

# **Diseño de un algoritmo de obtención y procesamiento de imágenes hiperespectrales**

## **1 .Introducción**

EMIT 4 Future es un portal Web diseñado para mostrar de forma clara e intuitiva toda la información derivada del análisis hiperespectral de imágenes por satélite, con el fin de aportar información clave para el análisis y toma de decisiones respecto a la conservación de la biodiversidad.

El área de estudio inicial es la región floral del cabo, Sudáfrica, y los datos serán las imágenes tomadas por el sensor EMIT de la estación espacial internacional (como se verá más adelante). El objetivo de este artículo es complementar la Demo del proyecto con una implementación conceptual del algoritmo que se encargará de aportar la información que se mostraría gráficamente en el portal Emit 4 Future.

## **2. Marco teórico**

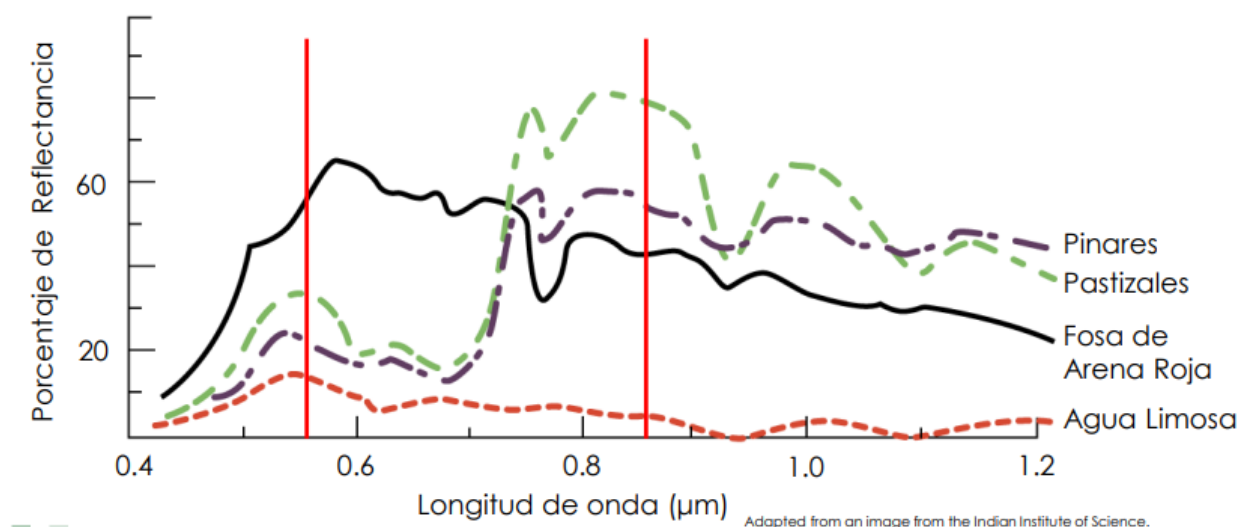
El análisis espectroscópico de imágenes es una rama muy amplia y compleja de la física. Con el objetivo de ilustrar el funcionamiento básico del algoritmo se ha preparado un marco teórico sobre las fórmulas y análisis que serán empleados más adelante en el proyecto.

### **2.1 Luz y longitudes de onda**

La luz blanca del sol que llega a la superficie terrestre está compuesta de multitud de ondas oscilando a diferentes frecuencias. De todas estas ondas las que nos interesan son los colores del espectro visible (desde el rojo hasta el violeta) y los rayos infrarrojos.

Esto es debido a que, cuando la luz llega a la superficie y rebota, algunas de estas longitudes de onda son absorbidas por distintos materiales. Sin entrar en el comportamiento que provoca este hecho, podemos utilizar la información de las

longitudes de onda que rebotan a la atmósfera para deducir contra qué material ha colisionado la luz.



**Figura 1.** Representación de la luz reflejada en función de la longitud de onda para diferentes materiales

De esta manera, identificando los patrones de reflectancia en las diferentes zonas de una imagen podemos identificar los materiales presentes en ella. Esto incluye desde depósitos minerales hasta vida vegetal y partículas en suspensión tanto en el aire como en el agua. Este patrón de reflectancia se conoce como la signatura espectroscópica de un material.

