

**VIII JORNADA  
DE PATRIMONI  
CULTURAL**

*LA INTEL·LIGÈNCIA  
ARTIFICIAL APLICADA  
AL PATRIMONI  
CULTURAL*



Generalitat de Catalunya  
Agència Catalana  
del Patrimoni Cultural

**Intel·ligència artificial i patrimoni cultural:  
Aplicacions actuals i futures perspectives**

Hèctor A. Orengo

Profesor de Recerca ICREA, BSC-CNS

La Intel·ligència Artificial es una camp molt ampli  
Normalment, les aplicacions dedicades al patrimoni cultural s'engloben dins de l'aprenentatge automàtic i l'aprenentatge profund, i es centren en tasques de detecció, identificació, agrupació, classificació i regressió.

Aquests s'apliquen predominantment en tres àrees interrelacionades:

Extracció i anàlisi de dades complexes/massives

Automatització de tasques simples

Predicció de dades

Les tecnologies de IA generativa constitueixen un cas particular

## Dades d'entrenament

El procés basic per l'aplicació de mètodes d'aprenentatge automàtic requereix:

- Dades d'entrenament i validació
- Algoritme
- Computació

A les disciplines relacionades amb el patrimoni cultural disposem de poques dades donada la natura del patrimoni... en molts casos únic!

**Les fonts poden ser extremadament diverses, però es poden classificar en:**

- Textos
- Imatges (imatges de banda única, RGB, multi- i híper-espectrals, geofísica, etc.)
- Àudio i vídeo
- Models 3D de diversos tipus
- Dades sensorials (temperatura, humitat, lidar, etc.)
- Metadades i bases de dades
- Fonts contextuais
- Interaccions d'usuari



## Principals problemes

Només detecta el que ha estat entrenat per detectar amb les fonts que s'han fet servir

Pot estar fent servir uns criteris diferents als teus per executar la feina

Els nostres prejudicis són els del detector

# Aplicacions de mètodes d'aprenentatge automàtic en el camp del patrimoni cultural

## 1. Conservació i Restauració

- Predicció de deteriorament
- Restauració digital
- Anàlisi de materials

## 2. Digitalització i Reconstrucció

- Reconstrucció 3D

## 3. Anàlisi i catalogació

- Classificació automàtica d'artefactes
- Reconeixement de patrons i estils
- Atribució d'obres

## 4. Interpretació i Accés

- Transcripció i traducció de textos antics
- Anàlisi de llenguatge natural en documents històrics

## 5. Detecció de Falsificacions

- Anàlisi forense digital
- Verificació d'autenticitat

## 6. Detecció i monitoratge de tràfic il·lícit d'elements patrimonials

- Identificació d'element patrimonials a la web i xarxes socials
- Detecció de tràfic d'aquest elements
- Identificació d'agents i xarxes de tràfic

## 7. Detecció i monitoratge de jaciments arqueològics

- Detecció de jaciments prèviament desconeguts
- Monitoratge dels jaciments per detectar impactes

## 8. Educació i Disseminació

- Assistents virtuals i xatbots
- Experiències educatives personalitzades
- Sistemes de recomanació cultural

# Aplicacions de mètodes d'aprenentatge automàtic en el camp del patrimoni cultural

## 1. Conservació i Restauració

- Predicció de deteriorament
- Restauració digital
- Anàlisi de materials

## 2. Digitalització i Reconstrucció

- **Reconstrucció 3D**

## 3. Anàlisi i catalogació

- **Classificació automàtica d'artefactes**
- **Reconeixement de patrons i estils**
- Atribució d'obres

## 4. Interpretació de txtos

- Transcripció i traducció de textos antics
- Anàlisi de llenguatge natural en documents històrics

## 5. Detecció de Falsificacions

- Anàlisi forense digital
- Verificació d'autenticitat

## 6. Detecció i monitoratge de tràfic il·lícit d'elements patrimonials

- **Identificació d'element patrimonials a la web i xarxes socials**
- Detecció de tràfic d'aquest elements
- Identificació d'agents i xarxes de tràfic

## 7. Detecció i monitoratge de jaciments arqueològics

- **Detecció de jaciments prèviament desconeguts**
- **Monitoratge dels jaciments per detectar impactes**

## 8. Educació i Disseminació

- Assistents virtuals i xatbots
- **Experiències educatives personalitzades**
- **Sistemes de recomanació cultural**

