



# Deep learning para la clasificación de usos de suelo agrícola con Sentinel-2

Campos-Taberner, M. <sup>\*</sup>, García-Haro, F.J. , Martínez, B. , Gilabert, M.A. 

*Departament de Física de la Terra i Termodinàmica, Universitat de València, Dr. Moliner 50, Burjassot, España.*

**Resumen:** En el campo de la teledetección se ha producido recientemente un incremento del uso de técnicas de aprendizaje profundo (*deep learning*). Estos algoritmos se utilizan con éxito principalmente en la estimación de parámetros y en la clasificación de imágenes. Sin embargo, se han realizado pocos esfuerzos encaminados a su comprensión, lo que lleva a ejecutarlos como si fueran “cajas negras”. Este trabajo pretende evaluar el rendimiento y acercarnos al entendimiento de un algoritmo de aprendizaje profundo, basado en una red recurrente bidireccional de memoria corta a largo plazo (2-BiLSTM), a través de un ejemplo de clasificación de usos de suelo agrícola de la Comunidad Valenciana dentro del marco de trabajo de la política agraria común (PAC) a partir de series temporales de imágenes Sentinel-2. En concreto, se ha comparado con otros algoritmos como los árboles de decisión (DT), los k-vecinos más cercanos (k-NN), redes neuronales (NN), máquinas de soporte vectorial (SVM) y bosques aleatorios (RF) para evaluar su precisión. Se comprueba que su precisión (98,6% de acierto global) es superior a la del resto en todos los casos. Por otra parte, se ha indagado cómo actúa el clasificador en función del tiempo y de los predictores utilizados. Este análisis pone de manifiesto que, sobre el área de estudio, la información espectral y espacial derivada de las bandas del rojo e infrarrojo cercano, y las imágenes correspondientes a las fechas del período de verano, son la fuente de información más relevante utilizada por la red en la clasificación. Estos resultados abren la puerta a nuevos estudios en el ámbito de la explicabilidad de los resultados proporcionados por los algoritmos de aprendizaje profundo en aplicaciones de teledetección.

**Palabras clave:** *deep learning*, BiLSTM, clasificación, series temporales, Sentinel-2.

## Deep learning for agricultural land use classification from Sentinel-2

**Abstract:** The use of deep learning techniques for remote sensing applications has recently increased. These algorithms have proven to be successful in estimation of parameters and classification of images. However, little effort has been made to make them understandable, leading to their implementation as “black boxes”. This work aims to evaluate the performance and clarify the operation of a deep learning algorithm, based on a bi-directional recurrent network of long short-term memory (2-BiLSTM). The land use classification in the Valencian Community based on Sentinel-2 image time series in the framework of the common agricultural policy (CAP) is used as an example. It is verified that the accuracy of the deep learning techniques is superior (98.6 % overall success) to that other algorithms such as decision trees (DT), k-nearest neighbors (k-NN), neural networks (NN), support vector machines (SVM) and random forests (RF). The performance of the classifier has been studied as a function of time and of the predictors used. It is concluded that, in the study area, the most relevant information used by the network in the classification are the images corresponding to summer and the spectral and spatial information

**To cite this article:** Campos-Taberner, M., García-Haro, F.J., Martínez, B., Gilabert, M.A. 2020. Deep learning for agricultural land use classification from Sentinel-2. *Revista de Teledetección*, 56, 35-48. <https://doi.org/10.4995/raet.2020.13337>

\* Corresponding author: [manuel.campos@uv.es](mailto:manuel.campos@uv.es)

derived from the red and near infrared bands. These results open the door to new studies in the field of the explainable deep learning in remote sensing applications.

**Key words:** deep learning, BiLSTM, classification, time series, Sentinel-2.

## 1. Introducción

El uso del suelo, es decir, la actividad o aprovechamiento que el ser humano realiza sobre él para producir, cambiar o mantener un determinado tipo de cobertura del suelo (Gregrio y Jansen, 2000), es un factor que tiene un impacto decisivo en aspectos socioeconómicos y medioambientales a escala regional y global (Watson et al., 2000).

La clasificación e identificación de usos del suelo sobre la superficie terrestre puede realizarse mediante técnicas de teledetección, que permiten inferir información sobre la naturaleza de las superficies, su estado y los procesos que pueden tener lugar en ellas. La comunidad científica lleva caracterizando usos del suelo desde hace décadas utilizando datos de sensores como el *Moderate Resolution Imaging Spectroradiometer* (MODIS) a bordo de los satélites Terra y Aqua con resoluciones espaciales desde los 250 m (Zhan et al., 2000) a 1 km (Friedl et al., 2002), y sensores de más alta resolución espacial (30 m) como el *Thematic Mapper* (TM) y el *Operational Land Imager* (OLI) a bordo de los satélites Landsat-5/7 (Zhu y Woodcock, 2014) y Landsat-8 respectivamente (Kussul et al., 2017). Gracias a la disponibilidad de datos del programa Copernicus de la Agencia Espacial Europea, han aparecido multitud de estudios de clasificación de usos del suelo utilizando datos del sensor *MultiSpectral Imager* (MSI) a bordo del satélite Sentinel-2 (Immitzer et al., 2016; Vuolo et al., 2018; Griffiths et al., 2019; Campos-Taberner et al., 2019a,b, González-Guerrero y Pons, 2020; Ruiz et al., 2020).

La clasificación digital implica categorizar una imagen o serie de imágenes a partir de sus características espectrales, espaciales y temporales (Chuvieco, 2008). La información en el dominio espacial se suele obtener a partir de medidas de variabilidad espacial de píxeles contiguos mediante el uso de texturas (Haralick et al., 1973). Por su parte, la información temporal se suele

extraer a partir de relaciones entre bandas espectrales, típicamente índices de vegetación (Gilbert et al., 1997), que evolucionan con el tiempo de distinta manera para distintas cubiertas vegetales (Wardlow et al., 2007). Para ello, es conveniente disponer de una serie temporal suficientemente larga para abarcar los ciclos fenológicos de los principales cultivos. Entre los algoritmos de clasificación supervisada más utilizados se encuentran los árboles de decisión (DT, *decision trees*) (Breiman et al., 1984), los *k*-vecinos más cercanos (*k*-NN, *k-nearest neighbors*) (Cover y Hart, 1967), redes neuronales (NN, *neural networks*) (Haykin, 1994), máquinas de soporte vectorial (SVM, *support vector machines*) (Mountrakis et al., 2011) y los bosques aleatorios (RF, *random forests*) (Breiman, 2001).

