

Estudio Comparativo de Técnicas de Clasificación de Imágenes Hiperspectrales

M.E. Paoletti^{a*}, J.M. Haut, J. Plaza, A. Plaza

Laboratorio de Computación Hiperspectral. Departamento de Tecnología de Computadores y de las Comunicaciones.

Escuela Politécnica. Universidad de Extremadura. CP:10003. España.

Resumen

Las imágenes hiperspectrales constituyen el núcleo de varios programas de observación remota de la Tierra. La cantidad de información que contienen estas imágenes, formadas por cientos de canales espectrales estrechos y casi continuos, resulta de gran utilidad en aplicaciones en las que la caracterización de los materiales observados en la superficie terrestre resulta de gran relevancia. Esto se debe a la posibilidad de caracterizar de forma inequívoca cada material a través de su firma espectral. Algunas de estas aplicaciones son la agricultura de precisión, la planificación de espacios urbanos, o la prevención y seguimiento de desastres naturales. Sin embargo, la gran dimensión de las imágenes hiperspectrales supone un reto en su tratamiento, almacenamiento y procesamiento, debido a la gran variabilidad espectral y la correlación existente en los datos. En la literatura se han desarrollado múltiples algoritmos de análisis de imágenes hiperspectrales. En este artículo revisamos los algoritmos más utilizados para la clasificación de este tipo de imágenes, realizando experimentos con tres imágenes públicas y presentando una comparativa entre los métodos más ampliamente utilizados en este campo.

Palabras Clave:

Inteligencia Artificial, Aplicaciones Satelitales, Análisis de Escenas, Ingeniería Informática, Aplicaciones Informáticas

A Comparative Study of Techniques for Hyperspectral Image Classification

Abstract

Hyperspectral images are very important in many Earth Observation programs. The large amount of information is contained in hyperspectral images (hundreds of narrow and continuous spectral channels) is very useful for applications in which the characterization of the Earth surface materials is relevant. This is due to the fact that each observed element can be uniquely characterized by its spectral signature, for instance in precision agriculture, urban planning or detection/prevention of natural disasters, among others. However, the large dimensionality of hyperspectral images represents a challenge for analysis algorithms, both from the storage and processing viewpoints, resulting from data variability and correlation. Several algorithms have been proposed in the literature for the analysis of hyperspectral images. In this paper, we review the most popular techniques for hyperspectral classification. These techniques are inter-compared using three publicly available hyperspectral data sets.

Keywords:

Artificial Intelligence, Satellite applications, Scene Analysis, Computer Engineering, Computer Applications

1. Introducción

La observación remota de la Tierra (teledetección) resulta de gran importancia en diversos campos científicos, como la agricultura de precisión, la administración de recursos forestales, hídricos y minerales, el análisis atmosférico, la planificación de zonas urbanas y rurales, la detección y prevención de

riesgos y catástrofes, o la detección de zonas y objetivos de interés dentro del ámbito militar y de inteligencia, entre otras muchas aplicaciones (Lee et al., 2011). Estas aplicaciones han permitido un gran avance en el procesamiento de datos obtenidos de forma remota (Liang, 2008), tanto en términos de hardware para la adquisición de imágenes, como de software para su procesamiento. Esto ha permitido un desarrollo notable en

*Autor para correspondencia: mpaoletti@unex.es

la captación de grandes cantidades de información acerca de la superficie terrestre, caracterizadas por su gran variedad en términos de resolución espacial, temporal y espectral (Chang et al., 2016).

Destaca en particular el auge de la teledetección hiperespectral, un campo multidisciplinario dentro de la teledetección basada en la captación de espectros, en particular en la teledetección electro-óptica e infrarroja (Eismann, 2012). Los avances en el desarrollo de sensores hiperespectrales, llamados comúnmente espectrómetros (Vane et al., 1989) con alta resolución espectral y espacial, han permitido la adquisición simultánea de cientos de imágenes sobre una misma área de la superficie terrestre en rangos del espectro continuos y muy estrechos, creando cubos de datos de grandes dimensiones donde cada píxel proporciona una firma espectral muy detallada que recoge el comportamiento del material observado a la reflexión y absorción de la luz solar, algo que sensores tradicionales como los pancromáticos, RGB o multispectrales no pueden ofrecer (Chutia et al., 2016). En este contexto, destacan ciertos espectrómetros cuyas imágenes son comúnmente utilizados por la comunidad científica, por ejemplo el sensor AVIRIS (*Jet Propulsion Laboratory's Airborne Visible/Infrared Imaging Spectrometer*) (Green et al., 1998) de la NASA, que es capaz de recoger información en el rango espectral 400nm-2500nm, adquiriendo cubos de datos con un total 224 bandas contiguas con un ancho nominal de 10nm, a una velocidad de 2.5 MB/s (casi 9 GB/hora), o el espectrómetro ubicado en el satélite EO-1 (*Earth Observing-1*), conocido como Hyperion (Plaza et al., 2011; Pour and Hashim, 2014), también de la NASA, que recoge información del infrarrojo visible (400nm-1000nm) y del infrarrojo de onda corta (900nm-2500nm) utilizando 242 bandas con un ratio de adquisición de casi 71.9 GB/hora (cerca de 1.6 TB/día).

Los cubos de datos que estos sensores captan forman lo que conocemos como imágenes hiperespectrales (Chang, 2003), caracterizadas por sus grandes dimensiones, que resulta muy útil para ciertas aplicaciones de teledetección. Por ejemplo, se pueden encontrar aplicaciones de la imágenes hiperespectrales en actividades como el análisis de cultivos para la detección de plagas y enfermedades (Bannari et al., 2006), la gestión de recursos forestales (Lawrence et al., 2006) o la explotación de recursos minerales (Kokaly et al., 2016), además de las referidas al campo militar, de seguridad-vigilancia (Banerjee et al., 2006) o de inteligencia gubernamental (Lee et al., 2011).

Sin embargo, el uso de estas imágenes presenta ciertas dificultades debido a sus grandes requerimientos en cuanto a almacenamiento. En particular, los cubos hiperespectrales ocupan mucho espacio en memoria. Además, su procesamiento es complicado ya que, debido a factores atmosféricos o propios del sensor, los datos pueden presentar una gran variabilidad. En este tipo de imágenes cobra especial importancia el *fenómeno de Hughes*, por el cual la precisión de los resultados obtenidos por los métodos de análisis y clasificación, a pesar de que aumenta inicialmente con el incremento del número de bandas espectrales, cae drásticamente cuando el número de bandas aumenta demasiado (Ma et al., 2013).

Hasta la fecha, se han desarrollado múltiples métodos de análisis de imágenes hiperespectrales, enmarcados en el campo del *machine learning* (Zhu and Basir, 2005), cubriendo un

amplio rango de objetivos, tales como desmezclado espectral (*unmixing*), recuperación de datos (*data recovery*), detección de objetos (*target/object detection*) o clasificación (*land-cover classification*). En este contexto, el presente artículo se centra en las técnicas de clasificación de imágenes hiperespectrales. En la literatura disponible, se pueden encontrar múltiples técnicas de clasificación como por ejemplo técnicas de soft computing (el ejemplo más claro son las redes neuronales artificiales (Atkinson and Tatnall, 1997), técnicas basadas en kernels (como la support vector machine -SVM- (Melgani and Bruzzone, 2004), el K-nearest neighbor -KNN- (Silverman and Jones, 1989)) o técnicas basadas en sistemas de lógica borrosa o sistemas de conjuntos difusos (Wang, 1990)), técnicas de data fusion (fusión a nivel de píxeles o fusión a nivel de decisión), técnicas de ensemble (como el random forest -RF- (Ham et al., 2005)) y técnicas con aproximación estadística (Serpico et al., 1996) (como el maximum likelihood (Paola and Schowengerdt, 1995) o multinomial logistic regression (Li et al., 2010)), entre otras. Todas ellas dan respuesta al mismo problema, analizándolo desde distintas perspectivas y empleando la información espectral contenida en los píxeles de las imágenes hiperespectrales de diversas formas. En este sentido, el objetivo de este trabajo es revisar los principales métodos de clasificación hiperespectral basados en el análisis de la firma espectral (*pixel-based o spectral-based*), con el objetivo de aportar un banco de pruebas sobre imágenes hiperespectrales públicas ampliamente utilizados por la comunidad científica, permitiendo un estudio comparativo.