Aprenentatge automàtic per la  
detecció de jaciments arqueològics

**VIII JORNADA  
DE PATRIMONI  
CULTURAL**

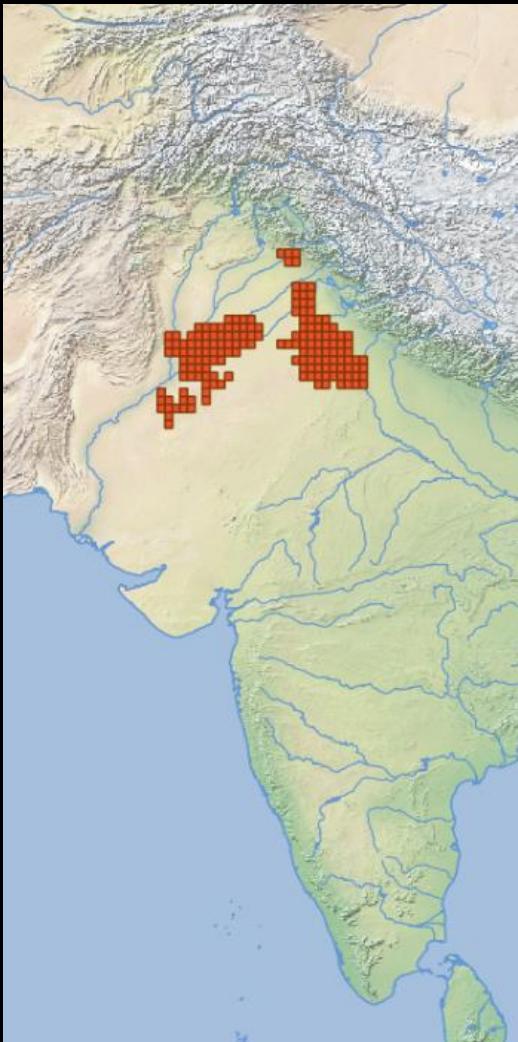
*LA INTEL·LIGÈNCIA  
ARTIFICIAL APLICADA  
AL PATRIMONI  
CULTURAL*

## Mapes Sol d'1 polzada

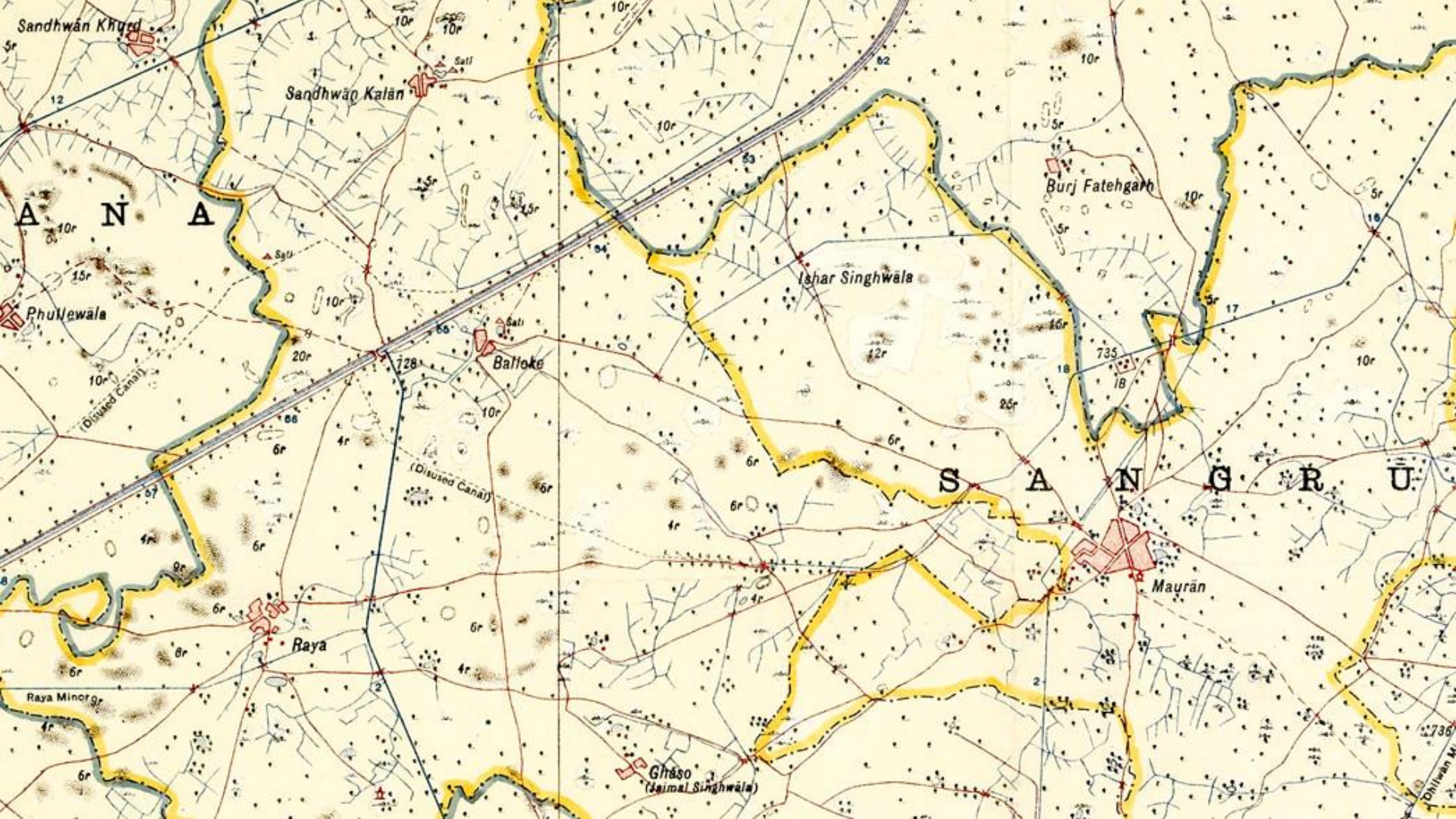
- Cobreixen la major part del país
- Anteriors a les modificacions paisatgístiques a gran escala del segle 20
- Es representen milers de túmuls corresponents a jaciments arqueològics desapareguts i altres elements d'interès arqueològic

## Tanmateix:

- Molt de temps per extreure manualment
- Difícil cobrir tota la zona sistemàticament
- Incoherències en les característiques extretes per diferents persones







Q Search location








Layers



Georeferencing Progress 

Zoom level 0 - 21

COLOR BASED ON  
status

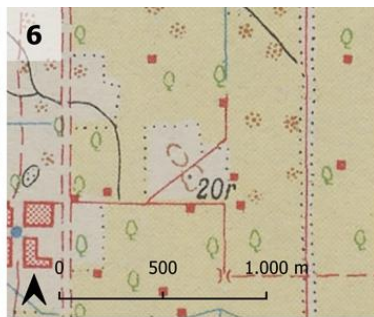
- NOT ORDERED
-  ORDERED
-  SCANNED
-  IN PROGRESS
-  GEOREFERENCED
-  MISSING / PROBLEMATIC

© CARTO, © OpenStreetMap contributors

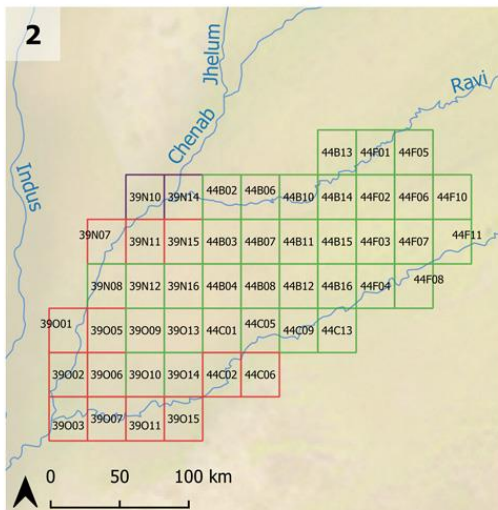




Tibba Bahawal Wah (5-8 C. AD)



Lohama Lal Tibba (3300-1400 BC)



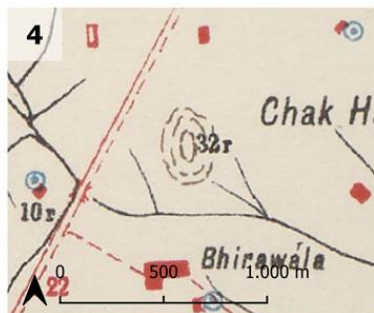
Test area SoI 1 inch/1 mile map sheets

Editions used in the test

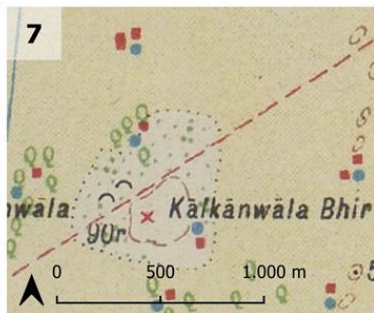
1900s editions

1930s editions

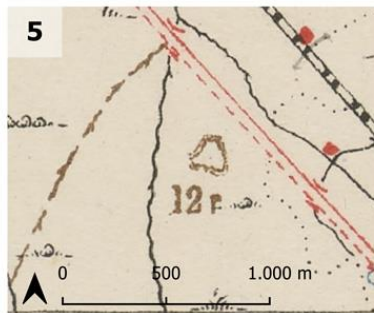
1940s editions (not used)



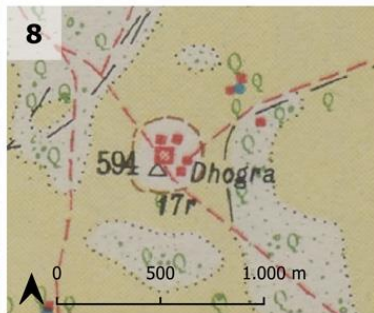
Nawan Shehr Bhir (8-15 C. AD)