Estos algoritmos de clasificación forman parte de lo que se conoce como aprendizaje automático o “*machine learning*”. El *machine learning* es una rama de la inteligencia artificial que, en general, trata de aprender patrones matemáticos entre las entradas a un algoritmo y sus salidas, para posteriormente ser aplicados a nuevos datos de entrada y generar predicciones. En el caso de la clasificación de imágenes digitales, las entradas son las características espectrales, espaciales y temporales de las imágenes, y las salidas las categorías clasificadas (Camps-Valls et al., 2013). En este contexto, y en el ámbito de la teledetección, recientemente se ha popularizado el uso de un tipo de técnicas de *machine learning* que forman parte de lo que se conoce como aprendizaje profundo o “*deep learning*” (Reichstein et al., 2019). Estas técnicas aprenden representaciones de los datos de entrada de manera sucesiva y jerarquizada mediante capas de manera similar a como lo hacen las NN, pero involucrando más de dos capas en la red (de ahí el término “*deep*”). *Grosso modo* en los algoritmos de *deep learning* se puede distinguir entre dos grandes grupos: técnicas que utilizan arquitecturas con redes neuronales convolucionales (CNN, *convolutional neural*

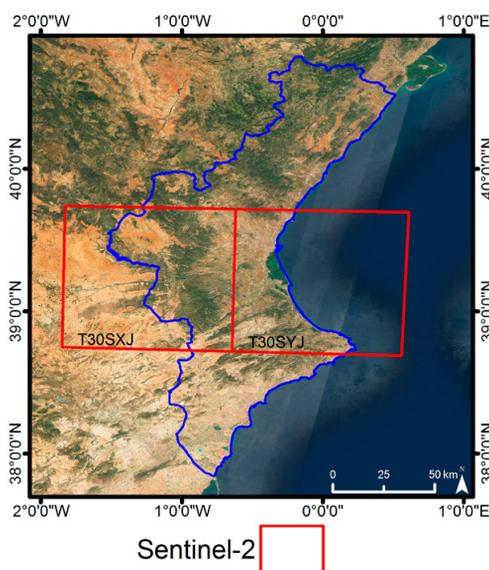
*networks*) y las redes de aprendizaje secuencial también llamadas redes neuronales recurrentes (RNN, *recurrent neural networks*). Las CNN realizan convoluciones en el dominio espacial de una imagen (Maggiore et al., 2016), mientras que las RNN también pueden utilizar el dominio temporal (Ndikumana et al., 2018) y el espacio hiperespectral (Mou et al., 2017) como característica fundamental de aprendizaje. Para más información sobre estas técnicas de *deep learning* en el ámbito de la teledetección, referimos al lector a los trabajos de Zhang et al. (2016), Zhu et al. (2017), Ma et al. (2019). En general, los algoritmos *deep learning* mejoran la precisión de las clasificaciones respecto de las técnicas *machine learning* (Campos-Taberner et al., 2015; Zhong et al., 2019). No obstante, la comprensibilidad de todas estas técnicas, y principalmente de las técnicas *deep learning* es limitada (Montavon et al., 2018), y en general, cuanto mejor es el aprendizaje de un algoritmo, más difícil es su interpretación (Gunning et al., 2019). Esta falta de comprensibilidad es un aspecto clave a la hora de utilizar estos algoritmos. Es conveniente, y muy enriquecedor, que el usuario sepa cómo el algoritmo está llegando a las predicciones proporcionadas (Samek et al., 2020).

El objetivo de este trabajo es generar, mediante un algoritmo de *deep learning*, un mapa de clasificación de usos de suelo agrícola en una zona de la provincia de Valencia que sea de utilidad para la toma de decisiones en el marco de la política agraria común (PAC) de la Unión Europea. Para este propósito se empleará la serie temporal de imágenes Sentinel-2 sin nubes disponibles a lo largo del año agronómico 2017/2018. En concreto se utilizará un algoritmo basado en una red recurrente bidireccional de memoria larga a corto plazo (BiLSTM, *bi-directional long short-term memory*) (Hochreiter y Schmidhuber, 1997). Además de proporcionar la clasificación, y con objeto de indagar someramente en el funcionamiento del algoritmo para mejorar su comprensión, se evaluará la influencia de diferentes arquitecturas de red así como la importancia de los predictores y de las fechas de las imágenes Sentinel-2 en el entrenamiento, y se comparará la precisión del algoritmo con la de otros algoritmos de *machine learning*.

## 2. Materiales y métodos

### 2.1. Zona de estudio

La zona de estudio, en la *Comunitat Valenciana*, es la contenida en las escenas Sentinel-2 correspondientes a los cuadrantes T30SYJ y T30SXJ, marcadas en rojo en la Figura 1. Las dos escenas Sentinel-2 abarcan conjuntamente una superficie de 220×220 km<sup>2</sup>. El objetivo del trabajo es obtener un mapa de los usos de suelo definidos por el Sistema de Información Geográfica de Parcelas Agrícolas (SIGPAC) (<https://www.fega.es>). SIGPAC categoriza los usos de suelo en treinta clases incluyendo diferentes tipos de vegetación natural, zonas urbanas, mixtas y edificaciones, corrientes de agua, y cultivos permanentes y anuales. En este estudio se han seleccionado dieciséis usos de suelo no urbanos que son los predominantes en la zona de estudio (Campos-Taberner et al., 2019a): cítricos (CI), pasto con arbolado (PR), pasto arbustivo (PA), olivar (OV), frutales (FY), frutos secos (FS), pastizal (PS), forestal (FO), viña (VI), y las tierras arables que incluyen cultivos de arroz (TA-AR), avena (TA-AV), cebada (TA-CE), trigo (TA-TR), triticale (TA-TT), girasol (TA-GI), y zonas de barbecho (TA-BA).



**Figura 1.** Localización de la zona de estudio (cuadrantes T30SXJ + T30SYJ de Sentinel-2, marcados en rojo) en la *Comunitat Valenciana*.

## 2.2. Series temporales de imágenes Sentinel-2

La misión Sentinel-2 está compuesta por dos satélites similares (Sentinel 2-A y Sentinel 2-B) de órbita polar, ubicados en la misma órbita heliosíncrona pero desfasados 180°. Gracias a ello, su periodo de revisita o resolución temporal pasa de 10 días en el Ecuador, con un solo satélite, a 5 días con los dos. Ambas plataformas llevan a bordo el sensor MSI que permite la adquisición de imágenes multispectrales (13 bandas en total, comprendiendo zonas del espectro visible, infrarrojo cercano e infrarrojo de onda corta) de alta resolución espacial (entre 10 m y 60 m, dependiendo de la banda; véase Tabla 1). Su campo de observación es de 290 km y su cobertura prácticamente global. En este estudio se han utilizado sesenta imágenes Sentinel-2 (treinta para cada cuadrante T30SYJ y T30SXJ) durante el año agronómico 2017/2018. La Tabla 2 muestra las fechas de las imágenes utilizadas.

Las imágenes se descargaron utilizando la plataforma facilitada por la ESA para la

diseminación de datos de las misiones Sentinel (<https://scihub.copernicus.eu>). Se descargó el producto de nivel 2A que corresponde a datos de reflectividad a nivel de la cubierta vegetal ya corregido de efectos atmosféricos mediante el uso de la herramienta *Sen2Cor*. Este producto proporciona 12 bandas (la banda B10 es omitida ya que se utiliza en el proceso de corrección atmosférica y no da información a nivel de superficie). Las bandas de 20 m y 60 m se remuestrearon a 10 m para disponer de imágenes multispectrales del mismo tamaño en todos los canales.

Para evaluar los algoritmos de clasificación se utilizaron como predictores catorce secuencias de información (o series temporales) en las treinta fechas seleccionadas, en concreto: doce secuencias correspondientes a todas las bandas Sentinel-2 disponibles en cada fecha, así como también la secuencia temporal del índice de vegetación de diferencia normalizada (NDVI, *normalized difference vegetation index*) y la de su entropía ( $E_{NDVI}$ ). La entropía (Baraldi y Parmiggiani, 1995) es un descriptor espacial que nos mide la aleatoriedad de la imagen. Para su cálculo se ha utilizado una ventana de 9×9 píxeles y 256 niveles de gris en la matriz de coocurrencia.