El artículo se organiza de la siguiente manera. La Sección 2 ofrece una revisión de cinco métodos espectrales ampliamente utilizados en la clasificación de este tipo de imágenes. La Sección 3 define un entorno experimental y discute el rendimiento y los resultados de clasificación obtenidos por cada método seleccionado sobre tres imágenes hiperespectrales muy conocidas: Indian Pines (IP), University of Pavia (UP) y Salinas Valley (SV). Finalmente, la Sección 4 concluye el artículo, resumiendo el contenido del mismo e indicando las principales observaciones resultantes del estudio comparativo, además de analizar las tendencias que se siguen actualmente en la clasificación hiperespectral, proponiendo nuevas líneas y trabajos futuros.

2. Métodos

Dentro de la amplia variedad de técnicas para el tratamiento de imágenes hiperespectrales, las basadas en inteligencia artificial (*machine learning*) han experimentado un importante auge su gran capacidad para extraer información útil de datos complejos, como es el caso de este tipo de imágenes. Dependiendo de cómo los algoritmos emplean las muestras etiquetadas (que forman la verdad terreno del dataset, *ground-truth* o GT) en su procesamiento se pueden agrupar en: i) algoritmos de clasificación supervisados, aquellos que presentan una fase de entrenamiento con un porcentaje de muestras etiquetadas pertenecientes a las distintas clases contenidas en el ground-truth con el que el algoritmo ajusta sus parámetros, ii) no supervisados, aquellos que no necesitan una fase de entrenamiento con datos etiquetados para encontrar la estructura interna de los datos con la que poder agruparlos en distintos grupos, como por ejemplo los basados en clustering, o iii) híbridos, también conocidos como se-

misupervisados, en los que se emplea un pequeño subconjunto de muestras etiquetadas frente a un subconjunto mayor de datos sin etiquetar para desarrollar la fase de entrenamiento y ajuste de parámetros del algoritmo.

Dentro del análisis hiperespectral, los algoritmos no supervisados tienen un especial interés en ciertos ámbitos como el desmezclado espectral, consistente en la obtención de los elementos básicos que componen la firma espectral recogida en un píxel (Chiang et al., 2000), la reducción de la dimensión espectral, por ejemplo para obtener un procesamiento más rápido y eficiente (Haut et al., 2018c) o para la selección de las bandas espectrales más discriminantes (Rodarmel and Shan, 2002; Yang et al., 2011), y como un pre-procesado de los datos para la extracción de características (He et al., 2018) que permitan a los clasificadores espectrales alcanzar un mejor desempeño. En clasificación, se pueden encontrar algunos ejemplos de algoritmos no supervisados basados en clustering en trabajos recientes como (Haut et al., 2017), aunque se puede observar una mayor variedad en la segmentación de imágenes (Theiler and Gisler, 1997; Ifarraguerri and Chang, 2000; Acito et al., 2003).

En cuanto a los algoritmos semisupervisados, éstos suponen un tipo de aprendizaje muy interesante ya que permiten obtener un modelo muy robusto con muy pocas muestras de entrenamiento etiquetadas, lo que resulta muy atractivo para el análisis de imágenes hiperespectrales, cuyos datasets normalmente contienen poca información de ground-truth (en el sentido de pocas muestras etiquetadas de antemano) debido al alto coste de categorizarlas mediante un experto (Chi and Bruzzone, 2007; Camps-Valls et al., 2007; Li et al., 2010; Haut et al., 2018b).

Finalmente, los algoritmos supervisados son los más ampliamente utilizados en las tareas de clasificación en general, debido a su gran rendimiento al inferir el modelo mediante un conjunto aceptable de muestras etiquetadas. Precisamente por su funcionamiento y por el grado de precisión que son capaces de alcanzar, suponen una herramienta potente para el análisis de imágenes hiperespectrales. En este trabajo revisamos los principales algoritmos de aprendizaje supervisado, discutiendo su aplicación al análisis de imágenes hiperespectrales. En particular se comparan los siguientes métodos: i) Support Vector Machines (SVM), ii) Multinomial Logistic Regression (MLR), iii) Perceptrón multicapa (*multilayer perceptron*, MLP) y iv) Red convolucional espectral (*spectral convolutional neural network*, CNN).

2.1. Support Vector Machines

Dentro de la familia de métodos basados en kernels (Schölkopf et al., 2002), la SVM ha sido popularmente estudiada y ampliamente utilizada en el campo del análisis hiperespectral debido a su gran capacidad para generar un modelo robusto con pocas muestras de entrenamiento, en especial cuando el número de dimensiones es mayor que la cantidad de datos (como es el caso de las imágenes hiperespectrales) (Mercier and Lennon, 2003; Melgani and Bruzzone, 2004; Pal and Mather, 2004; Banerjee et al., 2006; Mountrakis et al., 2011).

Desde un punto de vista geométrico, el objetivo principal de este clasificador es definir aquel hiperplano que divida las muestras en dos zonas separadas (se trata de un típico problema de clasificación binaria), maximizando los bordes marginales

existente entre las muestras y el hiperplano, es decir maximizando la distancia de las muestras más cercanas a la zona de decisión. Para ello, utiliza las muestras más discriminantes que más puedan afectar en el cálculo del hiperplano y que precisamente se encuentran más cerca de la zona de decisión (los llamados *support vectors*). Además de intuitivo, el clasificador SVM resulta muy flexible al soportar diferentes tipos de kernels para el cálculo del hiperplano óptimo (Mercier and Lennon, 2003) (Tabla 1). Así pues, en función del kernel empleado, la zona marginal puede ser construida mediante rectas, curvas, etc.

Tabla 1: Ejemplos de kernels más usados por el método SVM, donde \mathbf{x}_i y \mathbf{x}_j son dos muestras del dataset \mathbf{X}

Lineal	$\langle \mathbf{x}_i, \mathbf{x}_j \rangle$
Polinómico	$(\gamma \langle \mathbf{x}_i, \mathbf{x}_j \rangle + c)^n$
Gausiano	$\exp(-\gamma \ \mathbf{x}_i - \mathbf{x}_j\ ^2) \gamma$
Sigmoide	$\tanh(\gamma \langle \mathbf{x}_i, \mathbf{x}_j \rangle + c)^n$
Radial basis function (RBF)	$\exp\left(-\frac{\ \mathbf{x}_i - \mathbf{x}_j\ ^2}{2\sigma^2}\right)$

Sin embargo, la SVM debe hacer frente a la necesidad de ajustar de manera empírica los parámetros que componen el kernel para conseguir una solución óptima, por lo que los resultados finales dependen mucho del diseño del kernel. Además su coste computacional y de memoria es elevado cuando el problema a clasificar es complejo o presenta unas grandes dimensiones.