Kalkanwala Bhir (8-16 C. AD)



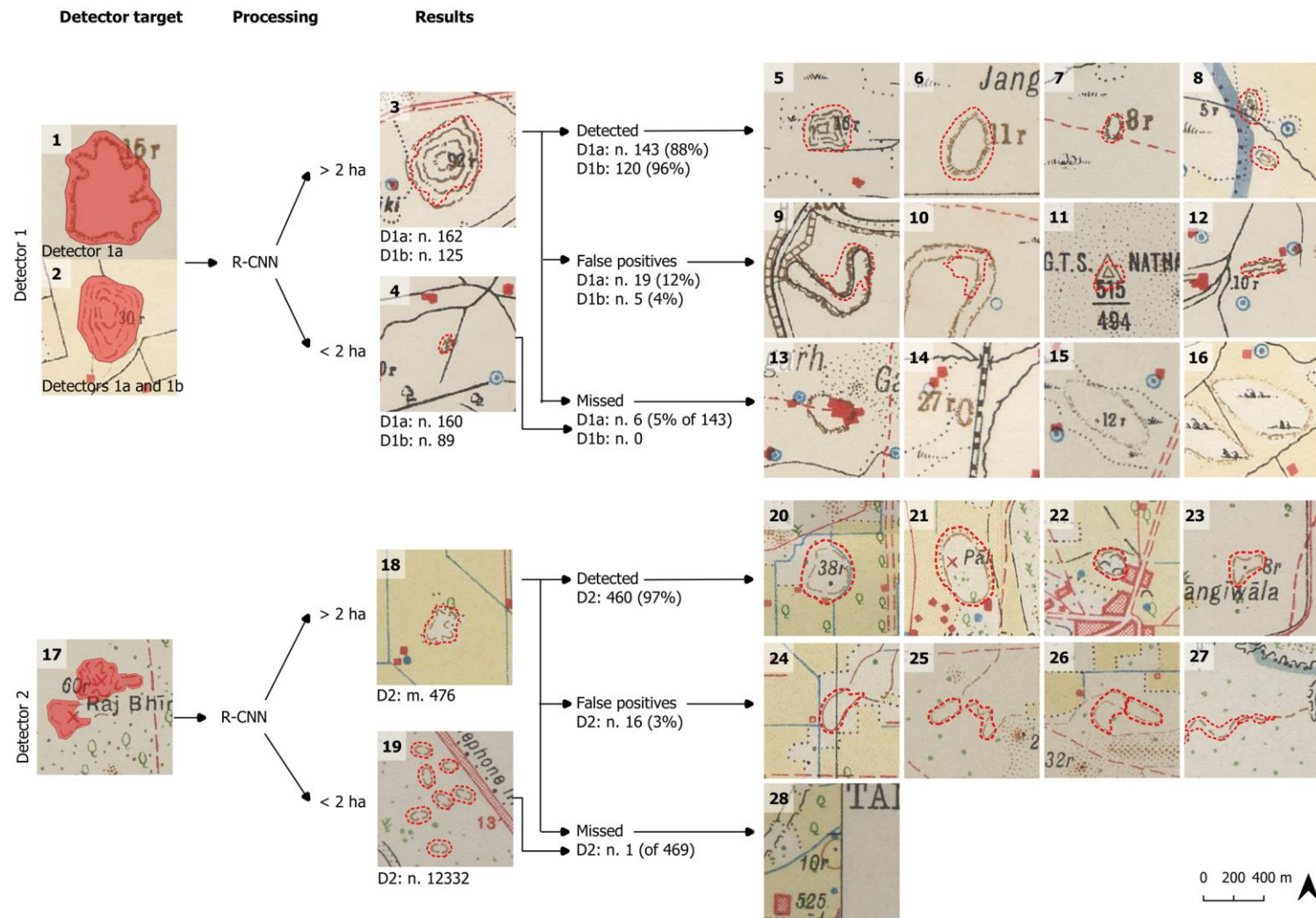
Mithu Di Basti (8-16 C. AD)



Fort of Burj Doruara (16-18 C. AD)



Garcia, A.; Orenge, H.A.; Lawrence, D.; Philip, G.; Hopper, K. & Petrie, C.A. 2020. Potential of deep learning segmentation for the extraction of archaeological features from historical map series. *Archaeological Prospection*.

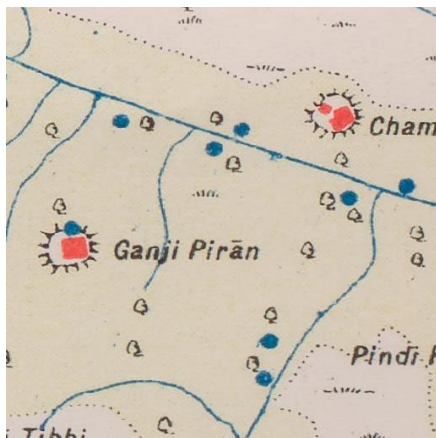


# Millora de la detecció

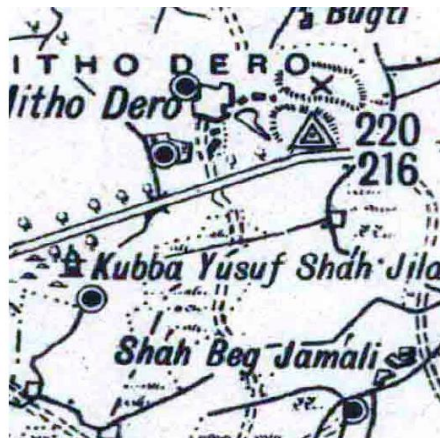
- Reducció significativa de la capacitat de detecció quan s'utilitzen mapes diferents d'aquells amb els quals es va entrenar l'algoritme
- Originalment es va utilitzar una plataforma d'aprenentatge automàtic de detecció pròpia:
  - No és gratuït
  - No reproduïble
  - No va ser possible proporcionar aquesta eina a les parts interessades
  - L'algoritme no es va poder modificar per adaptar-lo a la naturalesa específica de les fonts
  - L'algoritme no s'ha pogut modificar per filtrar falsos positius segons la forma o l'entorn del paisatge
- Era necessari desenvolupar els nostres propis detectors

# Aprentatge de currículum per a la detecció automatitzada de jaciments en mapes històrics del Sol

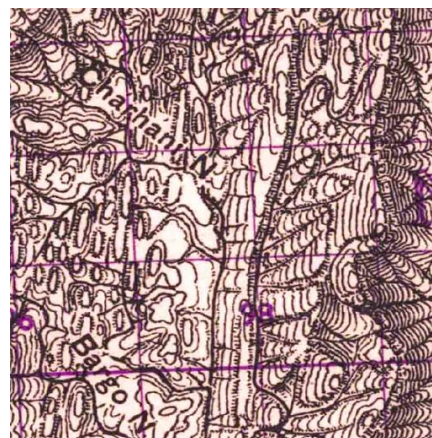
El principal problema era detectar jaciments en mapes molt variables:



(a)



(b)