Esta reflectancia, sin embargo, puede verse afectada por la propia absorción de la atmósfera, que también absorbe algunas longitudes de onda. Sin embargo, esto no se ha tenido en cuenta porque, como se explicará en el punto 3, las muestras utilizadas ya han sido corregidas para evitar este problema.

### 3. Obtención de las muestras

Las muestras han sido obtenidas desde el portal EarthData Search de la nasa. Pertenecen al recién instalado sensor EMIT de la estación espacial internacional y corresponden a una zona conocida como al región floral del cabo, Sudáfrica.

Los datos son descargados mediante la API de Common Metadata Repository del Earth Observing System. Para ello es necesario crearse una cuenta en dicha organización y utilizar las funciones de petición de los datos.

En el caso de los datos del EMIT, como se ha comentado previamente, no se trata de datos brutos sino que ya han sido preprocesados eliminando las capas de nubes y compensando la deformación del espectro causada por la atmósfera.

```
curl -v -i https://cmr.earthdata.nasa.gov/stac/LPCLLOUD?page=2
```

## 4. Librerías

Para la implementación del algoritmo que se explicara posteriormente hemos utilizado las siguientes librerías y herramientas:

QGIS: es un Sistema de Información Geográfica profesional de fácil uso, gratis y de código abierto, que posibilita la creación, visualización, análisis, edición y publicación de información geoespacial.

GeoJSON: es un formato estándar abierto diseñado para representar elementos geográficos sencillos, junto con sus atributos no espaciales

AVHYAS: es un plugin de QGIS basado en Python diseñado para el proceso y análisis de imágenes hiperespectrales.

## 5. Procesamiento de imágenes

A continuación procedemos a explicar la estructura y especificación del algoritmo encargado de obtener y procesar los datos que serán desplegados en la web. El objetivo de este algoritmo no es solo obtener la información, sino transformarla a un formato más liviano y manejable por navegadores web. Este algoritmo está compuesto de dos partes: obtención de datos y procesamiento de datos.

### 5.1 Algoritmo de obtención y procesamiento de datos

Como se explica brevemente en el apartado 3, hemos utilizado la API de CMR con el método post para poder utilizar shapefiles con el fin de delimitar la zona de nuestro estudio. Así, si queremos estudiar por ejemplo el amazonas, solo tendríamos que proveer al algoritmo en el parámetro shape un shapefile compuesto por puntos o polígonos. Con la API obtendríamos una lista de enlaces los cuales guardaremos en una lista, lista que a su vez utilizaremos para saber que gránulos nos faltan por

