

# **Ensayo del uso de la inteligencia artificial en la asignación automática de hábitats (LPEHT) a inventarios de vegetación**



Xavier Font  
Javier Peralta  
Raul Pérez  
Albert Ferré

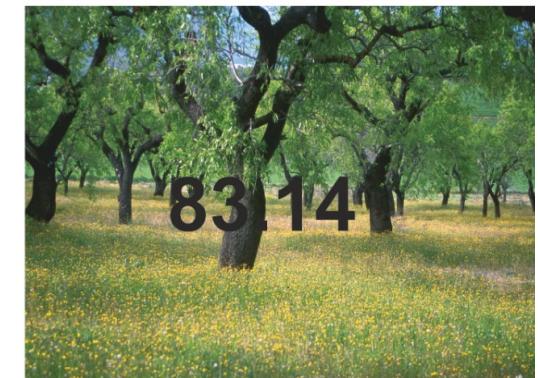


**upna**

Universidad Pública de Navarra  
Nafarroako Unibertsitate Publikoa

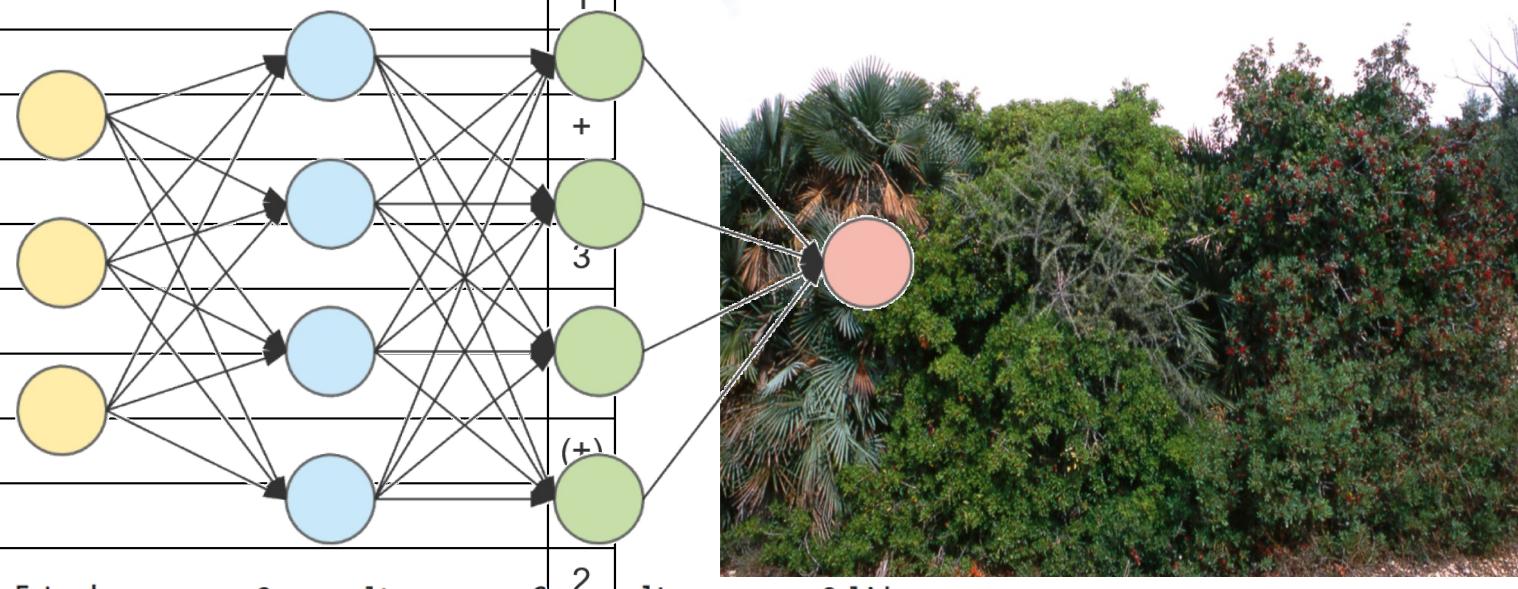


UNIVERSITAT DE  
BARCELONA



**El objetivo principal** de este ensayo es entrenar una IA (basada en redes neuronales artificiales) para obtener un modelo que sirva como una herramienta de reconocimiento automático de hábitats de la Lista Patrón Española de Hábitats Terrestres a partir de inventarios fitosociológicos