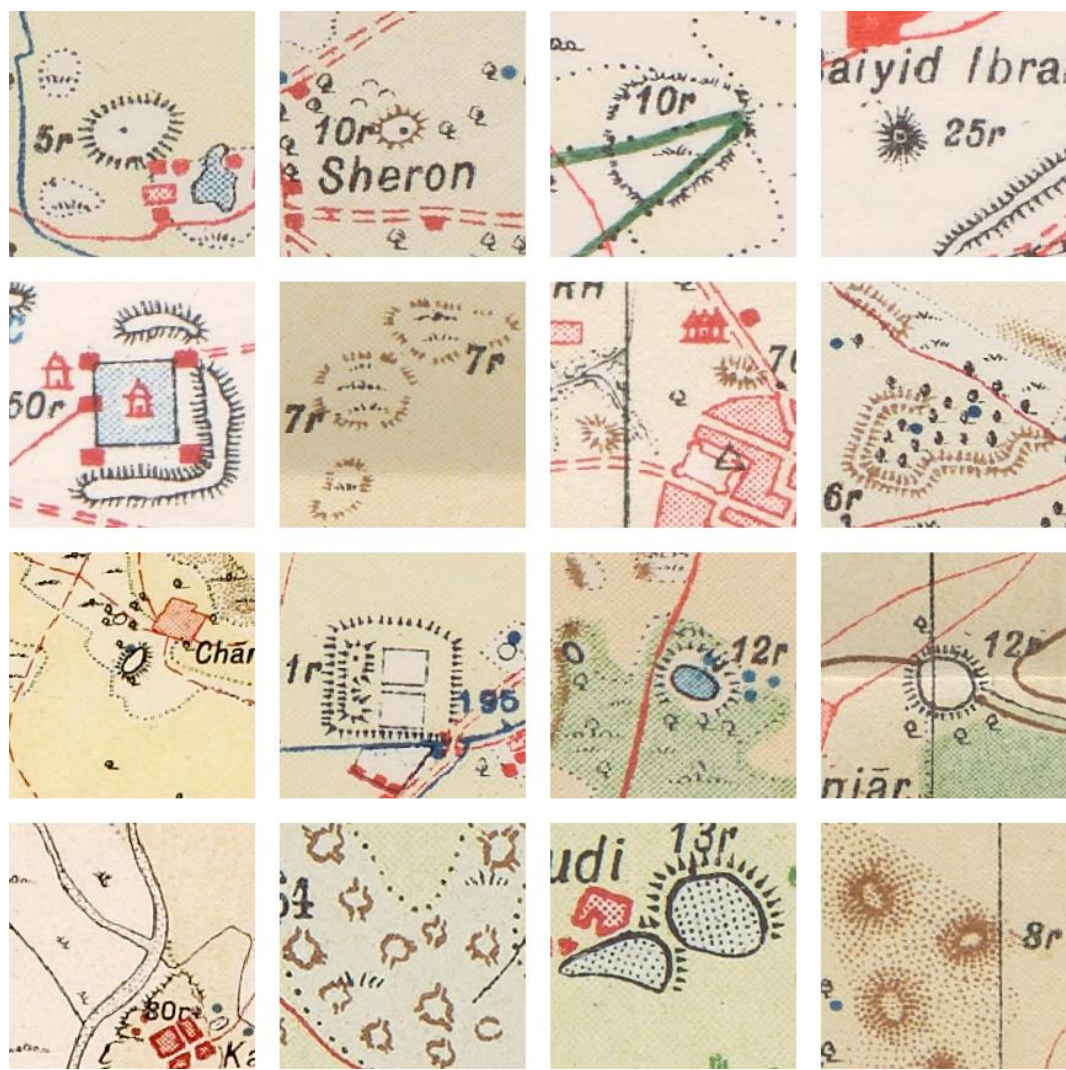
(a)



(b)

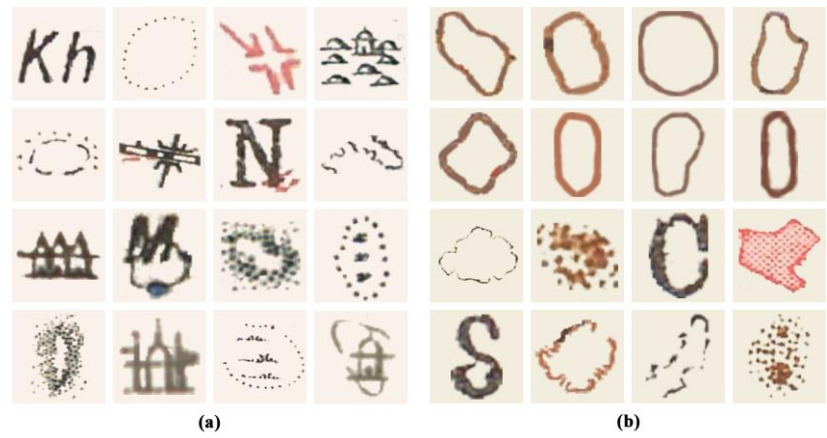
Iban Berganzo-Besga, Hector A. Orenco, Felipe Lumbreras, Aftab Alam, Rosie Campbell, Petrus J. Gerrits, Jonas Gregorio de Souza, Afifa Khan, María Suárez-Moreno, Jack Tomaney, Rebecca C. Roberts, Cameron A. Petrie. 2023. Curriculum Learning-based Strategy for Archaeological Mound Features Detection from Historical Maps in Low-Density Areas in India and Pakistan. *Scientific Reports*, 23: 11257