**Tabla 1.** Resolución espacial de las 13 bandas espectrales del Sentinel-2.

Resolución espacial (m)	Banda (longitud de onda central en nm)
10	B2 (490)
	B3 (560)
	B4 (665)
	B8 (842)
20	B5 (705)
	B6 (740)
	B7 (783)
	B8a (865)
	B11 (1610)
	B12 (2190)
60	B1 (443)*
	B9 (945)**
	B10 (1375)***

Especialmente enfocadas al estudio de aerosoles (\*), vapor de agua (\*\*) y cirrus (\*\*\*).

## 2.3. Verdad terreno

La selección de muestras utilizadas para el entrenamiento y validación del algoritmo de clasificación se realizó utilizando la verdad terreno proporcionada por la *Conselleria d'Agricultura, Medi Ambient, Canvi Climàtic i Desenvolupament Rural* de la *Generalitat Valenciana*. Esta información proviene de inspecciones de campo realizadas durante el año 2018 en las que se detalla el uso de suelo comprobado *in situ*. La Tabla 3 muestra el número de píxeles etiquetados por cada clase. Se utilizan dos tercios de los datos para el entrenamiento del algoritmo de clasificación y el tercio restante para la validación. Una vez seleccionadas las muestras de entrenamiento, se geolocalizan

**Tabla 2.** Fechas de las imágenes Sentinel-2 (año agronómico 2017/2018).

Año	2017					2018																								
Mes	Sep.	Oct.	Nov.	Dic.		Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.																	
Día	13	3	13	12	22	2	12	17	22	1	21	31	15	21	7	22	27	1	16	15	20	25	5	10	15	30	4	9	24	29

sobre las imágenes Sentinel-2 para asociar a cada píxel su correspondiente conjunto de catorce secuencias temporales.

**Tabla 3.** Número de píxeles de cada uso SIGPAC comprobado por visitas de campo *in situ*.

Uso SIGPAC	Nº de píxeles
OV	49 829
FO	495 223
FS	153 727
TA-A	230 935
TA-AV	28 754
TA-C	71 623
TA-G	10 104
TA-T	11 437
TA-TT	4 252
TA-B	84 491
PS	33 408
FY	29 859
PR	165 110
PA	663 995
VI	240 418
CI	125 161

## 2.4. Algoritmos de clasificación

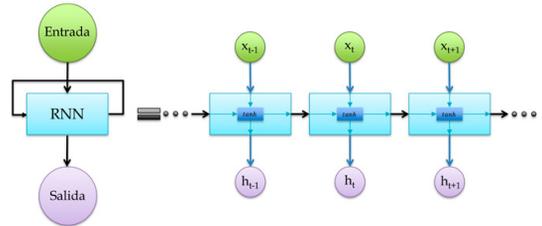
### 2.4.1. Redes neuronales recurrentes (RNN)

Las RNN son un tipo de redes neuronales que se retroalimentan ya que tienen una arquitectura en forma de cadena en la que se van repitiendo módulos de red (véase Figura 2). En estas redes, cada neurona es alimentada por los datos de entrada a la red en caso de ser la primera capa de la red, o de la capa anterior si no es la primera capa, además también recibe la información generada por esa misma neurona en la interacción anterior. En una RNN estándar el módulo de repetición puede contener simplemente una función de activación (típicamente una tangente hiperbólica) tal y como se muestra en la Figura 2. La salida de cada paso temporal ( $y_t$ ) viene dada por el estado oculto correspondiente que depende de la entrada en el instante  $t$  ( $x_t$ ) y de la salida o estado oculto previo ( $h_{t-1}$ ) de la siguiente manera:

$$y_t = \tanh(Wx_t + Uh_{t-1} + b) \quad (1)$$

donde  $W$ ,  $U$  y  $b$  son los pesos de la entrada, los pesos recurrentes y el sesgo, respectivamente, los cuales se obtienen mediante retropropagación de errores. Las RNN destacan por su poder de clasificación en problemas de secuencias como

el procesado de lenguaje natural (Mikolov et al., 2011), el reconocimiento de voz (Graves et al., 2013) y modelado generativo de imágenes (Van Den Oord, et al., 2016).



**Figura 2.** Arquitectura de una red neuronal recurrente típica.

### 2.4.2. Redes LSTM

Las RNN sufren de lo que se conoce como el “desvanecimiento del gradiente” (Bengio et al., 1994) ya que utilizan funciones de activación sigmoideas o tangentes hiperbólicas. Estas funciones ofrecen salidas en el rango 0-1 haciendo que, en los valores cercanos a las asíntotas, una variación grande de la entrada comporte una variación pequeña en la salida (i.e., un gradiente pequeño). Este efecto es mucho mayor cuando se incrementa el número de capas utilizadas en el diseño de la red, ya que en el entrenamiento (típicamente propagación hacia atrás) el gradiente se va desvaneciendo haciendo que las neuronas de las primeras capas no aprendan. Una manera de solucionar este problema de aprendizaje es mediante el uso de redes recurrentes LSTM, compuestas por módulos que proporcionan a la red “memoria” a largo plazo y evitan el desvanecimiento del gradiente. Para producir las salidas de la red, las LSTM tienen en cuenta, de una manera iterativa, los estados pasados de las series temporales utilizados como entrada en la red. Estas dependencias temporales se tienen en cuenta a través de lo que se conoce como memoria de la red ( $C_t$ ). Esta información fluye a través de cada uno de los módulos LSTM que, está compuesto por tres puertas: entrada ( $i_t$ ), olvido ( $f_t$ ), y salida ( $o_t$ ) (véase Figura 3). La puerta de olvido determina qué información se elimina en la memoria a través de una función sigmoide:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (2)$$

De manera similar, la puerta de entrada determina qué información se guarda en la memoria a través de otra función sigmoide:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (3)$$

La información en la memoria se actualiza mediante la suma de la información proveniente de la puerta de entrada y de olvido:

$$C_t = f_t \times C_{t-1} + i_t \times g_t \quad (4)$$

Finalmente, el estado oculto saliente se obtiene mediante la puerta de salida y la memoria actualizada a través de una función tangente hiperbólica:

$$h_t = o_t \times \tanh(C_t) \quad (5)$$

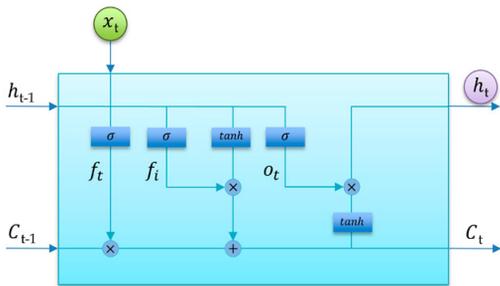


Figura 3. Módulo LSTM.

### 2.4.3. Redes BiLSTM

Los módulos de las redes LSTM se pueden combinar de tal manera que permiten obtener redes LSTM bidireccionales (BiLSTM) (Hochreiter y Schmidhuber, 1997). Las redes BiLSTM están compuestas por dos módulos LSTM que, en cada paso temporal de las series temporales que se introducen en la red, tienen en cuenta las dependencias temporales no solo en estados pasados sino también en estados futuros (Schuster y Paliwal, 1997). En este caso la información proporcionada por la memoria de los estados pasados y futuros se combina para producir la salida de la red.