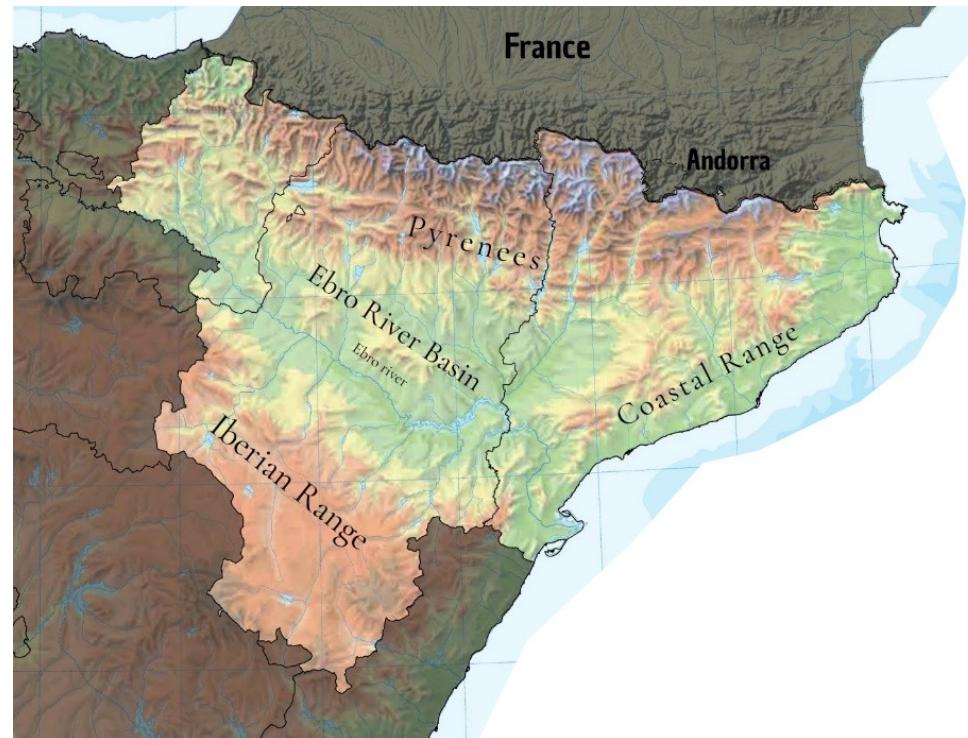
		1
<i>Chamaerops humilis</i> L.		1
<i>Rhamnus lycioides</i> L.		1
<i>Olea europaea</i> L.		
<i>Ceratonia siliqua</i> L.	+	
<i>Pistacia lentiscus</i> L.	3	
<i>Quercus coccifera</i> L.		
<i>Rubia peregrina</i> L.		
<i>Clematis flammula</i> L.		
<i>Asparagus acutifolius</i> L.	(+)	
<i>Euphorbia characias</i> L.		
<i>Myrtus communis</i> L.	2	
<i>Ampelodesmos mauritanicus</i> (Poir.) T. Durand & Schinz [sub]	+	
<i>Brachypodium retusum</i> (Pers.) P. Beauv. [sub]	1	
<i>Rosmarinus officinalis</i> L.	+	



## MATERIAL Y MÉTODOS

### Área de estudio

Las comunidades autónomas de **Aragón** y **Cataluña**, que en conjunto suman una superficie de 79.805 Km<sup>2</sup> (aproximadamente un 16% del territorio español).



Se trata de una zona bien estudiada desde el punto de vista fitosociológico y con una cartografía de los hábitats completa para el territorio catalán y bastante avanzada para Aragón

Disponer de la cartografía y sobre todo de **buenos manuales** de interpretación de los hábitats han sido determinantes para elegir estas dos comunidades autónomas para este estudio metodológico piloto. Ambos territorios han desarrollado clasificaciones de los hábitats desde hace más de dos décadas, y disponen de pasarelas de equivalencias entre distintas clasificaciones (CORINE Biotopes, EUNIS, HIC).



Lista de Hábitats de Aragón, versión 4.11, informe técnico de J.L. Benito (2012), Gobierno de Aragón

Actualización de la Leyenda del Mapa de Hábitats de Aragón, informe técnico de J. Peralta (2024); Gobierno de Aragón.

# Inventarios fitosociológicos

Para este territorio están disponibles, en SIVIM, 35.120 inventarios de vegetación, que incluyen datos de 3.852 especies de flora.

