, van Ruitenbeek FJA, Hecker CA, Bakker WH, Noomen MF, et al. Multi- and hyperspectral geologic remote sensing: A review. *Int J Appl Earth Obs Geoinf.* 2012 Feb;14(1):112-28. <https://doi.org/10.1016/j.jag.2011.08.002>
22. Peyghambari S, Zhang Y. Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: an updated review. *J Appl Remote Sens.* 2021 Jul 14;15(03). <https://doi.org/10.1117/1.JRS.15.031501>
23. Sabat M, Fares N, Mitri G, Kfouri A. Determination of asbestos cement rooftop surface composition using regression analysis and hyper-spectral reflectance data in the visible and near-infrared ranges. *J Hazard Mater.* 2024 May;469:134006. <https://doi.org/10.1016/j.jhazmat.2024.134006>
24. Bassani C, Cavalli RM, Cavalcante F, Cuomo V, Palombo A, Pascucci S, et al. Deterioration status of asbestos-cement roofing sheets assessed by analyzing hyperspectral data. *Remote Sens Environ.* 2007 Aug;109(3):361-78. <https://doi.org/10.1016/j.rse.2007.01.014>

25. Cilia C, Panigada C, Rossini M, Candiani G, Pepe M, Colombo R. Mapping of Asbestos Cement Roofs and Their Weathering Status Using Hyperspectral Aerial Images. *ISPRS Int J Geo-Information*. 2015 Jun 1;4(2):928-41.

<https://doi.org/10.3390/ijgi4020928>

26. Kaplan G, Gašparović M, Kaplan O, Adjiski V, Comert R, Mobariz MA. Machine Learning-Based Classification of Asbestos-Containing Roofs Using Airborne RGB and Thermal Imagery. *Sustainability*. 2023 Mar;15(7):6067. <https://doi.org/10.3390/su15076067>

27. Magnani C, Mensi C, Binazzi A, Marsili D, Grosso F, Ramos-Bonilla JP, et al. The Italian Experience in the Development of Mesothelioma Registries: A Pathway for Other Countries to Address the Negative Legacy of Asbestos. *Int J Environ Res Public Health*. 2023 Jan;20(2):936. <https://doi.org/10.3390/ijerph20020936>

28. Mensi C, Riboldi L, De Matteis S, Bertazzi PA, Consonni D. Impact of an asbestos cement factory on mesothelioma incidence: Global assessment of effects of occupational, familial, and environmental exposure. *Environ Int*. 2015 Jan;74:191-9.

<https://doi.org/10.1016/j.envint.2014.10.016>

29. Luberto F, Ferrante D, Silvestri S, Angelini A, Cuccaro F, Nannavecchia AM, et al. Cumulative asbestos exposure and mortality from asbestos related diseases in a pooled analysis of 21 asbestos cement cohorts in Italy. *Environ Heal*. 2019 Dec;18(1):71.

<https://doi.org/10.1186/s12940-019-0510-6>

30. Martel E, Guerra R, Lopez S, Sarmiento R. A GPU-Based Processing Chain for Linearly Unmixing Hyperspectral Images. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2017 Mar;10(3):818-34.

<https://doi.org/10.1109/JSTARS.2016.2614842>

31. Toma AC, Panica S, Zaharie D, Petcu D. Computational challenges in processing large hyperspectral images. In: 2012 5th Romania Tier 2 Federation Grid, Cloud & High Performance Computing Science (RQLCG). 2012. p. 111-4.

32. Paoletti ME, Haut JM, Plaza J, Plaza A. Estudio Comparativo de Técnicas de Clasificación de Imágenes Hiperespectrales. *Rev Iberoam Automática e Informática Ind*. 2019 Mar;16(2):129.

<https://doi.org/10.4995/riai.2019.11078>

33. Mohan BK, Porwal A. Hyperspectral Image Processing and Analysis. *Curr Sci*. 2015;108(5):833-41. <https://www.citedi.mx/percepcionremota/portal/files/documents/HA17082404.pdf>

34. Ibarrola-Ulzurrun E, Marcello-Ruiz FJ, Gonzalo-Martín C. Evaluation of dimensionality reduction techniques in hyperspectral imagery and their application for the classification of terrestrial ecosystems. In: Bruzzone L, Bovolo F, Benediktsson JA, editors. *Image and Signal Processing for Remote Sensing XXIII*. SPIE; 2017. p. 17. <https://doi.org/10.1117/12.2278501>

35. Race BR, Wittman T. A comparison of dimensionality reduction techniques for hyperspectral imagery. In: Messinger DW, Velez-Reyes M, editors. *Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imaging XXVIII*. SPIE; 2022. p. 53.