En este estudio se ha utilizado una combinación de redes BiLSTM tal y como se muestra en la Figura 4. Los componentes principales de esta arquitectura de red son: (i) la capa de entrada formada por el conjunto de las catorce series temporales seleccionadas en la sección 2.2, (ii) dos capas BiLSTM a la salida de las cuales se les aplica una función “dropout” que desactiva un número de neuronas, en nuestro caso el 50%, de forma aleatoria para evitar el sobreajuste de la red a los datos de entrenamiento, (iii) una capa “fully-connected” que conecta cada

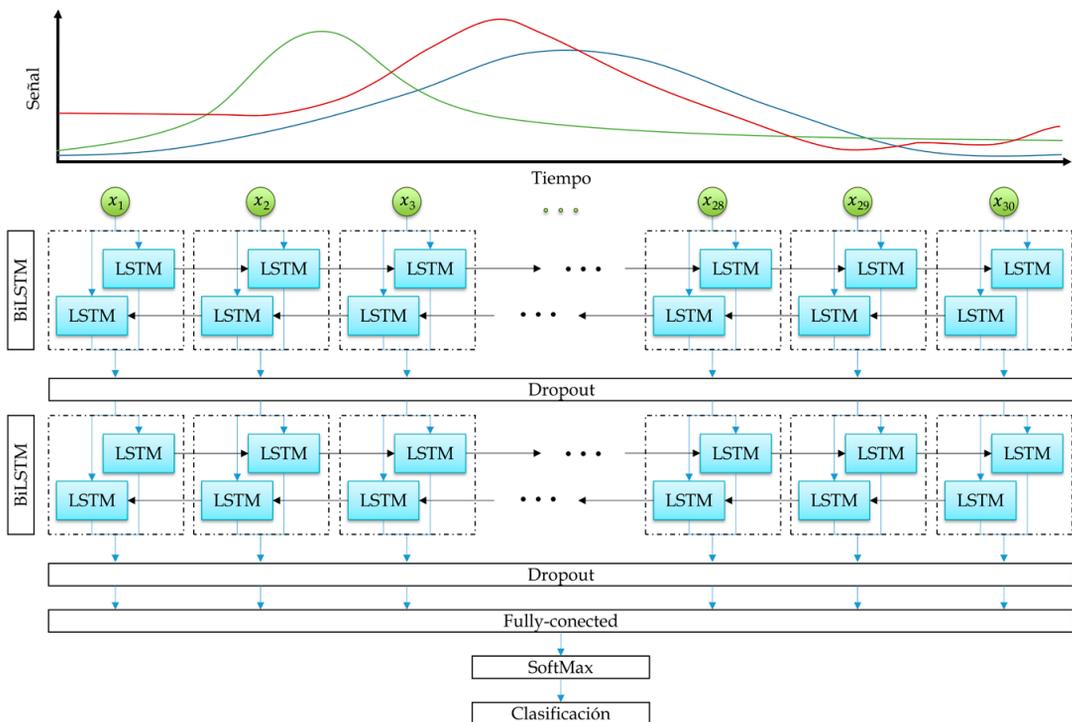


Figura 4. Arquitectura de la red 2-BiLSTM propuesta. La capa de entrada está formada por catorce secuencias de treinta pasos temporales.

neurona de la siguiente capa con todas las neuronas de la capa anterior, (iv) una función “*softmax*” que calcula la probabilidad de cada clase en la salida de la red, y (v) la capa de salida que contiene la clase asignada a cada píxel.

## 2.5. Evaluación

Para realizar el análisis de los resultados de la clasificación obtenida con la arquitectura de red descrita en la sección anterior (2-BiLSTM), se ha construido la matriz de confusión (Story y Congalton, 1986) y un índice de precisión derivado de ella, el acierto global (AG), obtenido sobre el conjunto de validación (no utilizado durante la fase de entrenamiento). Este acierto global se comparó con precisiones obtenidas utilizando otras arquitecturas de red, en concreto, tres redes similares con una, tres y cuatro capas BiLSTM (1-BiLSTM, 3-BiLSTM y 4- BiLSTM, respectivamente), así como también con la precisión obtenida utilizando los algoritmos *machine learning* DT, k-NN, NN, SVM y RF.

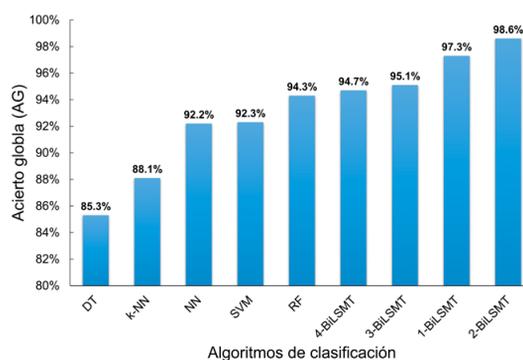
El comportamiento de la red 2-BiLSTM se abordó mediante la identificación de los predictores y fechas más relevantes en el entrenamiento de la red. La identificación se realizó mediante un análisis de permutación de ruido en cada predictor. Este análisis consistió en la adición de ruido Gaussiano  $\mathcal{N}(0, \sigma^2)$ , siendo  $\sigma$  el 3% de la amplitud de la señal perturbada, a todos los pasos temporales de un predictor seleccionado y dejando el resto de predictores sin perturbar. Este proceso se repitió para cada uno de los catorce predictores ejecutándose la red en cada caso y obteniéndose una precisión global distinta en cada ejecución. La relevancia de los predictores se calculó como la diferencia entre la precisión alcanzada sin perturbar ningún predictor y la obtenida perturbando cada uno de ellos por separado. Los resultados se normalizaron respecto al predictor más relevante. Este procedimiento de evaluación de relevancia se repitió para identificar la fecha más relevante en el proceso de clasificación. En este caso, se perturbaron todos los predictores en un único paso temporal dejando el resto sin perturbar. El proceso se repitió treinta veces permutando la adición de ruido en cada uno de los pasos temporales. Finalmente, se realizó una evaluación de la evolución temporal de la clasificación, lo que da una idea del comportamiento predictivo de la red a lo largo del tiempo.

## 2.6. Mapa de usos de suelo agrícola

El modelo entrenado con la red 2-BiLSTM se utilizó para obtener en la zona de estudio un mapa de usos de suelo agrícola y un mapa de confianza de la clasificación. Para ello fue necesario calcular los predictores para cada uno de los píxeles de las imágenes Sentinel-2 localizados dentro del área de estudio. Una vez calculados, los predictores fueron introducidos como entradas en el modelo, resultando como salidas del modelo los mapas de clasificación y confianza. En la realización de los mapas se enmascararon zonas de no interés (p. ej. zonas urbanas).

## 3. Resultados

La Figura 5 muestra un gráfico de barras que representa la precisión obtenida con los diferentes clasificadores que se han tenido en cuenta en este estudio. Los resultados más precisos se obtuvieron con el uso de la red 2-BiLSTM (98,6%) mientras que los menos precisos los produjo el uso del DT (85,3%).