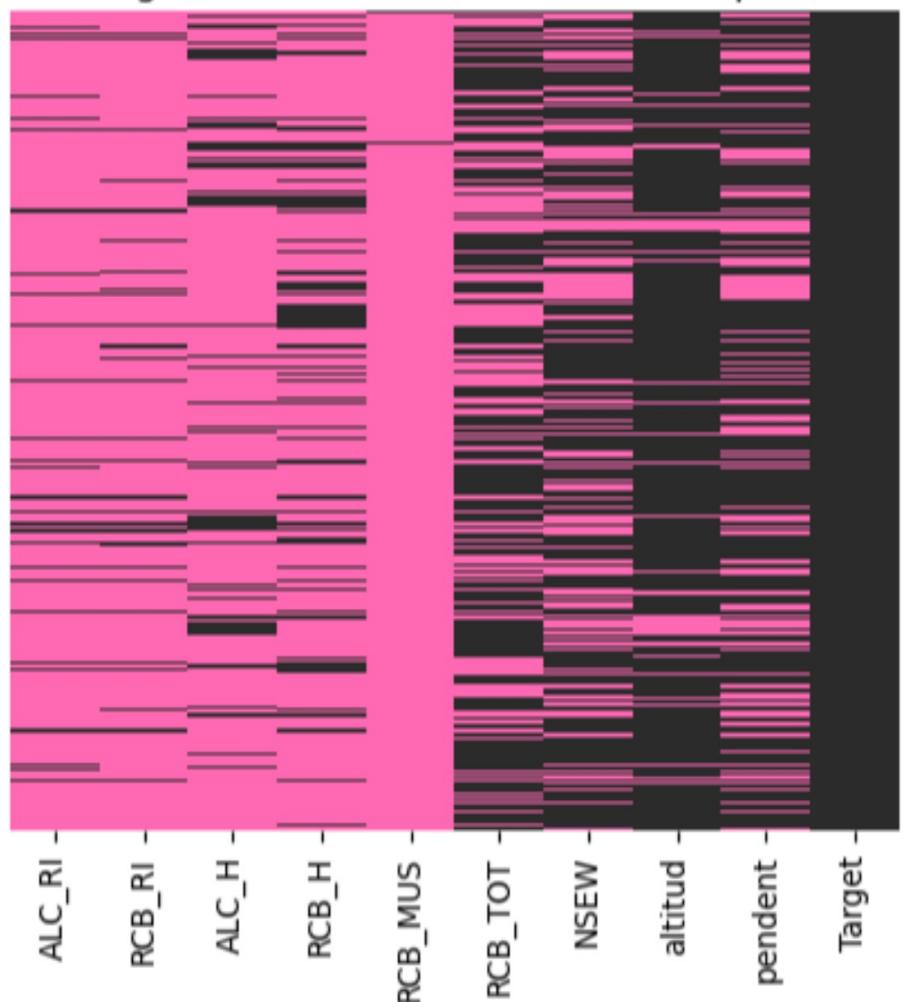
The image shows the SIVIM (Sistema de Información de la Vegetación Ibérica y Macaronésica) website. The header features the logo 'SIVIM' and the text 'Sistema de Información de la Vegetación Ibérica y Macaronésica'. Below the header is a photograph of a forest floor with purple flowers. The main content area includes a sidebar with links to 'Vegetación', 'Información sobre el proyecto', 'Novedades', 'Datos estadísticos', 'Descarga de tesauros', and 'B-VegAna'. The main area has tabs for 'Búsqueda sobre Mapa', 'Búsqueda por comunidad', and 'Búsqueda por obra'. Below these is a map of the Iberian Peninsula and surrounding regions, with labels for Portugal, Spain, Andorra, and the Balearic Islands. Cities like Porto, Lisboa, Madrid, Sevilla, Valencia, and Casablanca are marked. A legend in the top right corner indicates different vegetation types. The 'Mapa' tab is currently selected.

# Dos tipos de matrices de datos: Especies + recubrimiento / variables ecológicas + estructurales

## Unimodal vs Multimodal

~Sphagnum teres	3.
Parnassia palustris	aa+1.1.22*..+..1+..e..
Agrostis capillaris	+. .... *+a..
Phleum alpinum	++...a.+..
Selinum pyrenaeum	++1...+..+..2..+..
Caltha palustris	1...+..1+*..+..++..
Lychnis flos-cuculi	+. .... +..
~Calliergonella cuspidata	2...+..a..
Sedum villosum villosum	+. .... +..+..
~Sphagnum teres	.....
Succisa pratensis	++...*.....2a..
Trifolium pratense	+. .... +a.. .....*
Anthoxanthum odoratum	+. .... ++.....+
~Sphagnum subsecundum	f..
~Drepanocladus exannulatus	+
~Calliergon stramineum	+. .... 2..+
Scirpus cespitosus	.u**....2...*....i..
Calluna vulgaris	.*i...+..+..i....a..
Carex echinata	.a+..a.a22.1++....a..
Leontodon duboisii	.e.1.a.2...++....+a..
Potentilla erecta	.2a...+..1+....i2.a..
Allium schoenoprasum	.+.....
Pedicularis pyrenaica lasiocalyx	.a++.*..1a..+..a+.*+..
Selaginella selaginoides	.1...+..+..
Nardus stricta	.*2..a.2..f.....*
Orchis maculata	.++...+..+..+..
Primula integrifolia	.*...*....*2...+*2..

Missing Environmental Data Before Imputation



Avance de resultados: mejor con solo especies

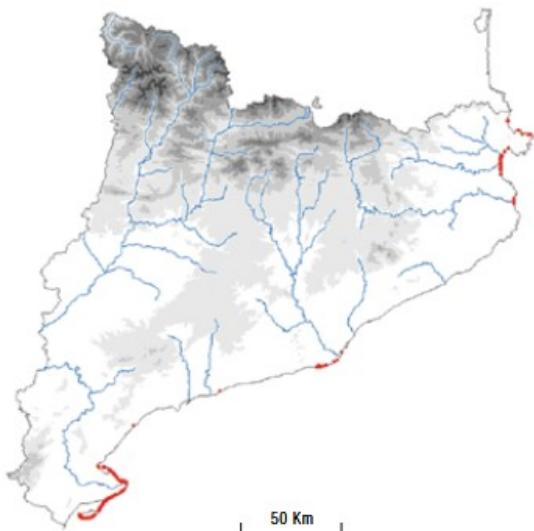
## Selección de inventarios de entreno de la IA

Para realizar la prueba de concepto y estudiar la viabilidad del proyecto se han seleccionado 24.989 inventarios (un 71% de todos los disponibles), representativos de todos los hábitats del territorio. Por criterio experto, se ha asignado un código de hábitat a cada uno de los inventarios considerados, según la clasificación CORINE Biotopes (adaptada para la LPEHT).