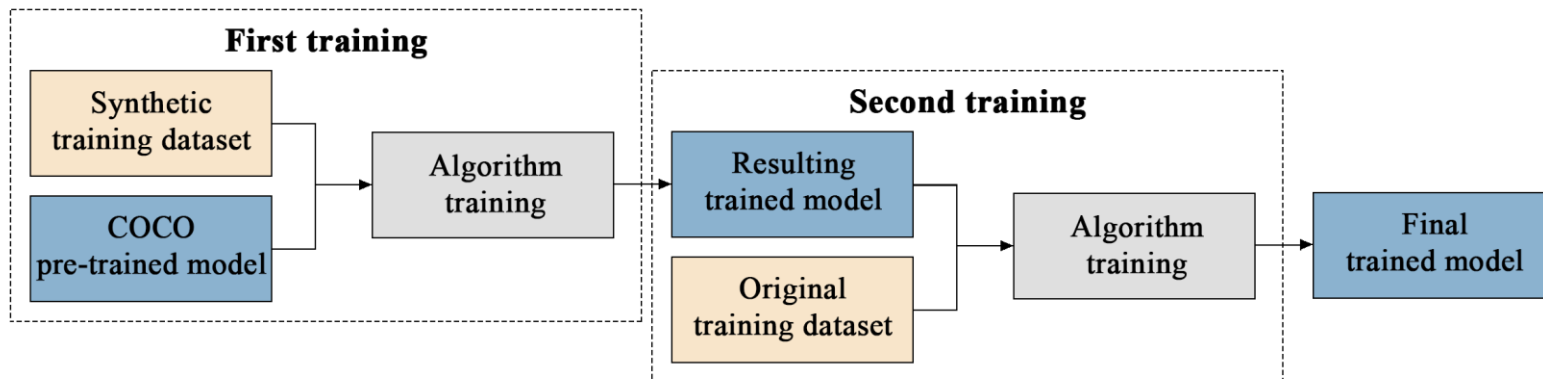
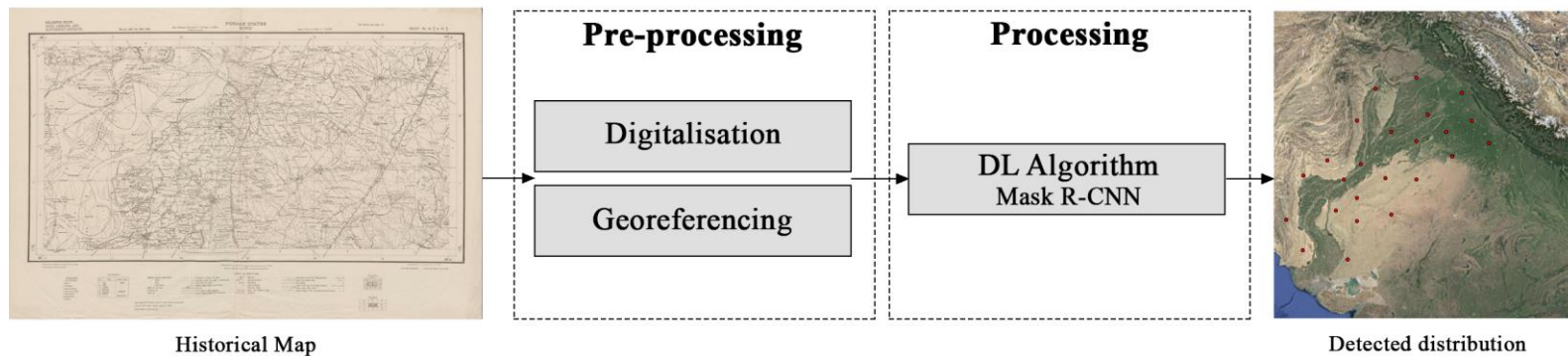
Problemes:

Alta diversitat de formes  
 Presència d'elements a l'interior dels monticles que també es van detectar  
 Les característiques topogràfiques també es podrien representar com a monticles

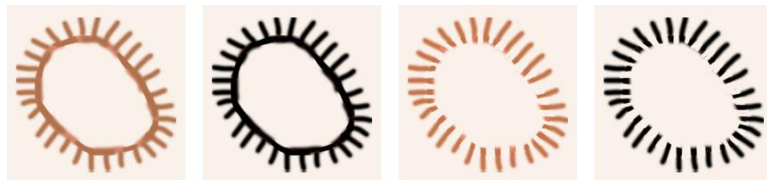


(a) (b)

# Aprentatge de currículum per a la detecció automatitzada de jaciments en mapes històrics del Sol



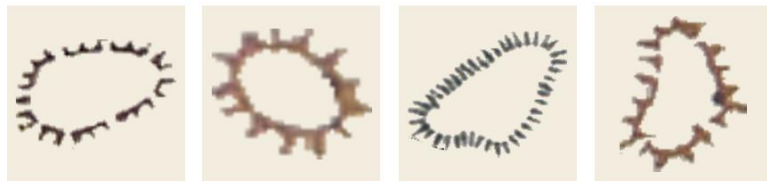
# Dades sintètiques i augment de dades



(a)



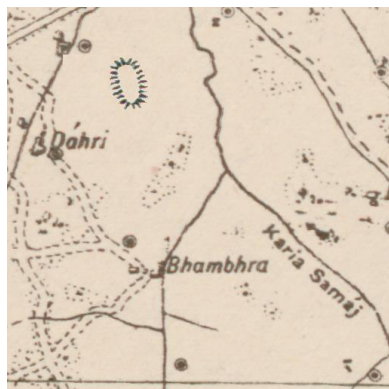
(c)



(b)



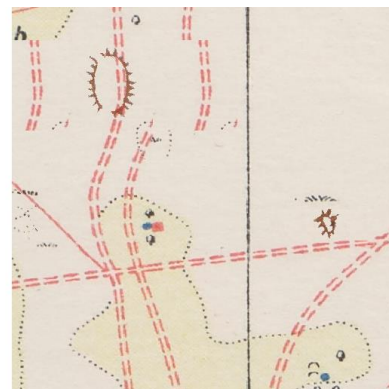
(d)



(a)



(b)



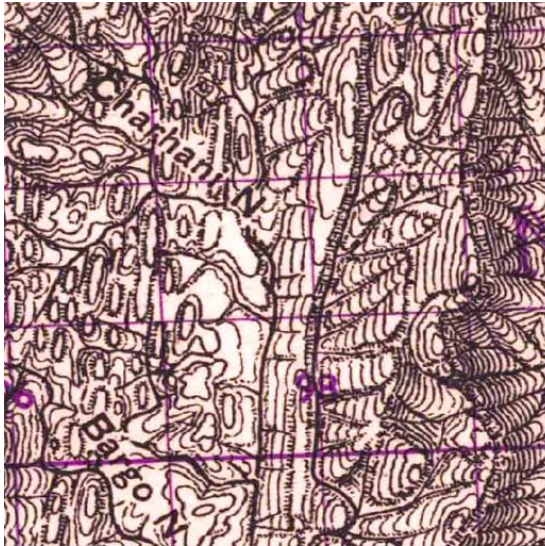
(c)



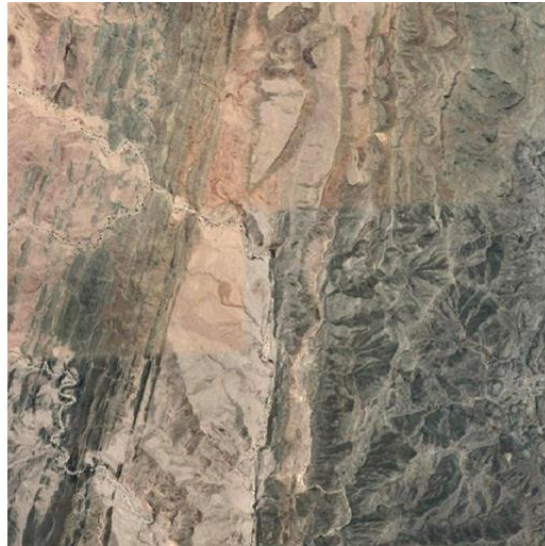
# Aprentatge de currículum per a la detecció automatitzada de jaciments en mapes històrics del Sol