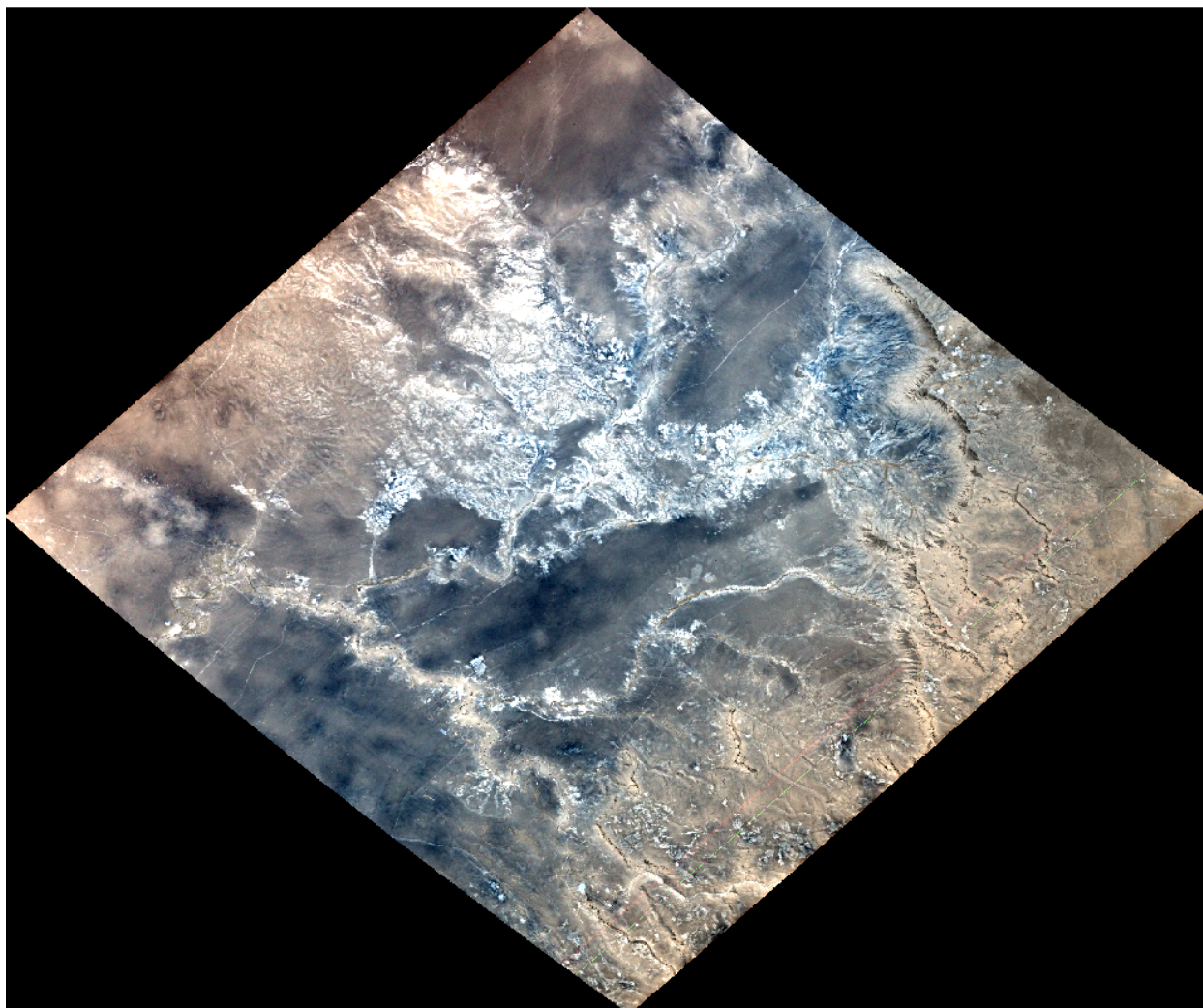
procesar. El algoritmo recorrerá la lista uno a uno, cogiendo cada enlace, descargando el granulo correspondiente y procesándolo con nuestro modelo de clasificación.

Además, existen dos archivos, pending y done. El archivo pending es un archivo pickle que tiene dentro la lista de los gránulos pendientes por descargar y procesar mientras que el archivo done no es mas que un archivo de texto donde se van escribiendo los gránulos procesados con éxito.

Este script da como resultado un numero de imágenes tiff procesadas y clasificadas correspondiente con el numero de gránulos.

## **5.2 Creación de un modelo IA**

Hemos utilizado QGIS, el cual se puede automatizar con python, para entrenar un modelo de inteligencia artificial, en concreto, un modelo de clasificación supervisada. El primer paso que se realiza es la transformación de la imagen NetCDF4 a ENVI, ya que no se puede usar NetCDF4 para entrenar este modelo IA.



**Figura 2.** ENVI de un gránulo

A este archivo ENVI se le pueden aplicar varias técnicas de preprocesado para reducir el ruido pero nosotros hemos decidido omitir este paso.

El siguiente paso es encontrar los endmembers. Para esto la extensión AVHYS nos proporciona varios métodos, nosotros hemos usado ATGP.