SIVIM → **35.120** → Javier Peralta / XavierFont → **24.989** con hábitat

16.2112

## Dunes embrionàries, amb comunitats obertes d'*Elymus farctus* (jull de platja), *Sporobolus pungens*..., de les platges arenoses



### Flora principal

*Elymus farctus* (jull de platja)  
*Sporobolus pungens*  
*Cyperus capitatus* (mansegà marina)  
*Polygonum maritimum* (passacamins mari)



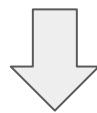
### Sintàxons o altres unitats tipològiques que hi corresponen

*Cypero mucronati-Agropyretum juncei* Kühnholtz-Lordat et Br.-Bl.  
1933

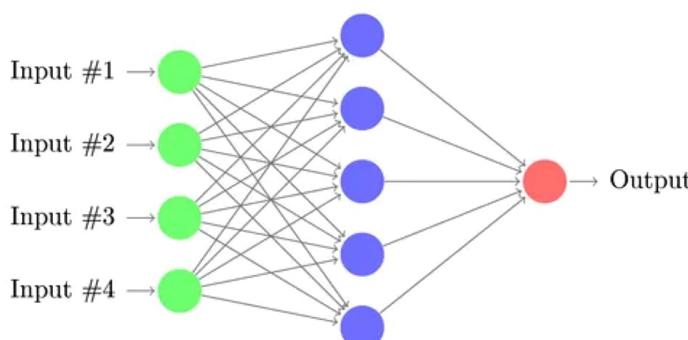
# Tipo de IA utilizada

Entre las distintas técnicas de IA disponibles, se han utilizado las redes neuronales conocidas como Perceptrones Multicapa (MLP) (Haykin, 1999): una red neuronal formada por múltiples capas interconectadas. Consideramos, a priori, que esta técnica puede ser la que dé mejores resultados, ya que ha sido utilizada, con bastante éxito, por Leblanc et al. (2024) con los hábitats EUNIS de Europa.

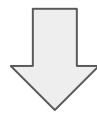
Vectorización



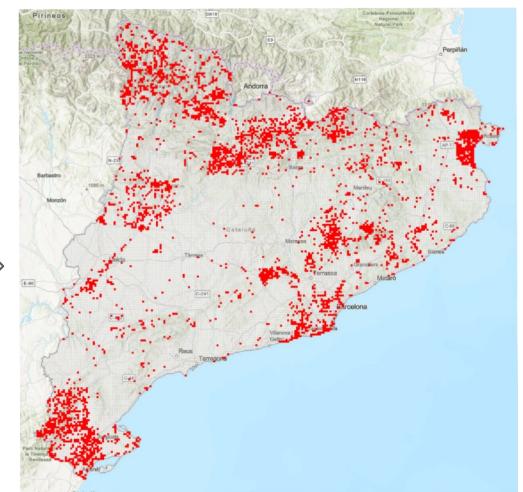
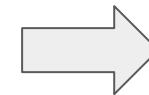
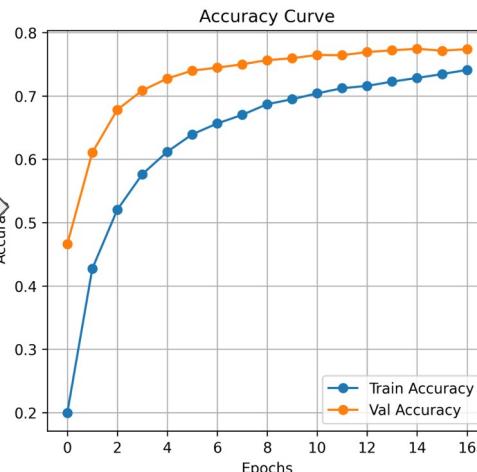
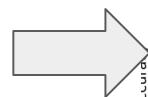
$[sp_1, sp_2, sp_3, \dots, sp_n]$



Entrenamiento MLP



Evaluación



Hemos usado una partición estratificada del conjunto de datos con un 70 % para entrenamiento y un 30% para su evaluación final

## Software utilizado

Todos los ensayos se han llevado a cabo utilizando *Jupyter Notebooks* programados en *Python*. Para el tratamiento y visualización de datos hemos utilizado las librerías *pandas*, *numpy*, *matplotlib* y *seaborn*. Para el entrenamiento de modelos de redes neuronales hemos utilizado las librerías *scikit-learn* y *tensorflow*



## Transformaciones de datos

También hemos realizado pruebas de asignación de dos escalas de recubrimiento para evaluar la importancia y el efecto que tienen las diferentes escalas de recubrimiento en el rendimiento de los modelos. Para ello hemos entrenado el modelo con la escala combinada de Van der Maarel y la escala de porcentaje medio de recubrimiento de Braun-Blanquet

Braun-Blanquet tradicional (dato original)	Van der Maarel (1979)	Coeficiente medio de recubrimiento de Braun-Blanquet
+	2	1
1	3	5
2	5	17
3	7	37
4	8	62
5	9	87

Avance de resultados: mejor con Van der Maarel

## Transformaciones de datos

Para abordar el desequilibrio de clases y la mala representación de algunos hábitats hemos establecido, una vez consultada la literatura especializada, un umbral de un mínimo de 20 inventarios por hábitat y solo incluimos aquellas categorías que cumplen con el número de muestras igual o superior al umbral en los datos de entrenamiento.

Resumen, según los tres niveles de precisión ensayados, y la cantidad de inventarios (muestras) y de hábitats que incluyen.