## Filtratge de falsos positius:

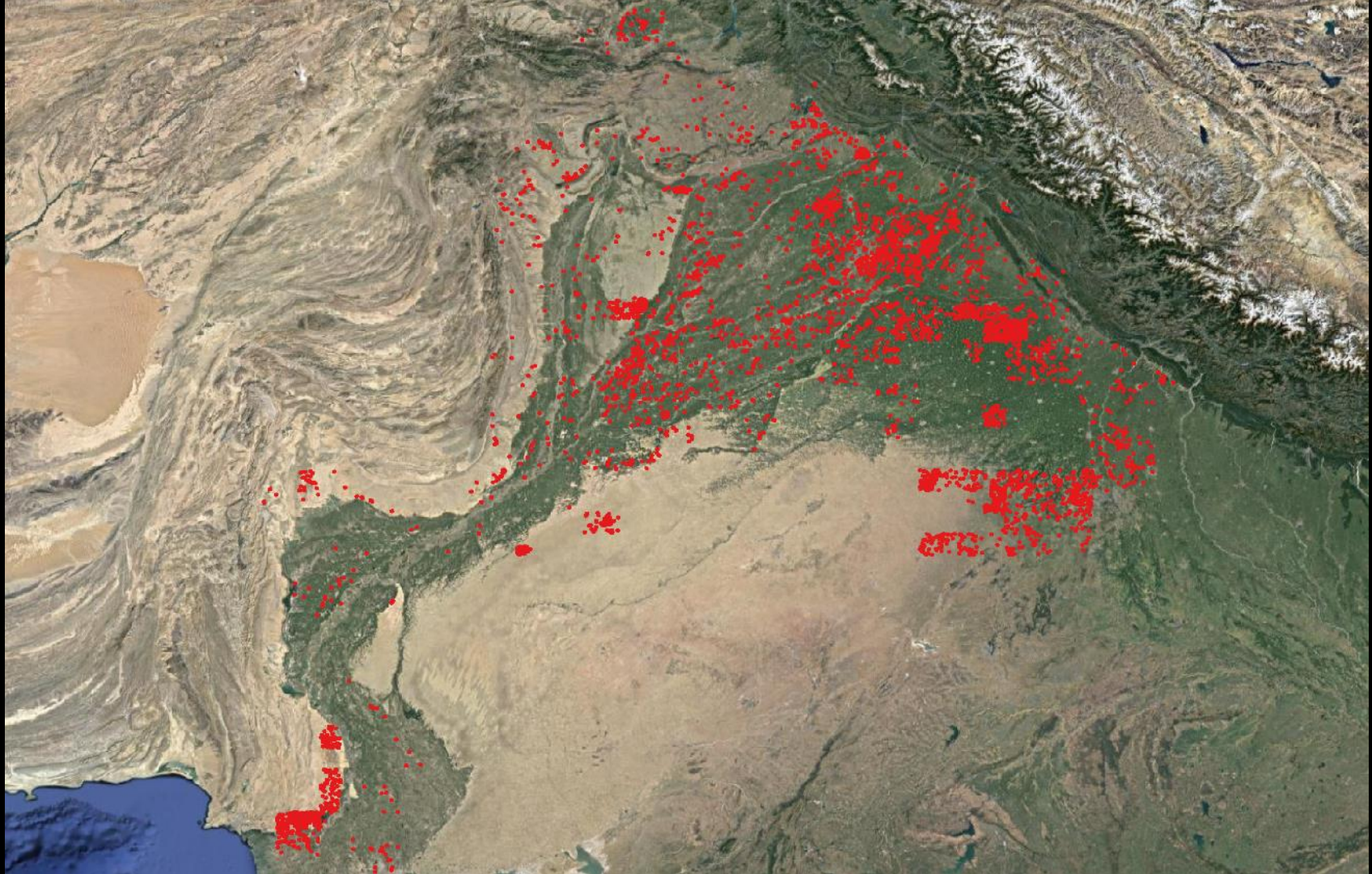
- Basat en la mida (es van filtrar els monticles més petits, que no podien correspondre amb els monticles)
- Topografia (talussos i cims continus corresponents a la topografia natural)
- Basat en formes (anàlisi de taques utilitzat per eliminar monticles allargats corresponents a dunes)



(a)



(b)