2.2. Random Forest

El algoritmo RF pertenece a la familia de los *ensemble methods* y se basa en una colección de árboles de clasificación, también llamados *decision trees classifiers* (DTs). Los DTs son métodos supervisados de aproximación no paramétricos que pueden usarse tanto para regresión como para clasificación de datos (Benediktsson and Ghamisi, 2015). En particular, se trata de un método recursivo, esquemático y en forma de árbol invertido. Esta estructura es utilizada para determinar el curso de una acción o para mostrar una probabilidad estadística siguiendo una estrategia de ramificación basada en decisiones de *divide y vencerás* que construye el árbol de arriba a abajo (Dixit and Bhatia, 2013).

El funcionamiento del RF es relativamente sencilla. Por cada DT implementado, durante la ejecución de cada DT el conjunto de datos de entrada se evalúa sucesivamente en cada uno de los nodos que componen el DT y se divide en subconjuntos más pequeños y homogéneos (que se basan en el divisor/diferenciador más significativo de las variables de entrada), enviándolos a las diferentes ramas del DT hasta que llegan a la parte inferior o nodo hoja, el cual proporciona la clase a la que pertenecen los datos. Finalmente, el RF recoge todas las salidas obtenidas y mediante *majority voting* obtiene la clase final a la que pertenecen los datos.

Su uso en el análisis de imágenes hiperespectrales es amplio (Ham et al., 2005; Lawrence et al., 2006; Chan and Paelinckx, 2008; Naidoo et al., 2012; Belgiu and Drăguț, 2016), sin embargo, a pesar de ser un método muy preciso cuando trabaja sobre conjuntos de datos complejos y de resultar simple a la hora de hacer la inferencia, la forma de expandir las ramas de cada DT hace que el RF sea computacionalmente lento de crear y costoso a la hora de ajustar todos los parámetros.

2.3. Multinomial Logistic Regression

Uno de los métodos supervisados estadísticos más utilizados en el aprendizaje automático es la regresión logística (*logistic regression*, LR), aplicada a la clasificación binaria, y su correspondiente extensión a datos multi-clase, la MLR.

Dado un dataset \mathbf{X} compuesto por N muestras, donde cada una es un vector de M características, $\mathbf{x}_i \in \mathbb{R}^M$, y existen K categorías distintas, el algoritmo LR (y, por lo tanto su extensión MLR) funciona como un clasificador lineal basado en la idea de *logits* o puntuaciones que funcionan como la probabilidad de que el dato de entrada \mathbf{x}_i pertenezca a la k -ésima categoría. Estas puntuaciones, indicadas en la Ecuación (1), son expresadas en términos de coeficientes de regresión, calculándose uno por cada clase considerada, β_k :

$$\text{score}(\mathbf{x}_i, k) = \beta_k \mathbf{x}_i = \beta_{k,1}x_{i,1} + \beta_{k,2}x_{i,2} + \dots + \beta_{k,M}x_{i,M} \quad (1)$$

Sabiendo que la probabilidad de que la etiqueta y_i del dato \mathbf{x}_i pertenezca a la k -ésima clase, $y_i = k$, puede expresarse mediante la Softmax Regression como:

$$P(y_i = k | \mathbf{x}_i, \beta_k) = \frac{\exp(\beta_k \mathbf{x}_i)}{\sum_{j=1}^K \exp(\beta_j \mathbf{x}_i)} \quad (2)$$

La MLR tiene como objetivo encontrar aquellos coeficientes β_k (organizados en la matriz de coeficientes \mathbf{B}) que minimicen la función de coste $f(\cdot)$ indicada por la Ecuación (3):

$$f(\mathbf{B}; \mathbf{X}, \mathbf{Y}) = -\frac{1}{N} \left[\sum_{i=1}^N \sum_{k=1}^K 1_{\{y_i = k\}} \log \frac{\exp(\beta_k \mathbf{x}_i)}{\sum_{j=1}^K \exp(\beta_j \mathbf{x}_i)} \right] \quad (3)$$

donde \mathbf{X} e \mathbf{Y} son los conjuntos de muestras y sus correspondientes categorías tal que $\{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_i, y_i\}_{i=1}^N$ donde $y_i = 1, 2, \dots, K$, siendo λ un factor de regularización para mitigar el sobreajuste.

La solución final puede ser encontrada utilizando un proceso iterativo a través de algoritmos de optimización basados en el gradiente descendente, como por ejemplo el Broyden-Fletcher-Goldfarb-Shanno con limitación de memoria (L-BFGS) (Nocedal, 1980; Haut et al., 2016), un método quasi-newton de minimización multivariable que es idóneo para el tratamiento de grandes volúmenes de datos, lo que lo hace un candidato ideal para imágenes hiperespectrales (Khodadadzadeh et al., 2014).

2.4. Redes neuronales: del perceptrón multicapa a la red convolucional

Las redes neuronales artificiales (*artificial neural networks*, ANNs) han sido estudiadas en profundidad y ampliamente utilizadas desde la aparición del algoritmo de retropropagación (*backpropagation*) (Rumelhart et al., 1986), por el que los métodos de optimización propagan la señal del gradiente a través de las neuronas ubicadas en las capas que componen la red.

El modelo más simple de ANN es el perceptrón multicapa (*multilayer perceptron*, MLP), compuesto por un conjunto de nodos o neuronas agrupadas en capas completamente interconectadas con las capas inmediatamente anterior y posterior. Para hacer sencilla la explicación, denotaremos todos los elementos relacionados con las capas de una red con el superíndice (l), indicando que son elementos (datos de entrada, pesos,

número de neuronas, sesgos...) pertenecientes a la l -ésima capa de la red.

Con las observaciones anteriores, y contemplando la l -ésima capa del MLP, compuesta por $n^{(l)}$ neuronas, es posible calcular su vector salida $\mathbf{x}_i^{(l)} \in \mathbb{R}^{n^{(l)}}$ para el originalmente insertado en la red \mathbf{x}_i como la transformación afín entre el vector de datos de entrada de la capa (proveniente de la capa inmediatamente anterior) $\mathbf{x}_i^{(l-1)} \in \mathbb{R}^{n^{(l-1)}}$ y el conjunto de pesos $\mathbf{W}^{(l)} \in \mathbb{R}^{n^{(l)} \times n^{(l-1)}}$ (donde cada peso asocia las salidas de la capa anterior con las neuronas de la capa actual) y la señal de sesgo (también llamada *bias*) $b^{(l)}$ pasadas por una función de activación $\mathcal{H}(\cdot)$:

$$\mathbf{x}_i^{(l)} = \mathcal{H}(\mathbf{W}^{(l)} \cdot \mathbf{x}_i^{(l-1)} + b^{(l)}) \quad (4)$$

La Ecuación (4) puede ser entendida a nivel de neurona como el producto lineal entre los datos de entrada de la capa y el peso asociado a la conexión sináptica entre cada una de las neuronas de la l -énima capa y las neuronas de la capa anterior, que es pasado por la función de activación (normalmente no lineal, por ejemplo la ampliamente utilizada *rectified linear unit* o ReLU, y sus derivados como la Leaky-ReLU, la *scaled exponential linear unit* o SELU, entre otras) que captura las relaciones y estructuras no-lineales de los datos.

Las redes neuronales convolucionales (*convolutional neural networks*, CNNs) siguen el mismo principio de aplicar unos pesos y sesgos a los datos de entrada y pasarlos por una función de activación para obtener la representación del dato en el espacio de características de cada capa. Sin embargo, a diferencia de las redes neuronales convencionales, las capas convolucionales operan a nivel de filtros, donde los pesos están localmente conectados a pequeñas zonas de los datos de entrada, definidas por un campo receptivo, y son deslizados por éstos emulando el comportamiento de los algoritmos de ventana. Esto hace posible la reducción del número de conexiones y el procesamiento grupal de la información, lo que añade robustez al método.