Precisión en la definición de los Hábitats	Nº de inventarios considerados finalmente en el modelo	Nº total de hábitats	Nº de hábitats considerados en el modelo
Todos los dígitos disponibles	22995	558	289
Nivel 4 LPEHT	24234	284	194
Nivel 3 LPEHT	24854	109	97

En total, por criterio experto y con la máxima precisión, se han encontrado 558 hábitats distintos de la LPEHT. En esta clasificación, basada en CORINE Biotopes, los hábitats están codificados por un código alfanumérico constituido por dos grupos de caracteres separados por un punto.

Se han realizado pruebas de modelado con tres niveles distintos de precisión:

- Todos los caracteres alfanuméricos considerados por los expertos (máxima precisión); 558 hábitats distintos.
- 2 caracteres alfanuméricos, nivel 4 de la LPEHT; 284 hábitats distintos.
- 1 carácter alfanumérico; nivel 3 de la LPEHT; 109 hábitats distintos.
-

# RESULTADOS

	A	B	C	D	E	F	
1	inventario	Top_1_Labels	Top_1_Probabilities	Top_2_Labels	Top_1_Probabilities	Top_3_Labels	Top_3_F
2	P-P08352	@15.721	0.98247	@22.3113	0.00404	@61.12	0.00276
3	S-P00449	@53.62	0.99992	@37.71	0.00001	@62.7	0.00001
4	P-P02358	@42.413	0.99075	@31.42	0.00923	@42.4241	0.00000
5	U-P17309	@44.81341	0.99952	@44.122	0.00010	@15.54	0.00006
6	S-P05339	@31.2261	0.98325	@31.8414	0.01580	@31.237	0.00063
7	T-P14115	@45.3415	0.44302	@45.3411	0.33827	@45.3416	0.21842
8	R-P02825	@31.872	0.99894	@31.8711	0.00096	@31.8712	0.00008
9	R-P03974	@38.23	0.99434	@38.31	0.00562	@38.112	0.00002
10	U-P09239	@37.241	0.77512	@37.242	0.19384	@37.4	0.02947
11	P-P01511	@35.91	0.94926	@35.12141	0.04010	@34.332G1	0.01042
12	P-P10723	@82.41	0.99685	@61.342	0.00058	@23.211	0.00039
13	P-P04681	@36.434	0.91165	@34.32614	0.01681	@34.7135	0.01359
14	S-P02381	@32.42	0.87451	@15.923	0.04227	@32.432	0.03150
15	T-P26975	@15.81	0.99616	@15.619	0.00092	@44.8131	0.00077
16	P-P08623	@62.1111	0.89388	@62.151	0.01974	@61.32	0.01376
17	Q-P06261	@24.53	0.99987	@15.52	0.00006	@82.31	0.00002
18	U-P01466	@61.345	0.99058	@36.434	0.00379	@61.341	0.00194
19	P-P10438	@15.53	0.96238	@15.81	0.01015	@15.51	0.00492
20	T-P37535	@87.2	0.89571	@62.151	0.04325	@62.1111	0.01687
21	R-P01668	@54.28	0.88355	@54.24_(+_54.2A	0.03866	@54.32	0.02448
22	P-P02729	@36.434	0.43188	@34.7135	0.17013	@36.4112	0.04933

## Rendimiento de los modelos, Van der Maarel vs Braun-Blanquet

En primer lugar, realizamos un ensayo para ver cómo afecta al rendimiento del modelo (MLP simple) el uso de distintas escalas de recubrimiento.

Escala de recubrimiento	Exactitud (Accuracy)
Van der Maarel	84,6
Braun-Blanquet	83,9

Tabla comparativa del rendimiento del mismo modelo (MLP simple) entrenado con diferentes escalas de recubrimiento. El valor de exactitud más alto indica un mejor modelo.