<https://doi.org/10.1117/12.2632014>

36. Bravo Martínez G, Silva Aceves JM, Torres Argüelles SV, Enríquez Aguilera FJ. Implementación de Algoritmos de Procesamiento Digital de Señales en Hardware Paralelo: Artículo de revisión. *Cult Científica y Tecnológica*. 2018;(66):80-100.

<https://doi.org/10.20983/culcyt.2018.3.10>

37. Altamimi A, Ben Youssef B. A Systematic Review of Hardware-Accelerated Compression of Remotely Sensed Hyperspectral Images. *Sensors*. 2021 Dec 30;22(1):263.  
<https://doi.org/10.3390/s22010263>
38. Nalepa J. Recent Advances in Multi- and Hyperspectral Image Analysis. *Sensors*. 2021 Sep 8;21(18):6002. <https://doi.org/10.3390/s21186002>
39. Viel F, Parreira WD, Susin AA, Zeferino CA. A Hardware Accelerator for Onboard Spatial Resolution Enhancement of Hyperspectral Images. *IEEE Geosci Remote Sens Lett*. 2021 Oct;18(10):1796-800. <https://doi.org/10.1109/LGRS.2020.3009019>
40. Reena Thakur E al. A Comprehensive Analysis to Image Classification: Understanding Techniques and Explore Data Preprocessing a Non-linear Approach. *Adv Nonlinear Var Inequalities*. 2023 Jun;26(2):110-22. <https://doi.org/10.52783/anvi.v26.i2.287>
41. Martinez-Plumed F, Contreras-Ochando L, Ferri C, Hernandez-Orallo J, Kull M, Lachiche N, et al. CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Trans Knowl Data Eng*. 2021 Aug 1;33(8):3048-61.  
<https://doi.org/10.1109/TKDE.2019.2962680>
42. Schröer C, Kruse F, Gómez JM. A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Comput Sci*. 2021;181:526-34. <https://doi.org/10.1016/j.procs.2021.01.199>
43. Maataoui S, Bencheikh G, Bencheikh G. Predictive Maintenance in the Industrial Sector: A CRISP-DM Approach for Developing Accurate Machine Failure Prediction Models. In: 2023 Fifth International Conference on Advances in Computational Tools for Engineering Applications (ACTEA). IEEE; 2023. p. 223-7. <https://doi.org/10.1109/ACTEA58025.2023.10193983>
44. Nava J, Hernández P. Optimization of a Hybrid Methodology (CRISP-DM). In: Logistics Management and Optimization through Hybrid Artificial Intelligence Systems. IGI Global; 2012. p. 356-79. <https://doi.org/10.4018/978-1-4666-0297-7.ch014>
45. Lin C-H, Hsieh C-Y, Lin J-T. CODE-IF: A Convex/Deep Image Fusion Algorithm for Efficient Hyperspectral Super-Resolution. *IEEE Trans Geosci Remote Sens*. 2024;62:1-18.  
<https://doi.org/10.1109/TGRS.2024.3384808>
46. Halimu C, Kasem A, Newaz SHS. Empirical Comparison of Area under ROC curve (AUC) and Mathew Correlation Coefficient (MCC) for Evaluating Machine Learning Algorithms on Imbalanced Datasets for Binary Classification. In: Proceedings of the 3rd International Conference on Machine Learning and Soft Computing. New York, NY, USA: ACM; 2019. p. 1-6.  
<https://doi.org/10.1145/3310986.3311023>
47. Chanchí Golondrino GE, Ospina Alarcón MA, Saba M. Vegetation Identification in Hyperspectral Images Using Distance/Correlation Metrics. *Atmosphere (Basel)*. 2023 Jul 14;14(7):1148.  
<https://doi.org/10.3390/atmos14071148>
48. Saba M, Valdelamar Martínez D, Torres Gil LK, Chanchí Golondrino GE, Alarcón MAO. Application of Supervised Learning Methods and Information Gain Methods in the Determination of Asbestos-Cement Roofs' Deterioration State. *Appl Sci*. 2024 Sep 19;14(18):8441.  
<https://doi.org/10.3390/app14188441>