Sin embargo, las redes neuronales en general presentan dos inconvenientes principales: el primero está relacionado con el número de parámetros, normalmente muy elevado, lo que las convierte en computacionalmente complejas. Por otra parte, un mal diseño de la arquitectura neuronal hace que el algoritmo de optimización tenga tendencia a quedar atascado en mínimos locales que impidan llegar a una solución óptima. Afortunadamente, en los últimos años (gracias al auge de las técnicas de *deep learning*), la comunidad científica ha hecho un gran esfuerzo en desarrollar no solo nuevas técnicas que permitan un mejor análisis de los datos, sino además en implementar nuevas mejoras que minimizan el impacto de los inconvenientes anteriormente señalados, solventando los problemas de mínimos locales (Kingma and Ba, 2014) y acelerando la computación (Paoletti et al., 2018; Haut et al., 2018a).

Por otra parte, el objetivo de cualquier clasificador neuronal es minimizar la función objetivo definida por la diferencia entre las etiquetas originales del conjunto de datos \mathbf{Y} y las etiquetas obtenidas al final de la red \mathbf{Y}' . Existen varias maneras de reflejar este objetivo, sin embargo, se ha demostrado, y así lo refleja su amplio uso, que la entropía cruzada de la señal de salida es una de las mejores, ver Ecuación (5). Lo que se busca con su aplicación al final de la red es que el método de optimización ajuste los parámetros del modelo teniendo en cuenta el ruido

de la señal de salida, lo que añade robustez al modelo de red neuronal:

$$\phi_c = - \sum_i y'_i - \log(y_i) \quad (5)$$

Todo ello hace que las ANNs sean una buena opción para procesar datos espectrales.

3. Resultados Experimentales

3.1. Configuración del entorno de pruebas

Para evaluar los distintos clasificadores anteriormente mencionados se ha usado un entorno hardware compuesto por un procesador Intel® Core™i7-6700K de sexta generación con 8M de Cache a 4.20GHz (4 cores/8 way multitask processing), 32GB de DDR4 RAM con velocidad serial de 2400MHz, una GPU NVIDIA GeForce GTX 1080 con 8GB GDDR5X de memoria de vídeo y 10Gbps de memoria, un Toshiba DT01ACA HDD a 7200RPM con 2TB de capacidad, y una placa base ASUS Z170 pro-gaming. Además, el entorno software se compone del sistema operativo Ubuntu 18.04 x64, la librería CUDA 9 y el lenguaje de programación Python versión 2.7.

3.2. Dataset Hiperespectrales

Los algoritmos de clasificación espectral han sido evaluados sobre tres imágenes hiperespectrales públicas y popularmente utilizadas dentro de la comunidad científica con objeto de hacer una comparativa funcional. Dichas imágenes son: Indian Pines (IP), University of Pavia (UP) y Salinas Valley (SV). A continuación se listan sus principales características:

- La imagen **Indian Pines** (IP) fue capturada por el sensor AVIRIS (Green et al., 1998) en 1992 sobre una zona agrícola, con parcelas rectangulares de distintos cultivos y ciertas zonas irregulares de arboledas en el noroeste de Indiana. La escena contiene 145×145 píxeles con 224 bandas espectrales que recogen la variación de la absorción solar en el rango de los 400 a los 2500nm, con 10nm de resolución espectral, 20nm de resolución espacial moderada y 16 bits de resolución radiométrica. De las bandas espectrales originales, 4 bandas con valor cero y 20 bandas ruidosas fueron eliminadas, quedando sólo 200 bandas espectrales. El dataset tiene en total 16 clases (Tabla 2).
- La imagen **University of Pavia** (UP) fue obtenida por el Reflective Optics System Imaging Spectrometer (ROSIS) (Kunkel et al., 1988) durante una campaña de vuelo sobre la ciudad italiana de Pavia, concretamente sobre parte del campus universitario. La imagen cubre un entorno urbano con varias estructuras sólidas, elementos naturales y sombras. Su ground-truth cuenta con 9 clases distintas en total (Tabla 2). La imagen contiene 610×340 píxeles con 103 bandas en el rango de los 430-860nm y resolución espacial de 1.3 metros por píxel.
- La última imagen considerada es **Salinas Valley** (SV) y también fue capturada por el sensor AVIRIS sobre el valle californiano de Salinas (Tabla 2). El área cubierta cuenta con 512×217 píxeles y 224 bandas espectrales.

La resolución espacial es de 3.7 metros por píxel. También se ha eliminado 20 bandas debido al efecto de la absorción del agua, quedando 204 bandas espectrales en total. La escena cuenta con un ground-truth de 16 clases (Tabla 2).

3.3. Experimentación

Para evaluar y validar los diferentes métodos propuestos se han llevado a cabo dos experimentos distintos sobre los conjuntos de datos señalados, tomando los datos de entrenamiento de forma balanceada. En dicho entrenamiento con datos balanceados, el conjunto de capacitación se forma con un número de muestras por cada clase. En este sentido, si el número de datos disponibles que hay en una clase es menor que el número de muestras por clase seleccionadas para el entrenamiento, se toman todas las muestras originales de la clase. El primer experimento consiste en una comparativa entre los métodos propuestos y ampliamente utilizados en la literatura del tratamiento de imágenes hiperespectrales, empleando como datos de entrenamiento un número fijo de muestras por cada clase. El segundo experimento se centra en el comportamiento de los métodos a medida que el número de muestras de entrenamiento varía. En ambos experimentos, la precisión alcanzada por los algoritmos se han medido en términos de precisión global ó Overall Accuracy (OA), precisión media ó Average Accuracy (AA) y coeficiente Kappa.

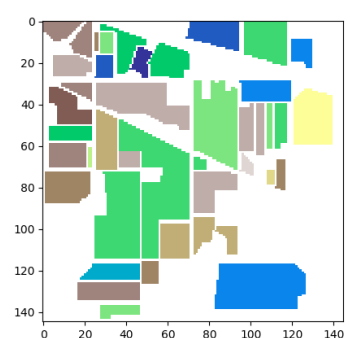
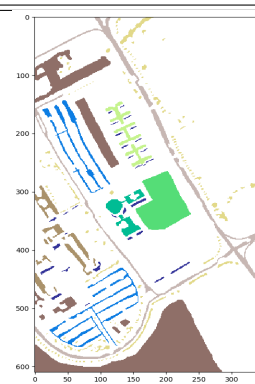
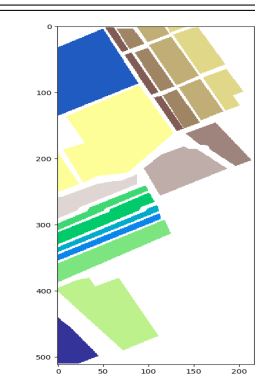












































3.3.1. Experimento 1: Número de muestras fijo

Para este experimento se han entrenado los clasificadores RF, MLR, SVM, MLP y CNN con 100 muestras por clase, empleándose el resto de datos para hacer la inferencia en la fase de test. Estos clasificadores se pueden clasificar en lineales (RF y MLR) y no-lineales (SVM, MLP y CNN). También pueden ser clasificados como: métodos de *ensemble* (RF), probabilísticos (MLR), basados en kernels (SVM) y en redes neuronales (MLP y CNN).