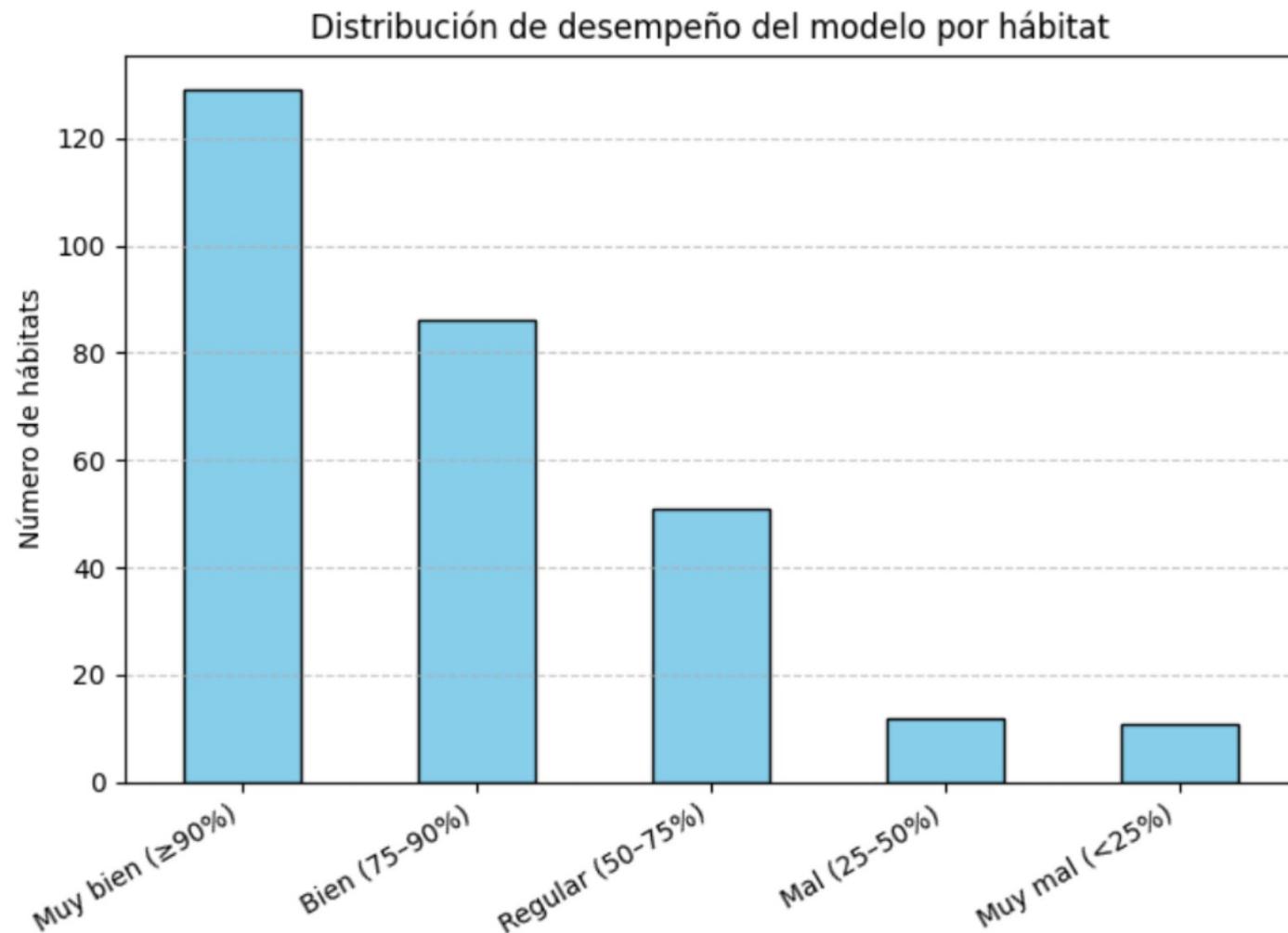
Tabla comparativa del rendimiento del MLP multimodal con diferentes combinaciones de imputación de datos ambientales y estructurales.

Tipo de ensayo	Exactitud (Accuracy)
Con todas las variables menos el recubrimiento muscinal; datos faltantes interpolados con LightGBM	78.64
Todos los datos faltante asignados al valor 0 por defecto	77.85
Todas las variables interpoladas con LightGBM	76.67
Solo los datos geográficos interpolados con <i>LightGBM</i> (sin datos de estructurales)	75.97

Comparación de la exactitud (Accuracy) del modelo MLP simple con diferentes niveles de precisión en la definición de los hábitats. También comparamos el número de inventarios utilizado para el entrenamiento con el filtro de, como mínimo, 20 inventarios por hábitat y, en último lugar, el número total de hábitats. La columna “top 3 accuracy” hace referencia a la exactitud tomando los tres hábitats predichos con la probabilidad asociada más alta

Nivel de precisión Corine	Accuracy	Top 3 Accuracy	Nº de inventarios utilizados	Nº de hábitats discriminados
<b>Nivel 3</b>	<b>90</b>	98,20	24.854	97
<b>Nivel 4</b>	86,38	96,67	24.234	194
<b>Precisión original</b>	84,60	97,22	22.995	289

Distribución de categorías de hábitat según su nivel de exactitud en el modelo MLP simple con máxima precisión. El histograma muestra que la clasificación de los hábitats puede categorizarse en 5 categorías: *accuracy* superior 90%, *accuracy* entre el 75 y el 90%, *accuracy* entre el 50 y el 75%, *accuracy* entre el 25 y el 50%) y *accuracy* inferior al 25%



Entre los tipos de hábitats con mejor reconocimiento destacan los tipos siguientes:

Catorce hábitats de “Matorrales, arbustadas y bosquetes de montaña y de ambientes frescos de baja altitud” (código 31. ).

Doce hábitats de “Bosques caducifolios planifolios” (código 41.).

Doce unidades de “Prados y pastizales alpinos, subalpinos y oromediterráneos” (código 36. ).

Nueve hábitats de “Canchales y campos de bloques “ (código 61.).

Ocho hábitats de “Matorrales y formaciones herbáceas de suelos salinos o yesosos” (código 15.).

## **En el lado opuesto encontramos seis hábitats que el modelo no es capaz de reconocer:**

**34.32614** Pastos calcícolas y mesoxerófilos, con *Astragalus sempervirens*, *Sideritis hyssopifolia*, *Festuca ovina*, *Avenula pratensis*, etc., del piso subalpino (y montano) de los Pirineos. El modelo los confunde con **36.434** Pastos de *Festuca gautieri* subsp. *scoparia* y comunidades análogas, calcícolas y mesoxerófilas, de la alta montaña pirenaica o bien **34.332G1** Pastos basófilos y xerófilos, con *Festuca ovina*, *Avenula iberica*, *Bromus erectus*, *Brachypodium phoenicoides*, *Seseli montanum*, *Teucrium pyrenaicum*, etc., montano-pirenaicos.