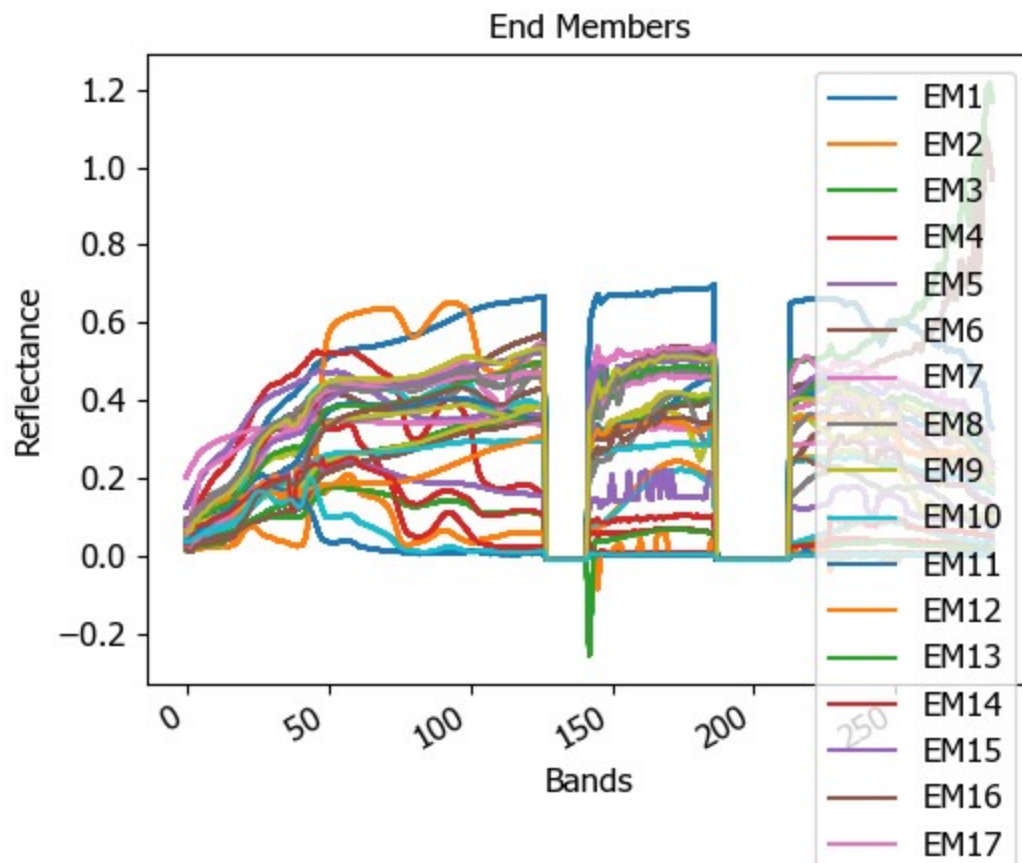
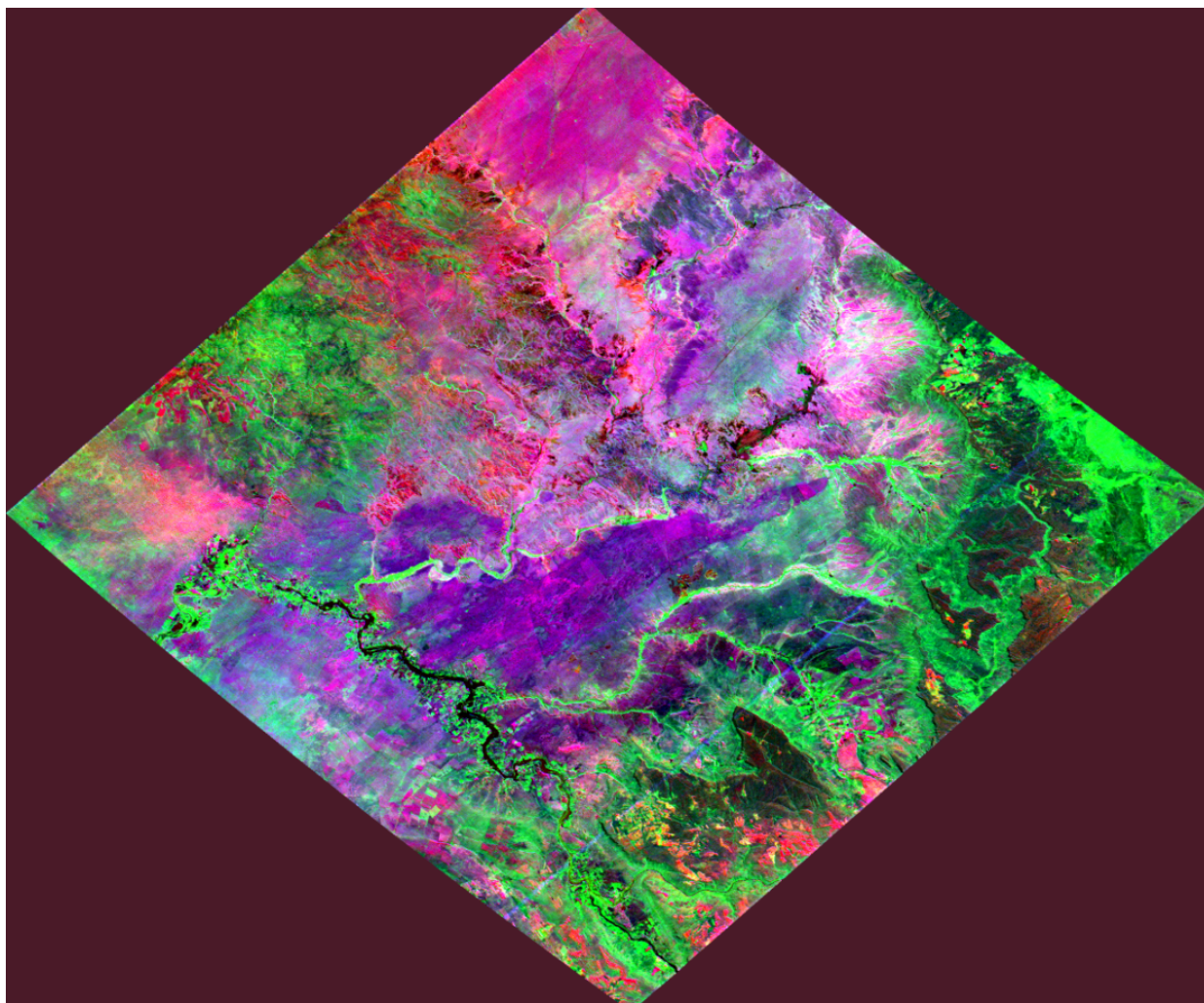