Los resultados obtenidos como la media de 10 ejecuciones Monte Carlo, en términos de OA, AA y Kappa pueden consultarse en la Tabla 3. Además las Figuras 1-3 ofrecen los mapas de clasificación de cada uno de los métodos. Como se puede observar, tanto en la Tabla como en las mapas de clasificación, los métodos lineales (RF y MLR) no tienen un rendimiento alto en comparación con sus rivales no lineales. Este hecho confirma que la composición de elementos que conforman la firma espectral del píxel sigue patrones no lineales. Aunque la precisión de estos métodos para imágenes sencillas, como puede ser SV donde la firma espectral es bastante limpia (ver Figura 3), es suficiente para ayudar a la toma de decisiones en tareas de clasificación, en otros casos ligeramente más complejos como UP (Figura 2) vemos claramente como sufren una gran degradación, haciéndose más patente en el caso de IP (Figura 1), donde las firmas espectrales se encuentran muy mezcladas y existe una gran variabilidad intra-clases.

Por otra parte, de los métodos no lineales también se puede observar que, para datasets más sencillos como SV y UP, el método basado en kernels de la SVM es capaz de alcanzar mejor OA que aquellos basados en ANNs. Sin embargo, ante los píxeles más ruidosos de IP, la SVM encuentra ciertas dificultades para separar las muestras mediante el cálculo de hiper-

Tabla 2: Distribución de muestras por clase de las imágenes hiperespectrales consideradas en el estudio.

INDIAN PINES			PAVIA UNIVERSITY			SALINAS		
								
Color	Tipo	Muestras	Color	Tipo	Muestras	Color	Tipo	Muestras
	Background	10776		Background	164624		Background	56975
	Alfalfa	46		Asphalt	6631		Brocoli-green-weeds-1	2009
	Corn-notill	1428		Meadows	18649		Brocoli-green-weeds-2	3726
	Corn-min	830		Gravel	2099		Fallow	1976
	Corn	237		Trees	3064		Fallow-rough-plow	1394
	Grass/Pasture	483		Painted metal sheets	1345		Fallow-smooth	2678
	Grass/Trees	730		Bare Soil	5029		Stubble	3959
	Grass/pasture-mowed	28		Bitumen	1330		Celery	3579
	Hay-windrowed	478		Self-Blocking Bricks	3682		Grapes-untrained	11271
	Oats	20		Shadows	947		Soil-vinyard-develop	6203
	Soybeans-notill	972					Corn-senesced-green-weeds	3278
	Soybeans-min	2455					Lettuce-romaine-4wk	1068
	Soybean-clean	593					Lettuce-romaine-5wk	1927
	Wheat	205					Lettuce-romaine-6wk	916
	Woods	1265					Lettuce-romaine-7wk	1070
	Bldg-Grass-Tree-Drives	386					Vinyard-untrained	7268
	Stone-steel towers	93					Vinyard-vertical-trellis	1807
	Total de muestras	21025		Total de muestras	207400		Total de muestras	111104

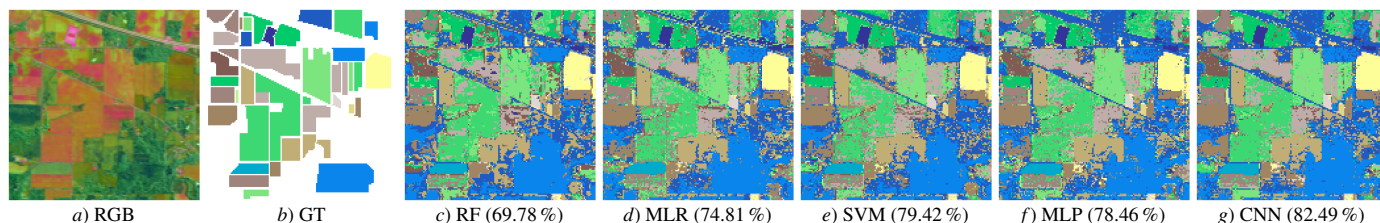


Figura 1: Mapas de clasificación para el dataset hiperespectral *Indian Pines* (IP). El primer mapa (a) representa una imagen sintética en RGB. El segundo (b) contiene la verdad terreno de los datos. Finalmente, el resto de las imágenes (c) a (g) muestran los mapas de clasificación correspondientes a la Tabla 3. El porcentaje de acierto ó Overall Accuracy (OA) se muestra entre paréntesis.

planos, siendo las ANNs las que mejor precisión alcanzan, en particular la CNN. En cuanto a la complejidad computacional, se puede observar que la SVM es el clasificar más rápido de todos, seguida del RF y la MLR, siendo los basados en ANNs los más altos, debido al ajuste de parámetros que hacen en tiempo de entrenamiento. Sin embargo, el tiempo de búsqueda de los hiperparámetros de la SVM, el RF y la MLR óptimos para resolver el problema de clasificación, son muy grandes y su ajuste resulta complicado, necesiándose una búsqueda intensiva en un rango de valores hasta encontrar aquel con el que mejor prestación de el método.

De este experimento se puede concluir que, cuando se tienen mezclas no lineales en los datos, los modelos más precisos y robustos son aquellos diseñados expresamente para explotar esas relaciones no lineales contenidas en los datos. En este sentido la SVM-RBF, el perceptrón multicapa (MLP) y la red con-

volucional (CNN) se postulan como modelos precisos para el procesado eficaz de imágenes hiperespectrales. Sin embargo, si además de existir relaciones no lineales, los datos presentan cierta complejidad (como presencia de ruido o gran variabilidad intra-clases), son los métodos basados en ANNs los que pueden extraer con mayor precisión las características más discriminantes de las imágenes, alcanzando una mayor precisión.

3.3.2. Experimento 2: Número de muestras variable

En el segundo experimento se han entrenado los clasificadores RF, MLR, SVM, MLP y CNN con 50, 100 y 150 muestras por clase, obteniéndose la media y la desviación estándar de 10 ejecuciones Monte Carlo. La Figura 4 muestra las gráficas con la evolución del OA para IP, UP y SV.

En general, la disminución de muestras de entrenamiento afecta negativamente a todos los clasificadores espectrales,

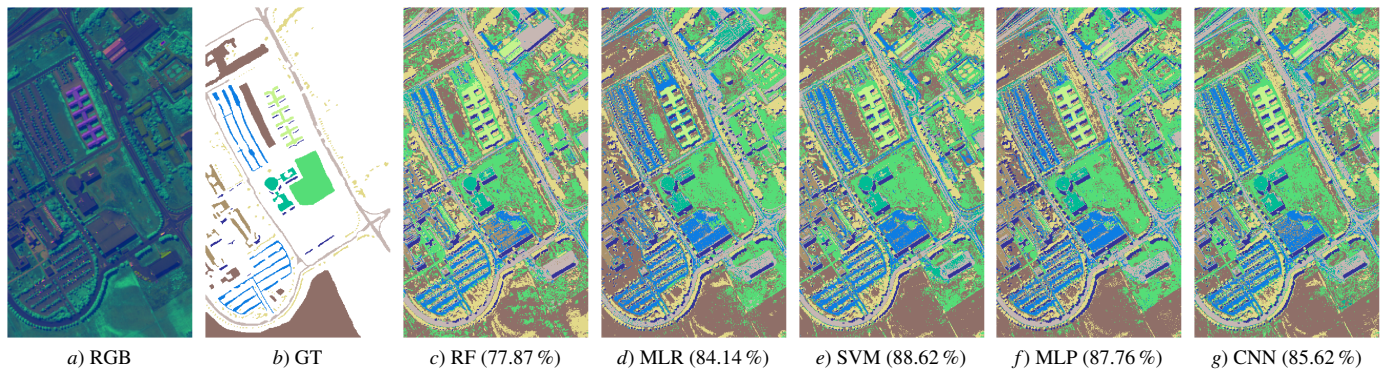


Figura 2: Mapas de clasificación para el dataset hiperspectral *Pavia University* (PU). El primer mapa (a) representa una imagen sintética en RGB. El segundo (b) contiene la verdad terreno de los datos. Finalmente, el resto de las imágenes (c) a (g) muestran los mapas de clasificación correspondientes a la Tabla3. El porcentaje de acierto ó Overall Accuracy (OA) se muestra entre paréntesis