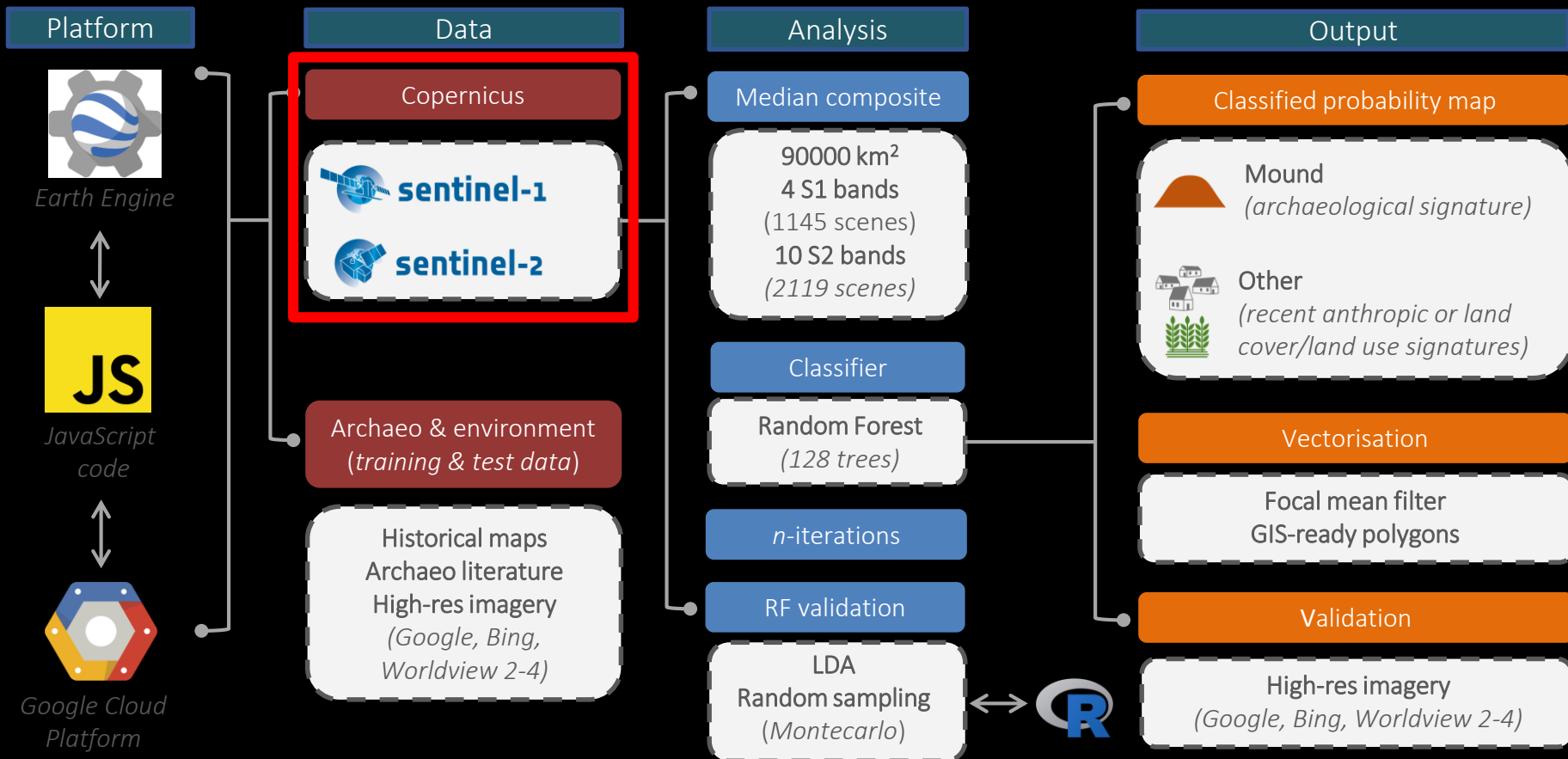


















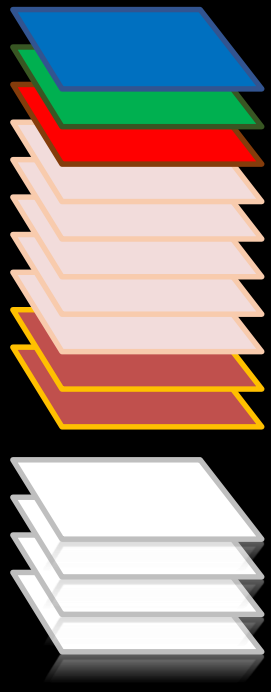
Detecció d'aprenentatge automàtic multitemporal i multifont de jaciments arqueològics

Orengo et al. 2020. Automated detection of archaeological mounds using machine learning classification of multi-sensor and multi-temporal satellite data. *PNAS*, 117(31): 18240-18250.

# Methodological workflow



Type	Source	Bands	Proxy	m/px		
Passive	Reflectance Sentinel 2 multitemp cloud-free Averages (MuTeC-FAv)	B	Visible	Colour		10
		G				10
		R				10
		RE1	Red Edge	Vegetation		20
		RE2				20
		RE3				20
		NIR				10
		RE4	Short Wave Infrared	Chemical composition		20
		SWIR1				20
		SWIR2				20
Active	Sentinel 1 MT-med	Cb A-VV	SAR	Physical/material Soil penetration		10
		Cb A-VH				10
		Cb D-VV				10
		Cb D-VH				10



1105 SAR images (4 bands x image)

- 1) Removal of low intensity noise and invalid data on scene edges
- 2) Thermal noise removal,
- 3) Radiometric calibration
- 4) Terrain correction using SRTM 30
- 5) Speckle filtering by aggregation and median values



Big Data

Data cleaning

3201 Multispectral images (10 bands x image)

- 1) Cloud filtering using the cirrus and cloud band
- 2) Mean of multitemporal stacks seasonally grouped