**34.7135** Céspedes de *Festuca gautieri*, calcícolas, generalmente en umbrías y pie de roquedos. El modelo los confunde normalmente con **36.434** Pastos de *Festuca gautieri* subsp. *scoparia* y comunidades análogas, calcícolas y mesoxerófilas, de la alta montaña pirenaica.

**34.73** Pastos xerófilos supramediterráneos de *Festuca hystrix* y *Poa ligulata*. El modelo los confunde siempre con **36.435** Pastos oromediterráneos de *Festuca hystrix* y *Poa ligulata*, de las zonas culminales de las montañas calizas ibéricas. La confusión parece bastante razonable ya que comparten dos de las especies más significativas.

41.331 Fresnedas de *Fraxinus excelsior* del Pirineo. El modelo las confunde con diversos tipos de bosques de ribera.

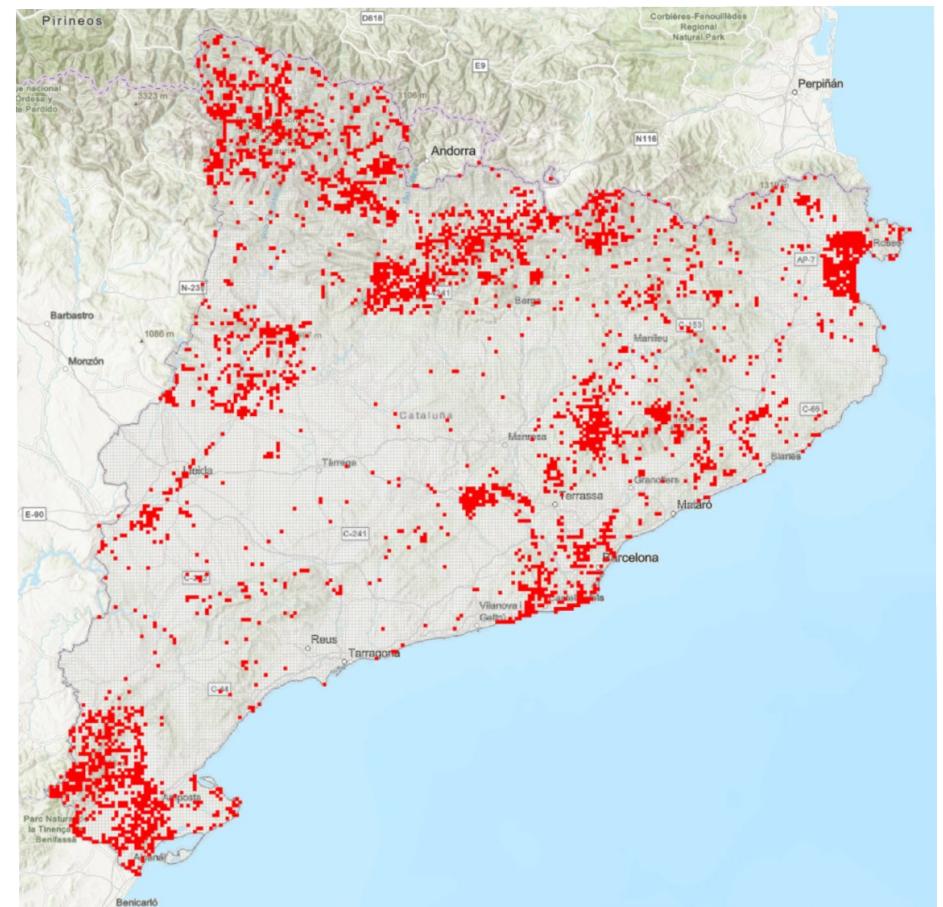
**42.5921** Bosques de *Pinus sylvestris*, calcícolas y xerófilos, del Prepirineo y Pirineo. El modelo los confunde casi siempre con **42.561** Bosques de pino albar (*Pinus sylvestris*), calcícolas y mesófilos, de las umbrías del piso montano de los Pirineos.

**62.143** Rocas calcáreas sombrías, con vegetación cismofítica de musgos y helechos, de las áreas mediterráneas. El modelo las confunde con **62.7** Rocas y muros colonizados por vegetación subnitrófila.

»»» Revisar classficación experta !!!

## Validación del modelo a través del mapa de hábitats de Cataluña

De los 11.019 inventario catalanes con asignación de hábitat según el modelo de máxima precisión (289 hábitats) un total de 5.020 se encuentran en cuadrículas UTM de 1 km que contienen el hábitat asignado al inventario (un 46 %).



Porcentaje de coincidencia (columna "%") igual o superior al 90 % entre el hábitat asignado a los inventarios y la cartografía, para los hábitats que cuentan con más de 10 inventarios originales (columna "Inv.").