**Figura 5.** Precisión obtenida sobre el conjunto de validación utilizando cada uno de los clasificadores.

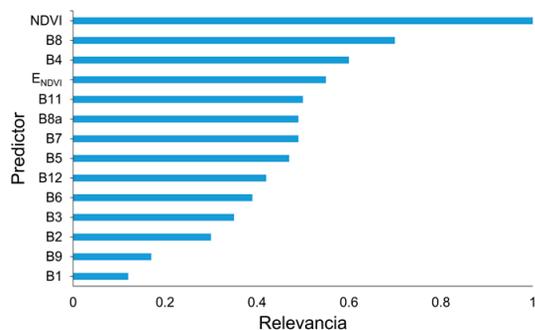
Con el objetivo de evaluar el acierto logrado en cada una de las clases, la Tabla 4 muestra la matriz de confusión obtenida sobre el conjunto de validación utilizando el algoritmo con mayor acierto global (i.e., el 2-BiLSTM). La mayor precisión, 99,9%, se obtuvo sobre la clase tierra arable-arroz (TA-A), mientras que la mayor confusión se produjo sobre la clase de frutales (FY), con precisiones de usuario y productor del 91,1% y 92,6% respectivamente.

La relevancia normalizada de cada serie temporal que actúa como predictor en la red se muestra

**Tabla 4.** Matriz de confusión obtenida con la red 2-BiLSTM. El acierto global es del 98,6%.

Verdad Terreno	TA-A	FO	PA	PS	PR	TA-T	VI	TA-TT	CI	TA-C	FS	TA-B	TA-AV	TA-G	OV	FY													
TA-A	78382	3		2													41	2					4	99.9%	0.1%				
FO	10	167784	75	41													60	50	21	41				5	12	99.8%	0.2%		
PA		47	224824	11	579	19	1										3		20	244	179			20		99.5%	0.5%		
PS	1	132	14	10969	35		2										78	1	15	6	3			4	1	97.4%	2.6%		
PR			531	26	55477		2										1	1	2	6	3			1	2	99.0%	1.0%		
TA-T		4	12				3721	22									7	6		17	21			1		97.6%	2.4%		
VI		14	10	3	12	5	80431										86	125	495	253	10	37	188	50		98.4%	1.6%		
TA-TT							1	1424												18	19						97.4%	2.6%	
CI	18	65	5	16				41									41835		106	42				17	357	98.4%	1.6%		
TA-C		62	22		2			120											23850	55	136	30	67	50	16		97.7%	2.3%	
FS		53	36		6		567										63	45	51482	60	20	1	369	18		97.7%	2.3%		
TA-B	1	42	306	9	19	22	462	30	115	139	179	27145	130	69	59	33										94.4%	5.6%		
TA-AV			195	1	15	18	11										1	38	36	183	9174			34	7	94.5%	5.5%		
TA-G		2		3			20										3	116	1	45				3197	2	4	94.2%	5.8%	
OV		53	69		4	1	290										73	72	700	69	35	1	15294	239		90.5%	9.5%		
FY	1	43	2	7			29										484	26	46	26	11	6	235	9352		91.1%	8.9%		
	99.9%	99.7%	99.4%	98.9%	98.8%	98.3%	98.1%	97.9%	97.6%	97.5%	96.8%	95.9%	95.4%	94.6%	93.9%	92.6%													
	0.1%	0.3%	0.6%	1.1%	1.2%	1.7%	1.9%	2.1%	2.4%	2.5%	3.2%	4.1%	4.6%	5.4%	6.1%	7.4%													
	TA-A	FO	PA	PS	PR	TA-T	VI	TA-TT	CI	TA-C	FS	TA-B	TA-AV	TA-G	OV	FY													
	Clasificación																												

en la Figura 6. La serie temporal del NDVI es el predictor más relevante de los utilizados en la red, mientras que la información de la evolución temporal de la banda 1 (B1) es la menos utilizada en las predicciones. De manera similar la Figura 7 muestra la relevancia normalizada de cada una de las fechas o pasos temporales. La fecha 28, correspondiente al 24 de agosto de 2018, es la fecha cuya información es más relevante en el entrenamiento de la red. Por su parte, la imagen Sentinel-2 correspondiente al 17 de diciembre de 2017 (fecha 8) es la menos relevante en las predicciones de la red.



**Figura 6.** Relevancia de cada predictor utilizado en las predicciones de la red 2-BiLSTM.

La Figura 8 es un ejemplo del comportamiento predictivo de la red a lo largo del tiempo. Se muestra la evolución temporal de la clasificación (Figura 8a), utilizando para ello información de una única fecha en cada instante temporal sobre un píxel etiquetado como viña (VI). Se obtendría una asignación correcta a dicha clase al considerar las primeras y últimas fechas de la secuencia temporal, i.e. los meses de verano, y una asignación incorrecta al considerar únicamente las imágenes de las fechas de invierno. En el caso que se muestra en la figura, la clase finalmente asignada por la red corresponde con la misma clase verificada en el campo. Asociada a esta evolución temporal, la figura 8b muestra la probabilidad asignada a dicho píxel para cada posible clase en cada fecha. Se observa que en el periodo comprendido entre principios de enero (fecha 11) y finales de junio (fecha 23) de 2018 la red no es capaz de identificar correctamente el píxel, mostrando su probabilidad máxima en tres clases distintas (PS, TA-AV y TA-B) (véase Figura 8b). Este comportamiento predictivo de la red se debe a que durante dicho periodo los cultivos de viña en la zona de estudio no son claramente identificables mediante la señal de Sentinel-2, ya que la viña es un cultivo que presenta su principal actividad fotosintética en verano, alcanzando su maduración a finales de septiembre/principios de octubre.

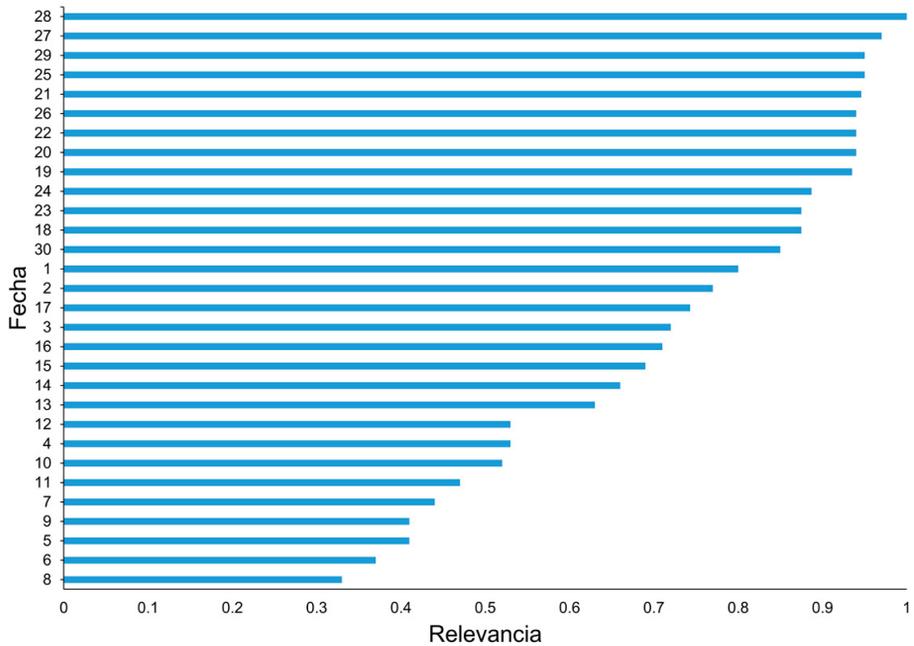


Figura 7. Relevancia de cada fecha utilizada en las predicciones de la red 2-BiLSTM.