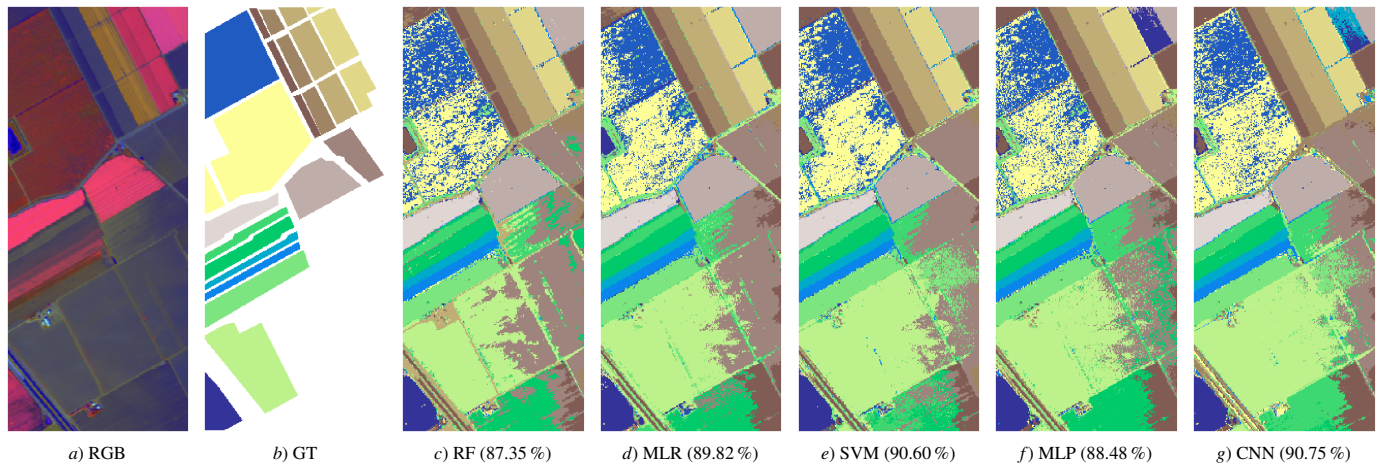


Figura 3: Mapas de clasificación para el dataset hiperspectral *Salinas Valley* (SV). El primer mapa (a) representa una imagen sintética en RGB. El segundo (b) contiene la verdad terreno de los datos. Finalmente, el resto de las imágenes (c) a (g) muestran los mapas de clasificación correspondientes a la Tabla3. El porcentaje de acierto ó Overall Accuracy (OA) se muestra entre paréntesis.

Tabla 3: Resultados de clasificación utilizando 100 muestras por clase sobre tres datasets hiperspectrales reales.

Clase	INDIAN PINES					PAVIA UNIVERSITY					SALINAS VALLEY				
	RF	MLR	SVM	MLP	CNN	RF	MLR	SVM	MLP	CNN	RF	MLR	SVM	MLP	CNN
1	65.85	71.43	76.60	75.49	80.04	71.70	79.47	85.06	83.99	81.31	85.92	88.66	89.52	87.19	89.70
2	99.13	100.00	100.00	100.00	100.00	75.82	80.02	83.49	83.26	83.65	99.39	98.92	99.41	99.25	99.54
3	46.70	71.78	75.72	72.68	78.45	73.81	81.97	89.57	88.66	83.95	99.28	99.71	99.59	99.58	99.69
4	59.65	59.48	74.77	70.41	77.29	76.38	83.54	84.78	85.77	83.54	97.25	99.27	99.57	98.53	99.52
5	81.46	85.69	87.44	83.50	88.61	95.12	91.25	95.05	94.76	94.51	99.46	99.32	99.42	99.49	99.32
6	87.10	90.60	93.05	90.60	92.01	99.28	99.39	99.42	99.50	99.55	98.03	98.12	98.37	98.72	98.87
7	90.03	94.25	95.78	95.87	97.05	74.39	87.42	87.20	89.40	85.03	99.48	99.73	99.84	99.75	99.77
8	100.00	100.00	100.00	100.00	100.00	90.02	87.55	93.30	91.69	93.14	98.31	99.64	99.57	99.53	99.56
9	96.67	98.36	97.09	97.78	98.09	77.04	82.62	83.85	76.18	81.77	70.46	74.32	77.85	66.79	75.08
10	100.00	100.00	100.00	100.00	100.00	100.00	99.88	99.91	99.55	99.48	98.59	99.52	99.44	99.34	99.63
11	77.34	79.01	82.52	80.18	87.07	-	-	-	-	-	88.79	93.89	94.87	93.42	94.58
12	58.29	57.31	64.97	66.05	70.12	-	-	-	-	-	94.09	97.21	98.68	96.24	99.01
13	66.13	75.34	84.22	82.39	86.17	-	-	-	-	-	98.11	99.78	99.74	99.71	99.68
14	95.24	98.67	99.05	99.05	99.24	-	-	-	-	-	97.55	98.95	99.24	99.07	99.07
15	88.29	88.02	86.75	86.68	89.77	-	-	-	-	-	94.41	97.13	97.22	97.11	97.42
16	67.34	79.16	77.62	78.39	80.42	-	-	-	-	-	65.68	71.58	70.86	73.97	76.12
OA	69.78	74.81	79.42	78.46	82.49	77.87	84.14	88.62	87.76	85.62	87.35	89.82	90.60	88.48	90.75
AA	82.08	86.10	88.69	87.72	90.27	84.65	88.18	90.73	89.86	89.40	93.49	95.36	95.78	94.94	95.94
Kappa	65.85	71.43	76.60	75.49	80.04	71.70	79.47	85.06	83.99	81.31	85.92	88.66	89.52	87.19	89.70
Time (s)	2.91	12.62	0.16	26.44	26.11	0.75	1.50	0.03	16.94	17.00	2.14	9.98	0.07	30.31	29.82

cumpléndose la regla de a menor *training*, menor precisión. Sin embargo, también se puede observar que el uso de un número de muestras muy elevado tampoco es recomendable, ya que los algoritmos sobreajustan sus resultados. En este sentido podemos

observar la evolución de la desviación estándar de los distintos métodos: al tener mayor complejidad el dataset IP, los clasificadores se ven obligados a procesar unas características más robustas, por lo que la disminución de los datos de entrenamiento