LPEHT	Hábitat	Inv.	%
44.81341	Tarayales hipersalinos con <i>Tamarix boveana</i>	11	100
51.1117	Convexidades de turberas dominadas por <i>Sphagnum magellanicum</i>	31	97
36.422	Pastos de <i>Kobresia myosuroides</i> , calcícolas, del piso alpino de los Pirineos	21	95
41.4C	Bosques mixtos con abundancia de arce campestre ( <i>Acer campestre</i> ), mesófilos y eutróficos, de la Cataluña central	63	92
41.1751	Hayedos calcícolas <u>submediterráneos</u> con boj ( <i>Buxus sempervirens</i> )	47	91
45.216	Alcornocales catalanes	21	90

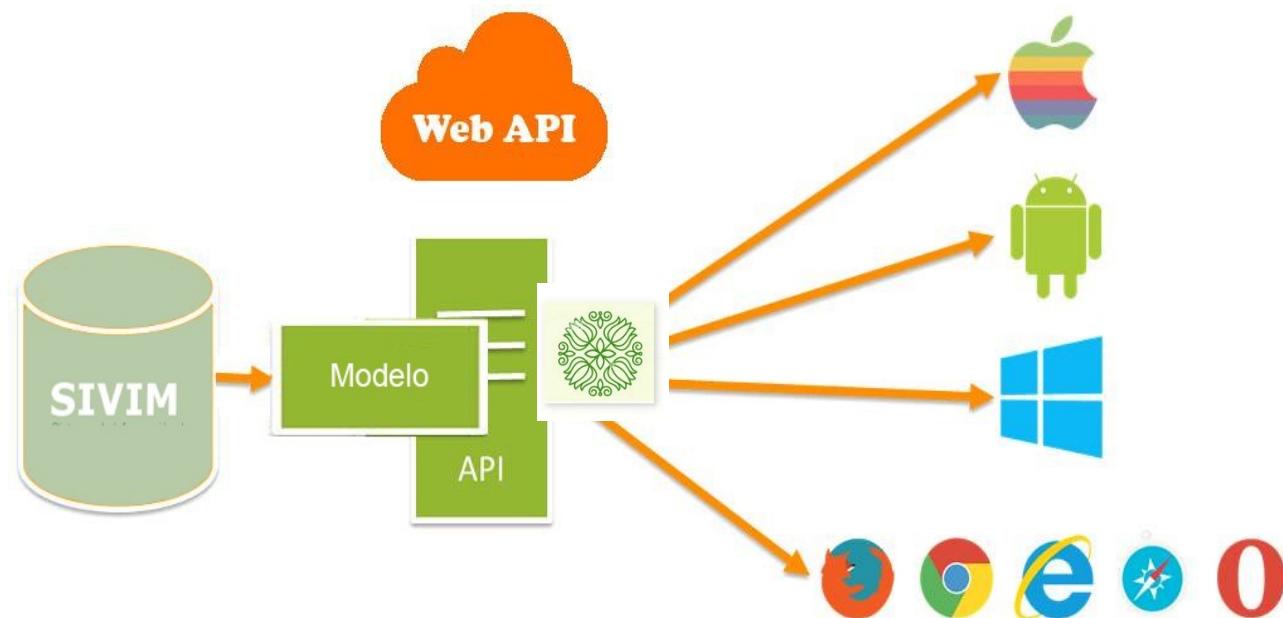
El mayor número de casos de no coincidencia entre el hábitat del inventario y la cartografía se da en aquellas unidades que tienden a ser poco representadas en el mapa, por razones de escala: **comunidades ruderales, orlas espinosas, comunidades dominadas por terófitos**, etc. Se trata de hábitats frecuentes en el territorio y en general bien inventariados, pero que ocupan habitualmente superficies reducidas, difícilmente recogidas en la cartografía, el área mínima de los elementos geométricos del mapa es de 1,5 hectáreas, y solamente se representan superficies menores en forma de puntos cuando se considera que se trata de hábitats o situaciones relevantes para la conservación)

Este valor de concidència del 46 % puede parecer bajo, pero solo es un **1% menor al que obtenemos de comparar las determinaciones de los inventarios por criterio experto** y el mapa de hábitats de Cataluña.

A pesar de las limitaciones de la metodología, pensamos que puede ser útil a la práctica (complemento a la cartografía, ayuda en el trabajo de campo, definición de pasarelas automáticas entre asociaciones y hábitats, etc.)

## ¿En qué trabajamos ahora?

Las redes neuronales generan un modelo que permite clasificar nuevos inventarios de forma rápida y bastante precisa



La API permite **clasificar automáticamente inventarios** en asociaciones (y hábitats en Cataluña y Aragón) y visualizar para cada inventario **hábitat y probabilidad**, junto con las **top-3** predicciones



## SIVIM Relevé IA Classifier

Region Target Upload .txt

Catalunya-Aragó ▾ Habitat Navega... test\_data\_long.txt

Upload and Predict Download CSV IG steps 50 Top N 20 Download Interpretability

**Predictions**

Relevé ID	Prediction Code	Readable Name	Top 1 Probability	Top 3
P-P00002	@35.91		1	@35.91 (1) @35.12141 (0) @34.332G1 (0)
P-P00004	@35.91		1	@35.91 (1) @35.12141 (0) @32.4I (0)

Además, ofrece la descarga de resultados en **Excel** y un módulo de **interpretabilidad** que muestra qué especies influyen más en cada decisión mediante **gráficos e informes** basados en *integrated gradients*.

@31.42 Matorrales de *Rhododendron ferrugineum*, acidófilos, de lugares largamente innivados de la alta montaña alpino-pirenaica