Finalmente, el modelo entrenado con la red 2-BiLSTM se utilizó para obtener en la zona de estudio un mapa de usos de suelo agrícola y un mapa de confianza de la clasificación. Para ello fue necesario calcular los predictores para cada uno de los píxeles de las imágenes Sentinel-2 localizados dentro del área de estudio. Una vez calculados, los predictores fueron introducidos como entradas en el modelo, resultando como salidas del modelo las

clasificaciones y las probabilidades asociadas a las mismas. Con estas salidas se derivaron los mapas de clasificación (Figura 9a) y confianza (Figura 9b). El mapa de confianza muestra la probabilidad con la que cada píxel es clasificado. En la realización de los mapas se enmascararon zonas de no interés (p. ej. zonas urbanas).

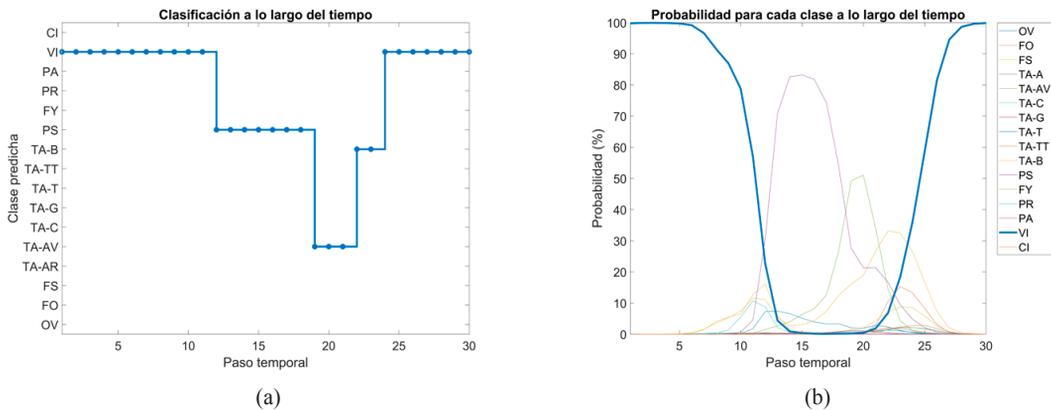
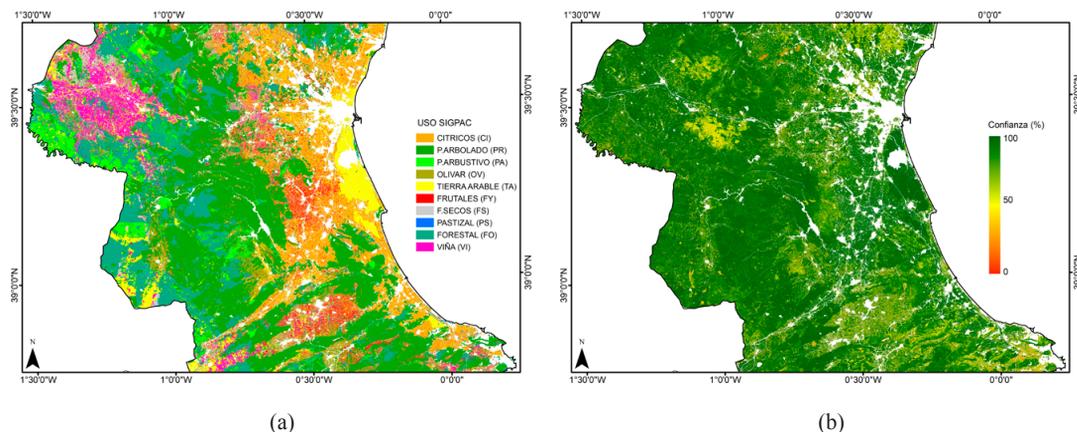


Figura 8. Para un píxel etiquetado en campo como viña (VI): (a) evolución temporal de la clasificación y, (b) probabilidad de clasificación en cada clase a lo largo del tiempo.



**Figura 9.** (a) Mapa de clasificación de usos del suelo y (b) mapa de confianza obtenidos sobre la zona de estudio. A efectos de visualización, se han agrupado las tierras arables en una sola clase (TA).

#### 4. Discusión

La clasificación, mediante datos de teledetección, de una escena en la que coexisten muchos tipos de cultivos se beneficia enormemente del uso de una serie temporal de imágenes lo suficientemente larga y con alta frecuencia temporal como para abarcar los ciclos fenológicos de todos los cultivos y aprovechar las diferencias que se establecen entre las distintas clases a lo largo del tiempo. La elevada resolución temporal de los satélites Sentinel-2 resulta especialmente apropiada para estudiar la dinámica de la vegetación y su catalogación tipológica. Además de la información espectral y temporal, se ha utilizado una textura para extraer información espacial de las imágenes Sentinel-2.

En estos problemas de discriminación de clases en los que se dispone de información de series temporales a partir de las cuales extraer conocimiento, como es el caso que nos ocupa, el uso de redes neuronales recurrentes es una ventaja dado que incorporan en su diseño la retroalimentación de información. Las RNN convencionales presentan inestabilidades y problemas durante la fase de entrenamiento porque los gradientes retropropagados tienden a desvanecerse con el tiempo. Esto provoca dificultades para utilizar dependencias a largo plazo que, sin embargo, son mitigadas por las redes LSTM incorporando una serie de pasos para decidir qué información va a ser almacenada y cuál borrada. Se consigue así que la red tenga cierta “memoria”. Si además esta memoria aporta información tanto de estados pasados como de

estados futuros (caso de las BiLSTM), su uso en aplicaciones donde se requiere extraer conocimiento de la serie temporal completa en ambas direcciones para tener una visión global de toda la evolución temporal es de especial interés.

Los resultados de clasificación obtenidos confirman que el uso de capas BiLSTM mejora la precisión global del resto de clasificadores evaluados. Respecto a la arquitectura de las redes recurrentes evaluadas, la red 2-BiLSTM fue la que mejores resultados produjo. Por lo tanto, en el caso que nos ocupa, no siempre incrementar el número de capas en una red neuronal profunda es sinónimo de mejores resultados de clasificación. Esto se debe en parte a que en las arquitecturas con muchas capas, aunque se utilicen capas *dropout*, las redes pueden tender a sobreajustarse a los datos de entrenamiento y perder capacidad de generalización con lo que la precisión de las predicciones decrece. Además, puede persistir el problema del desvanecimiento del gradiente que, aunque en las arquitecturas en las que se usan capas LSTM y BiLSTM se tiende a mitigar, no desaparece completamente cuando se usan arquitecturas con muchas capas.

Usando la red 2-BiLSTM todas las clases fueron clasificadas con un acierto mayor al 91% tanto en precisión de usuario como de productor. La mayor precisión se obtuvo sobre cultivos de arroz (TA-A) donde prácticamente la totalidad de los píxeles son clasificados correctamente lo que destaca la utilidad de los datos multitemporales de Sentinel-2 en la caracterización de este tipo de

cultivos (Campos-Taberner et al., 2019a). Por su parte, la clase frutales (FY) presenta una confusión más elevada con las clases CI y OV, haciendo que el poder discriminativo del clasificador sea menor entre dichos usos.