Figura 3. Endmembers de un gránulo

Con los endmembers podemos generar un abundance map el cual nos servira como base para crear un ground truth para el modelo IA.

Este abundance map no es muy certero como ground truth por lo que tenemos que corregirlo a mano.





**Figura 4.** Abundance map de un gránulo

Una vez obtenido un ground truth preciso podemos proceder a entrenar nuestro modelo.

La librería AVHYAS nos da varios modelos de clasificación, en este caso hemos usado Random Forest. Cuando aplicamos este método a nuestro raster obtenemos como resultado un raster clasificado mediante el modelo, modelo que se puede guardar en formato Pickle para su posterior uso en el algoritmo de obtención y procesamiento de datos.

## 6. Consideraciones

Dada la gran extensión en memoria de las imágenes del satélite y a las limitaciones físicas de nuestro equipo, el algoritmo expuesto debe ir eliminando las imágenes conforme las acaba de procesar para ahorrar recursos. Sin embargo en equipos más potentes puede obviarse esta parte del diseño.

Además también cabe comentar que podría diseñarse una versión paralela del algoritmo, de manera que varios nodos podrían descargar y procesar imágenes a la vez reduciendo en gran medida el coste temporal del proceso.

## **7. Conclusiones**

La espectrometría es una rama muy compleja, con una gran variedad de enfoques, sensores e imágenes diferentes a estudiar. En este estudio se ha presentado tan solo una versión reducida para ilustrar como se obtendrían los datos para cargarlos posteriormente en la demo de la página web.

En un futuro podrían ampliarse tanto los recursos como la profundidad del análisis para ofrecer una mayor cantidad de información de cada imagen, así como para mejorar la calidad de la visualización y precisión de la representación.