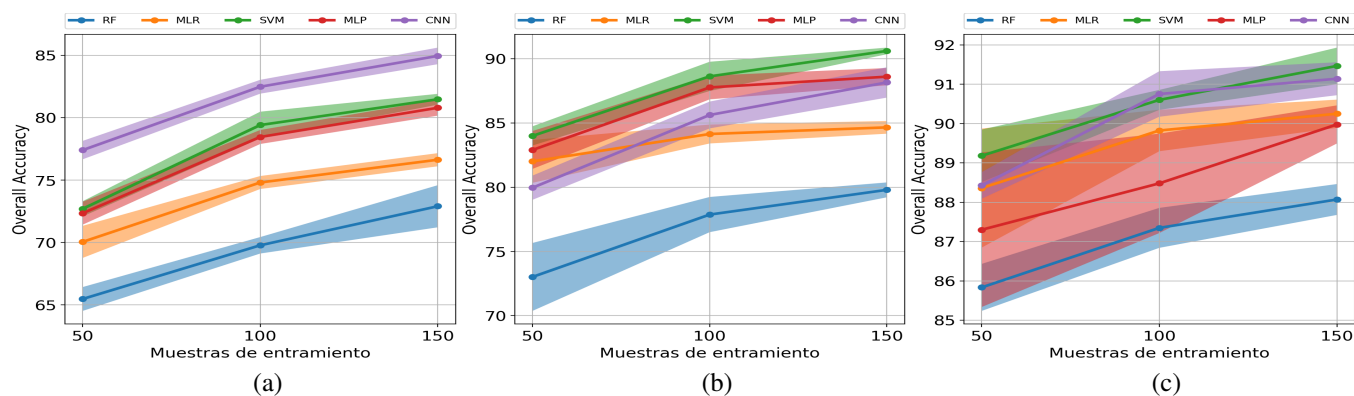


Figura 4: Evolución del Overall Accuracy de los modelos estudiados para los datasets IP (a), UP (b) y SV (c), comparando la precisión alcanzada con diferentes tamaños de muestras por clase

no suponen un aumento de la desviación, manteniéndose generalmente estable también cuando más datos son empleados en el entrenamiento, a excepción del MLP cuya desviación parece aumentar. Sin embargo, observando los resultados obtenidos con UP y SV podemos ver cómo la desviación aumenta sensiblemente cuando hay pocos datos de entrenamiento e incluso cuando se usan más de los necesarios (como es el caso de SV, cuya simplicidad provoca un sobreajuste de parámetros).

4. Conclusiones

En este artículo se han comparado los resultados obtenidos por un total de cinco de los clasificadores más utilizados para el análisis de imágenes hiperespectrales de la superficie terrestre. Se ha realizado una revisión de los principios básicos del funcionamiento de cada uno de los métodos, destacando tanto sus puntos fuertes como sus debilidades y carencias a la hora de analizar datos de gran complejidad. Los experimentos desarrollados facilitan la elección de algoritmos ante distintas situaciones, mostrando el comportamiento de los mismos ante la disponibilidad de diferentes número de muestras en el conjunto de entrenamiento. Una de las principales conclusiones que podemos extraer de este estudio es que no hay ningún clasificador espectral que provea una precisión alta y robusta a cambios en el conjunto de entrenamiento, es por ello que se requieren nuevos desarrollos futuros para aumentar la robustez de los métodos existentes ante esta circunstancia.

Agradecimientos

Este trabajo ha sido financiado por el Ministerio de Educación (Resolución de 26 de diciembre de 2014 y de 19 de noviembre de 2015, de la Secretaría de Estado de Educación, Formación Profesional y Universidades, por la que se convocan ayudas para la formación de profesorado universitario, de los subprogramas de Formación y de Movilidad incluidos en el Programa Estatal de Promoción del Talento y su Empleabilidad, en el marco del Plan Estatal de Investigación Científica y Técnica y de Innovación 2013-2016). Este trabajo también ha sido financiado por la Junta de Extremadura (decreto 279/2014, ayudas para la realización de actividades de investigación y desarrollo tecnológico, de divulgación y de transferencia de cono-

cimiento por los Grupos de Investigación de Extremadura, Ref. GR15005) y por el MINECO (TIN2015-63646-C5-5-R).

Referencias

- Acito, N., Corsini, G., Diani, M., 2003. An unsupervised algorithm for hyperspectral image segmentation based on the gaussian mixture model. In: Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International. Vol. 6. IEEE, pp. 3745–3747.
- Atkinson, P. M., Tatnall, A. R. L., 1997. Introduction Neural networks in remote sensing. *International Journal of Remote Sensing* 18 (4), 699–709. URL: <http://dx.doi.org/10.1080/014311697218700> DOI: 10.1080/014311697218700
- Banerjee, A., Burlina, P., Diehl, C., 2006. A Support Vector Method for Anomaly Detection in Hyperspectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing* 44 (8), 2282–2291. DOI: 10.1109/TGRS.2006.873019
- Bannari, A., Pacheco, A., Staenz, K., McNairn, H., Omari, K., 2006. Estimating and mapping crop residues cover on agricultural lands using hyperspectral and IKONOS data. *Remote Sensing of Environment* 104 (4), 447–459. DOI: 10.1016/j.rse.2006.05.018
- Belgiu, M., Drăguț, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* 114, 24–31.
- Benediktsson, J. A., Ghamisi, P., 2015. *Spectral-Spatial Classification of Hyperspectral Remote Sensing Images*. Artech House.
- Camps-Valls, G., Marsheva, T. V. B., Zhou, D., 2007. Semi-supervised graph-based hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* 45 (10), 3044–3054.
- Chan, J. C.-W., Paelinckx, D., 2008. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment* 112 (6), 2999–3011.
- Chang, C.-I., 2003. *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Springer US. DOI: 10.1007/978-1-4419-9170-6
- Chang, N.-B., Bai, K., Imen, S., Chen, C.-F., Gao, W., 2016. Multisensor Satellite Image Fusion and Networking for All-Weather Environmental Monitoring. *IEEE Systems Journal* PP (99), 1–17. DOI: 10.1109/JSYST.2016.2565900
- Chi, M., Bruzzone, L., 2007. Semisupervised classification of hyperspectral images by svms optimized in the primal. *IEEE Transactions on Geoscience and Remote Sensing* 45 (6), 1870–1880.
- Chiang, S.-S., Chang, C.-I., Ginsberg, I. W., 2000. Unsupervised hyperspectral image analysis using independent component analysis. In: Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International. Vol. 7. IEEE, pp. 3136–3138.
- Chutia, D., Bhattacharyya, D. K., Sarma, K. K., Kalita, R., Sudhakar, S., 2016. Hyperspectral Remote Sensing Classifications: A Perspective Survey. *Transactions in GIS* 20 (4), 463–490. DOI: 10.1111/tgis.12164
- Dixit, V. S., Bhatia, S. K., 2013. Cross Project Validation for Refined Clusters Using Machine Learning Techniques. In: *Computational Science and Its Ap-*