Es conveniente que, a la hora de realizar una clasificación, no solo se documenten los resultados en términos de precisión, sino que se aporte también la explicación de lo que está ocurriendo internamente en el clasificador para posteriormente interpretar los resultados. Esta evaluación es uno de los principales retos que presenta el uso de algoritmos *deep learning* ya que estos algoritmos pueden ser, hasta cierto punto, cajas negras que realizan clasificaciones con gran precisión sin que el operador llegue a poder interpretar lo que está sucediendo en el algoritmo ni pueda establecer cuál es la contribución de cada una de las entradas (predictores) en él. En este aspecto, el procedimiento realizado para evaluar la importancia de los predictores en las clasificaciones de la red, sacó a la luz que la red extrae la información principalmente de la evolución temporal del NDVI, de las bandas B8 y B4 y de la información espacial proporcionada por la entropía del NDVI. Esto tiene sentido dado que la curva temporal del NDVI refleja la evolución fenológica, que es un rasgo identificativo de los distintos tipos de cubiertas presentes en la zona de estudio. De hecho, según el análisis de relevancia por fecha, la red utiliza principalmente en su entrenamiento la información de las imágenes Sentinel-2 adquiridas en verano, lo que concuerda con la actividad fenológica de la mayor parte de clases identificadas en la zona de estudio. Por otra parte, la banda B8 está ubicada en el infrarrojo cercano, zona del espectro que contiene información muy relevante acerca de la vegetación.

El mapa de usos de suelo obtenido sigue la distribución espacial esperada en la zona de estudio mientras que el mapa de confianza muestra que la gran mayoría de zonas fueron clasificadas con alta confianza (colores verdes). Sin embargo, existen pequeñas zonas (colores rojizos) en las que las clasificaciones deberían tomarse con cautela ya que la confianza de la clasificación en ellas es baja. La generación y actualización de estos mapas puede ofrecer una herramienta de gran valor en el marco de trabajo de asignación de ayudas de la PAC (Política Agraria Común) a los agricultores

(Campos-Taberner et al., 2019a). Los resultados de la precisión del clasificador (98,6 %) están ligeramente por encima de los alcanzados en estudios similares en el marco de trabajo de la PAC. Por ejemplo, Estrada et al. (2017) obtuvieron un 95% y 85% de acierto en dos zonas de estudio, respectivamente, Sitokonstantinou et al. (2018) un 91,3% de acierto sobre nueve clases; Schmedtmann y Campagnolo (2015) un 68% de acierto clasificando 12 clases y Tapsall et al. (2010) un 81% de precisión sobre seis clases.

## 5. Conclusiones

Se ha evaluado el comportamiento de una red recurrente bidireccional de memoria corta a largo plazo (2-BiLSTM) en la clasificación de usos de suelo agrícola de la Comunidad Valenciana a partir de series temporales de datos de Sentinel-2. La red 2-BiLSTM fue utilizada para generar un mapa de usos de suelo agrícola que se puede utilizar como una fuente independiente de información en el marco de las ayudas procedentes de la PAC.

En problemas de clasificación se suele elegir un algoritmo u otro, básicamente en función de su precisión global. Dicha elección debería fundamentarse no solo en su capacidad predictiva, sino también en su comportamiento y adecuación al tipo de aplicación o problema. En el caso que ocupa a este trabajo, se eligió el uso de redes BiLSTM porque son capaces de extraer información de las series temporales completas en ambas direcciones para identificar diferentes patrones con diferentes usos de suelo.

El acierto global alcanzado por la red 2-BiLSTM fue del 98,6% mejorando la precisión del resto de algoritmos de clasificación evaluados. En los dieciséis usos de suelo agrícola clasificados la precisión superó el 91% tanto en precisión del usuario como del productor, lo que pone de manifiesto la robustez del algoritmo. De las arquitecturas diseñadas, los mejores resultados se alcanzaron con la utilización de dos capas BiLSTM, lo cual indica que en los algoritmos de aprendizaje profundo o *deep learning*, el hecho de aumentar las capas no es sinónimo de mejor rendimiento. El estudio del comportamiento de la red reveló que la información más relevante en el entrenamiento de la red es la que se extrae de los predictores NDVI, B8, B4 y  $E_{NDVI}$ , así como de las imágenes correspondientes

a las fechas del período estival. Estos resultados abren la puerta a nuevos estudios en el ámbito de la evaluación de los algoritmos de aprendizaje profundo en aplicaciones de teledetección.

## Agradecimientos

Este trabajo ha sido subvencionado gracias al Convenio 2019 y 2020 de colaboración entre la Generalitat Valenciana, a través de la *Conselleria d'Agricultura, Medi Ambient, Canvi Climàtic i Desenvolupament Rural*, y la *Universitat de València – Estudi General*.

## Referencias

Baraldi, A., Parmiggiani, F. 1995. An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE Transactions on Geoscience and Remote Sensing*, 33(2), 293-304. <https://doi.org/10.1109/36.377929>

Bengio, Y., Simard, P., Frasconi, P. 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157-166. <https://doi.org/10.1109/72.279181>

Breiman, L., Friedman, J., Olshen, R.A., Stone, C.J. 1984. *Classification and regression trees*. Taylor & Francis: London, UK.

Breiman, L. 2001. Random forests. *Machine learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>

Campos-Taberner, M., Romero-Soriano, A., Gatta, C., Camps-Valls, G., Lagrange, A., Le Saux, B., Beaupère, A., Boulch, A., Chan-Hon-Tong, A., Herbin, S., Randrianarivo, H., Ferecatu, M., Shimoni, M., Moser, G., Tuia, D. 2016. Processing of extremely high-resolution Lidar and RGB data: outcome of the 2015 IEEE GRSS data fusion contest–part a: 2-D contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(12), 5547-5559. <https://doi.org/10.1109/JSTARS.2016.2569162>

Campos-Taberner, M., García-Haro, F.J., Martínez, B., Sánchez-Ruiz, S., Gilabert, M.A. 2019a. A Copernicus Sentinel-1 and Sentinel-2 Classification Framework for the 2020+ European Common Agricultural Policy: A Case Study in València (Spain). *Agronomy*, 9(9), 556. <https://doi.org/10.3390/agronomy9090556>

Campos-Taberner, M., García-Haro, F.J., Martínez, B., Sánchez-Ruiz, S., Gilabert, M.A. 2019b. Evaluación del potencial de Sentinel-2 para actualizar el SIGPAC de la Comunitat Valenciana. En: *XVIII Congreso de la Asociación Española de Teledetección*. Valladolid, España, 24-27, septiembre. pp 11-14.

Camps-Valls, G., Tuia, D., Bruzzone, L., Benediktsson, J.A. 2013. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE Signal Processing Magazine*, 31(1), 45-54. <https://doi.org/10.1109/MSP.2013.2279179>

Chuvieco, E. 2008. *Teledetección Ambiental. La observación de la Tierra desde el espacio*. Madrid: Ariel.