- plications - ICCSA. Springer, Berlin, Heidelberg, pp. 498–512.
DOI: 10.1007/978-3-642-39643-4_36
- Eismann, M. T., 2012. Hyperspectral Remote Sensing. SPIE.
DOI: 10.1117/3.899758
- Green, R. O., Eastwood, M. L., Sarture, C. M., Chrien, T. G., Aronsson, M., Chippendale, B. J., Faust, J. A., Pavri, B. E., Chovit, C. J., Solis, M., Olah, M. R., Williams, O., 1998. Imaging spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). *Remote Sensing of Environment* 65 (3), 227–248.
DOI: 10.1016/S0034-4257(98)00064-9
- Ham, J., Chen, Y., Crawford, M. M., Ghosh, J., 2005. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing* 43 (3), 492–501.
- Haut, J. M., Bernabé, S., Paoletti, M. E., Fernandez-Beltran, R., Plaza, J., Plaza, A., Pla, F., 2018a. Low-high power consumption architectures for deep learning models applied to hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*.
DOI: 10.1109/LGRS.2018.2881045
- Haut, J. M., Paoletti, M., Plaza, J., Plaza, A., 2016. Evaluación del rendimiento de una implementación Cloud para un clasificador neuronal aplicado a imágenes hiperespectrales. *Actas Jornadas Sarteco*, 127–134.
- Haut, J. M., Paoletti, M., Plaza, J., Plaza, A., Jan 2017. Cloud implementation of the k-means algorithm for hyperspectral image analysis. *The Journal of Supercomputing* 73 (1), 514–529.
URL: <https://doi.org/10.1007/s11227-016-1896-3>
DOI: 10.1007/s11227-016-1896-3
- Haut, J. M., Paoletti, M. E., Plaza, J., Li, J., Plaza, A., Nov 2018b. Active learning with convolutional neural networks for hyperspectral image classification using a new bayesian approach. *IEEE Transactions on Geoscience and Remote Sensing* 56 (11), 6440–6461.
DOI: 10.1109/TGRS.2018.2838665
- Haut, J. M., Paoletti, M. E., Plaza, J., Plaza, A., 2018c. Fast dimensionality reduction and classification of hyperspectral images with extreme learning machines. *Journal of Real-Time Image Processing*, 1–24.
- He, N., Paoletti, M. E., Fang, L., Li, S., Plaza, A., Plaza, J., et al., 2018. Feature extraction with multiscale covariance maps for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 1–15.
- Ifarraguerri, A., Chang, C.-I., 2000. Unsupervised hyperspectral image analysis with projection pursuit. *IEEE Transactions on Geoscience and Remote Sensing* 38 (6), 2529–2538.
- Khodadadzadeh, M., Li, J., Plaza, A., Bioucas-Dias, J. M., 2014. A subspace-based multinomial logistic regression for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters* 11 (12), 2105–2109.
- Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kokaly, R. F., Hoefen, T. M., Graham, G. E., Kelley, K. D., Johnson, M. R., Hubbard B. E., Goldfarb, R. J., Buchhorn, M., Prakash, A., 2016. Mineral information at micron to kilometer scales: Laboratory, field, and remote sensing imaging spectrometer data from the orange hill porphyry copper deposit, Alaska, USA. In: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). pp. 5418–5421.
DOI: 10.1109/IGARSS.2016.7730411
- Kunkel, B., Blechinger, F., Lutz, R., Doerffer, R., van der Piepen, H., 1988. ROSIS (Reflective Optics System Imaging Spectrometer) - A candidate instrument for polar platform missions. In: Seeley, J., Bowyer, S. (Eds.), *Optoelectronic technologies for remote sensing from space*. pp. 134–141.
DOI: 10.1117/12.943611
- Lawrence, R. L., Wood, S. D., Sheley, R. L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). *Remote Sensing of Environment* 100 (3), 356–362.
DOI: 10.1016/j.rse.2005.10.014
- Lee, C. A., Gasster, S. D., Plaza, A., Chang, C. I., Huang, B., 2011. Recent developments in high performance computing for remote sensing: A review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4 (3), 508–527.
DOI: 10.1109/JSTARS.2011.2162643
- Li, J., Bioucas-Dias, J. M., Plaza, A., 2010. Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Transactions on Geoscience and Remote Sensing* 48 (11), 4085–4098.
- Liang, S., 2008. *Advances in land remote sensing: system, modeling, inversion and application*. Springer Science & Business Media.
- Ma, W., Gong, C., Hu, Y., Meng, P., Xu, F., 2013. The Hughes phenomenon in hyperspectral classification based on the ground spectrum of grasslands in the region around Qinghai Lake. In: *Proc. SPIE 8910, International Symposium on Photoelectronic Detection and Imaging 2013: Imaging Spectrometer Technologies and Applications*, 89101G. pp. 89101G–89101G–11.
DOI: 10.1117/12.2034457
- Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* 42 (8), 1778–1790.
- Mercier, G., Lennon, M., 2003. Support vector machines for hyperspectral image classification with spectral-based kernels. In: *Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International. Vol. 1. IEEE*, pp. 288–290.
- Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (3), 247–259.
- Naidoo, L., Cho, M. A., Mathieu, R., Asner, G., 2012. Classification of savanna tree species, in the greater kruger national park region, by integrating hyperspectral and lidar data in a random forest data mining environment. *ISPRS Journal of Photogrammetry and Remote Sensing* 69, 167–179.
- Nocedal, J., 1980. Updating Quasi-Newton Matrices With Limited Storage. *Math. of Computation* 35 (151), 773–782.
- Pal, M., Mather, P. M., 2004. Assessment of the effectiveness of support vector machines for hyperspectral data. *Future Generation Computer Systems* 20 (7), 1215–1225.
- Paola, J. D., Schowengerdt, R. A., 1995. A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification. *IEEE Transactions on Geoscience and Remote Sensing* 33 (4), 981–996.
DOI: 10.1109/36.406684
- Paoletti, M., Haut, J., Plaza, J., Plaza, A., 2018. A new deep convolutional neural network for fast hyperspectral image classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 145, 120–147, deep Learning RS Data.
DOI: <https://doi.org/10.1016/j.isprsjprs.2017.11.021>
- Plaza, A., Plaza, J., Paz, A., Sanchez, S., 2011. Parallel Hyperspectral Image and Signal Processing. *IEEE Signal Processing Magazine* 28 (3), 119–126.
- Pour, A. B., Hashim, M., 2014. ASTER, ALI and Hyperion sensors data for lithological mapping and ore minerals exploration. *SpringerPlus* 3 (1), 130.
DOI: 10.1186/2193-1801-3-130
- Rodarmel, C., Shan, J., 2002. Principal component analysis for hyperspectral image classification. *Surveying and Land Information Science* 62 (2), 115–122.
- Rumelhart, D. E., Hinton, G. E., Williams, R. J., 1986. Learning representations by back-propagating errors. *Nature* 323 (6088), 533–536.
- Schölkopf, B., Smola, A. J., Bach, F., et al., 2002. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.
- Serpico, S. B., Bruzzone, L., Roli, F., 1996. An experimental comparison of neural and statistical non-parametric algorithms for supervised classification of remote-sensing images. *Pattern Recognition Letters* 17 (13), 1331–1341.
DOI: 10.1016/S0167-8655(96)00090-6
- Silverman, B., Jones, M. C., 1989. E. Fix and J.L. Hodges(1951): an important contribution to nonparametric discriminant analysis and density estimation. *International Statistical Review* 57 (3), 233–247.
DOI: 10.2307/1403796
- Theiler, J. P., Gislser, G., 1997. Contiguity-enhanced k-means clustering algorithm for unsupervised multispectral image segmentation. In: *Algorithms, Devices, and Systems for Optical Information Processing. Vol. 3159. International Society for Optics and Photonics*, pp. 108–119.
- Vane, G., Evans, D. L., Kahle, A. B., 1989. Recent Advances In Airborne Terrestrial Remote Sensing With The Nasa Airborne Visible/infrared Imaging Spectrometer (aviris), Airborne Synthetic Aperture Radar (sar), And Thermal Infrared Multispectral Scanner (tims). In: *12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium*. pp. 942–943.
DOI: 10.1109/IGARSS.1989.579044
- Wang, F., 1990. Fuzzy classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 28 (2), 194–201.
DOI: 10.1109/36.46698
- Yang, H., Du, Q., Chen, G., 2011. Unsupervised hyperspectral band selection using graphics processing units. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4 (3), 660–668.
- Zhu, H., Basir, O., 2005. An adaptive fuzzy evidential nearest neighbor formulation for classifying remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 43 (8), 1874–1889.
DOI: 10.1109/TGRS.2005.848706