Cover, T., Hart, P. 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27. <https://doi.org/10.1109/TIT.1967.1053964>

Estrada, J., Sánchez, H., Hernanz, L., Checa, M.J., Roman, D. 2017. Enabling the Use of Sentinel-2 and LiDAR Data for Common Agriculture Policy Funds Assignment. *ISPRS International Journal of Geo-Information*, 6(8), 255. <https://doi.org/10.3390/ijgi6080255>

Friedl, M.A., McIver, D.K., Hodges, J.C., Zhang, X.Y., Muchoney, D., Strahler, A.H., Baccini, A. 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote sensing of Environment*, 83(1-2), 287-302. [https://doi.org/10.1016/S0034-4257\(02\)00078-0](https://doi.org/10.1016/S0034-4257(02)00078-0)

Graves, A., Mohamed, A.R., Hinton, G. 2013. Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6645-6649. <https://doi.org/10.1109/ICASSP.2013.6638947>

Gilabert, M.A., González-Piqueras, J., García-Haro, J., 1997. Acerca de los índices de vegetación. *Revista de teledetección*, 8, 1-10. Disponible en <http://www.aet.org.es/?q=revista8-4>

González-Guerrero, O., y Pons, X., 2020. The 2017 Land Use/Land Cover Map of Catalonia based on Sentinel-2 images and auxiliary data. *Revista de Teledetección*, 55, 81-92. <https://doi.org/10.4995/raet.2020.13112>

Gregorio, A., Jansen, J. 2000. *Land cover classification system (LCCS): Classification concepts and user manual for software version 2*. Roma: FAO.

Griffiths, P., Nendel, C., Hostert, P. 2019. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote Sensing of Environment*, 220, 135-151. <https://doi.org/10.1016/j.rse.2018.10.031>

- Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., Yang, G.Z. 2019. XAI—Explainable artificial intelligence. *Science Robotics*, 4(37). <https://doi.org/10.1126/scirobotics.aay7120>
- Haralick, R.M., Shanmugam, K., Dinstein, I.H. 1973. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, 3(6), 610-621. <https://doi.org/10.1109/TSMC.1973.4309314>
- Haykin, S. 1994. *Neural networks: a comprehensive foundation*. River: Prentice Hall.
- Immitzer, M., Vuolo, F., Atzberger, C. 2016. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sensing*, 8(3), 166. <https://doi.org/10.3390/rs8030166>
- Hochreiter, S., Schmidhuber, J. 1997. Long short-term memory. *Neural computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A. 2017. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 778-782. <https://doi.org/10.1109/LGRS.2017.2681128>
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A. 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 166-177. <https://doi.org/10.1016/j.isprsjprs.2019.04.015>
- Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P. 2016. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2), 645-657. <https://doi.org/10.1109/TGRS.2016.2612821>
- Mikolov, T., Kombrink, S., Burget, L., Černocký, J., Khudanpur, S. 2011. Extensions of recurrent neural network language model. En *2011 IEEE International Conference on acoustics, speech and signal processing (ICASSP)*. Praga, República Checa, 22-27 Mayo. pp. 5528-5531. <https://doi.org/10.1109/ICASSP.2011.5947611>
- Montavon, G., Samek, W., Müller, K.R. 2018. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73, 1-15. <https://doi.org/10.1016/j.dsp.2017.10.011>
- Mou, L., Ghamisi, P., Zhu, X.X. 2017. Deep recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7), 3639-3655. <https://doi.org/10.1109/TGRS.2016.2636241>
- 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. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>
- Ndikumana, E., Ho Tong Minh, D., Baghdadi, N., Courault, D., Hossard, L. 2018. Deep recurrent neural network for agricultural classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sensing*, 10(8), 1217. <https://doi.org/10.3390/rs10081217>
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. 2019. Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195-204. <https://doi.org/10.1038/s41586-019-0912-1>
- Ruiz, L.A., Almonacid-Caballer, J., Crespo-Peremarch, P., Recio, J.A., Pardo-Pascual, J.E., Sánchez-García, E. 2020. Automated classification of crop types and condition in a mediterranean area using a fine-tuned convolutional neural network. En *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*. Niza, Francia, 31 Agosto - 2 Septiembre (en línea). pp. 1061-1068. <https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-1061-2020>
- Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., Müller, K.R. (Eds.). 2020. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Cham: Springer Nature. <https://doi.org/10.1007/978-3-030-28954-6>
- Schmedtmann, J., Campagnolo, M.L. 2015. Reliable crop identification with satellite imagery in the context of common agriculture policy subsidy control. *Remote Sensing*, 7(7), 9325-9346. <https://doi.org/10.3390/rs70709325>
- Schuster, M., Paliwal, K.K. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11), 2673-2681. <https://doi.org/10.1109/78.650093>
- Sitokonstantinou, V., Papoutsis, I., Kontoes, C., Lafarga Arnal, A., Armesto Andrés, A.P., Garraza Zurbano, J.A. 2018. Scalable parcel-based crop identification scheme using sentinel-2 data time-series for the monitoring of the common agricultural policy. *Remote Sensing*, 10(6), 911. <https://doi.org/10.3390/rs10060911>
- Story, M., Congalton, R.G. 1986. Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing*, 52(3), 397-399.

- Tapsall, B., Milenov, P., Tasdemir, K. 2010. Analysis of RapidEye imagery for annual landcover mapping as an aid to European Union (EU) common agricultural polic. En *ISPRS Technical Commission VII Symposium – 100 Years ISPRS*. Viena, Austria, 5-7 Julio. pp. 568-573.
- Van Den Oord, A., Kalchbrenner, N., Kavukcuoglu, K. 2016. Pixel recurrent neural networks. En *Proceedings of the 33rd International Conference on International Conference on Machine Learning*. New York, EEUU., 20-22 Junio. pp. 1747-1756.
- Vuolo, F., Neuwirth, M., Immitzer, M., Atzberger, C., Ng, W.T. 2018. How much does multi-temporal Sentinel-2 data improve crop type classification? *International Journal of Applied Earth Observation and Geoinformation*, 72, 122-130. <https://doi.org/10.1016/j.jag.2018.06.007>
- Wardlow, B.D., Egbert, S.L., Kastens, J.H. 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. *Remote Sensing of Environment*, 108(3), 290-310. <https://doi.org/10.1016/j.rse.2006.11.021>
- Watson, R.T., Noble, I.R., Bolin, B., Ravindranath, N.H., Verardo, D.J., Dokken, D.J. 2000. *Land use, land-use change and forestry: a special report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.
- Zhan, X., Defries, R., Townshend, J.R.G., Dimiceli, C., Hansen, M., Huang, C., Sohlberg, R. 2000. The 250 m global land cover change product from the Moderate Resolution Imaging Spectroradiometer of NASA's Earth Observing System. *International Journal of Remote Sensing*, 21(6-7), 1433-1460. <https://doi.org/10.1080/014311600210254>
- Zhang, L., Zhang, L., Du, B. 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2), 22-40. <https://doi.org/10.1109/MGRS.2016.2540798>
- Zhong, L., Hu, L., Zhou, H. 2019. Deep learning based multi-temporal crop classification. *Remote sensing of environment*, 221, 430-443. <https://doi.org/10.1016/j.rse.2018.11.032>
- Zhu, Z., Woodcock, C.E. 2014. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sensing of Environment*, 144, 152-171. <https://doi.org/10.1016/j.rse.2014.01.011>
- Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., Fraundorfer, F. 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8-36. <https://doi.org/10.1109/MGRS.2017.2762307>