Big Data

Data cleaning

Training data

- Historical maps
- Survey
- Arch literature
- High-res imagery



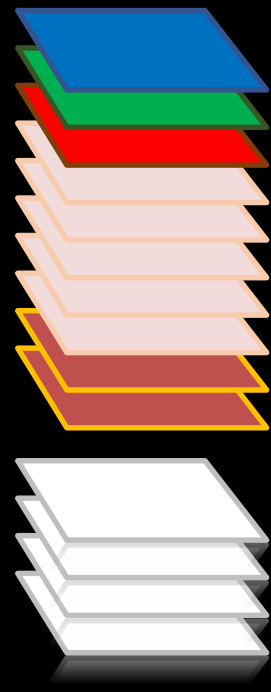
Train ML classifier

- RF
- 300 trees
- Probability

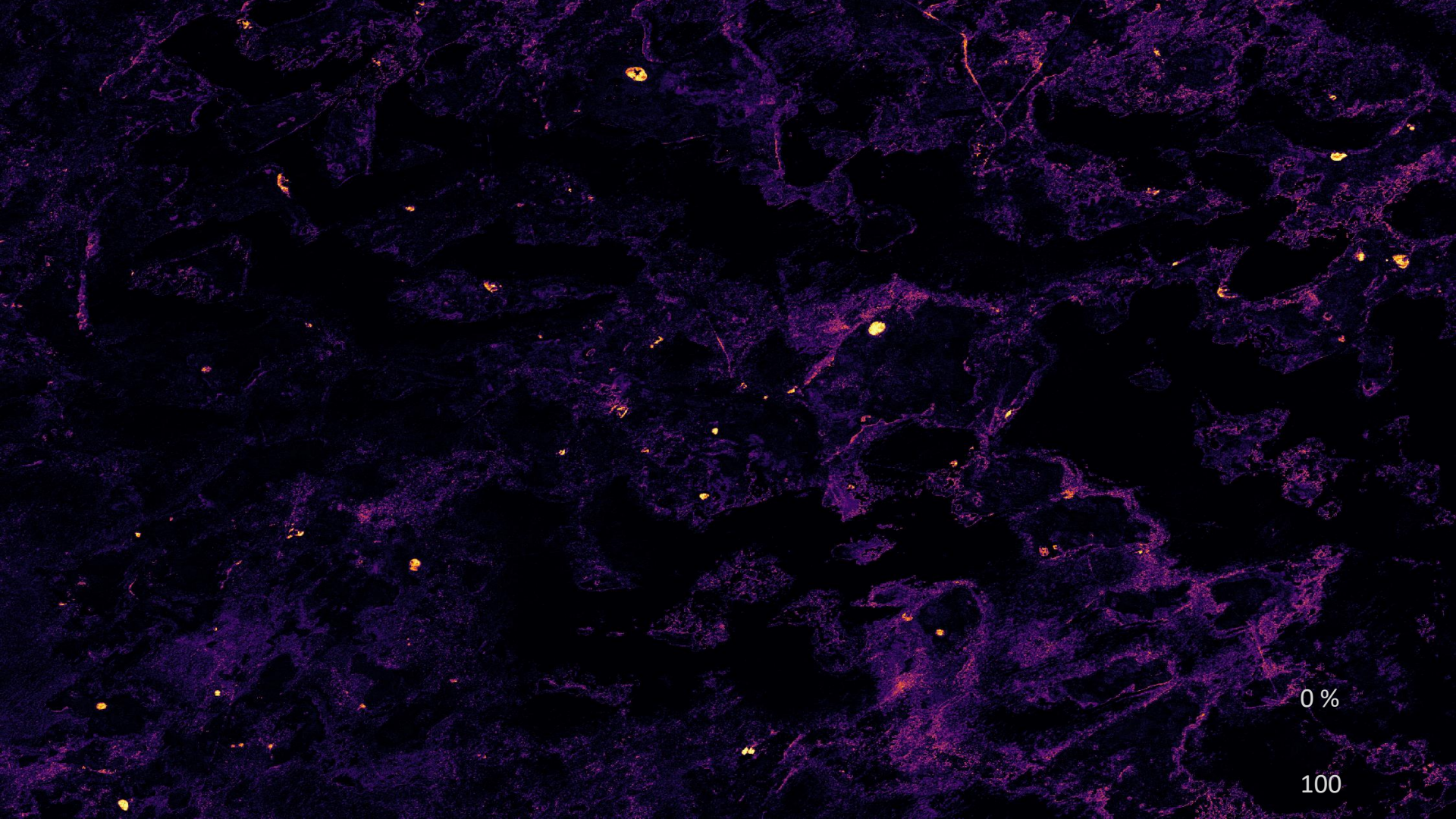


Apply classifier

- n iterations

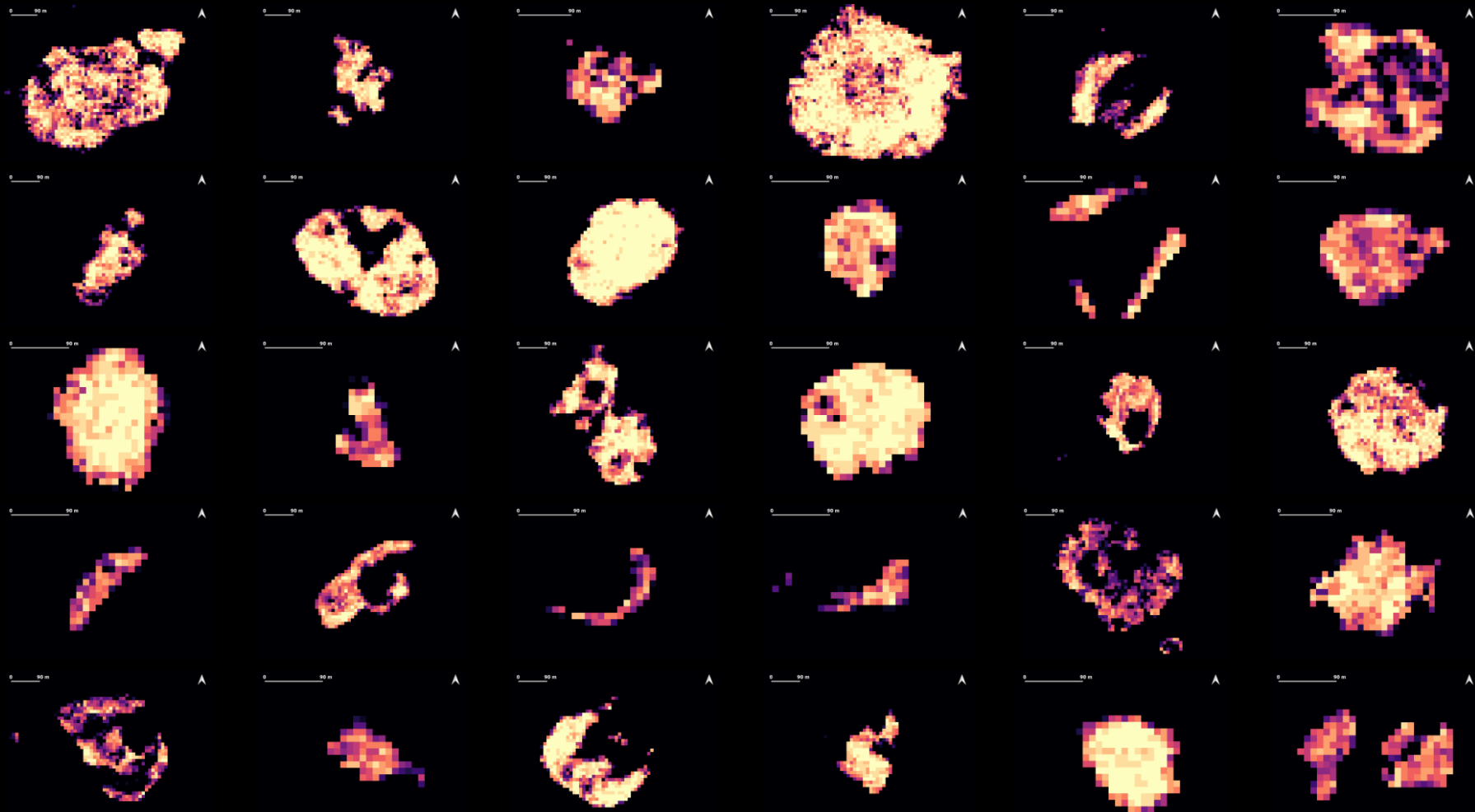


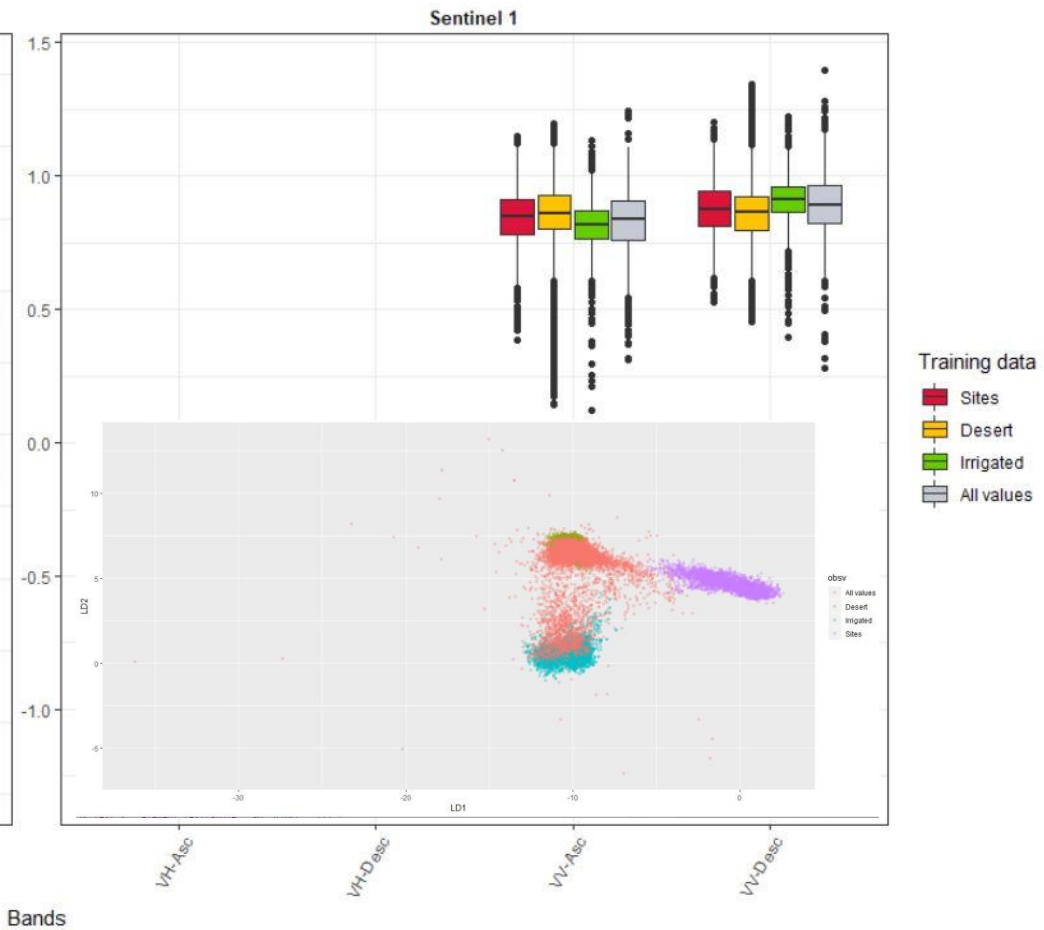
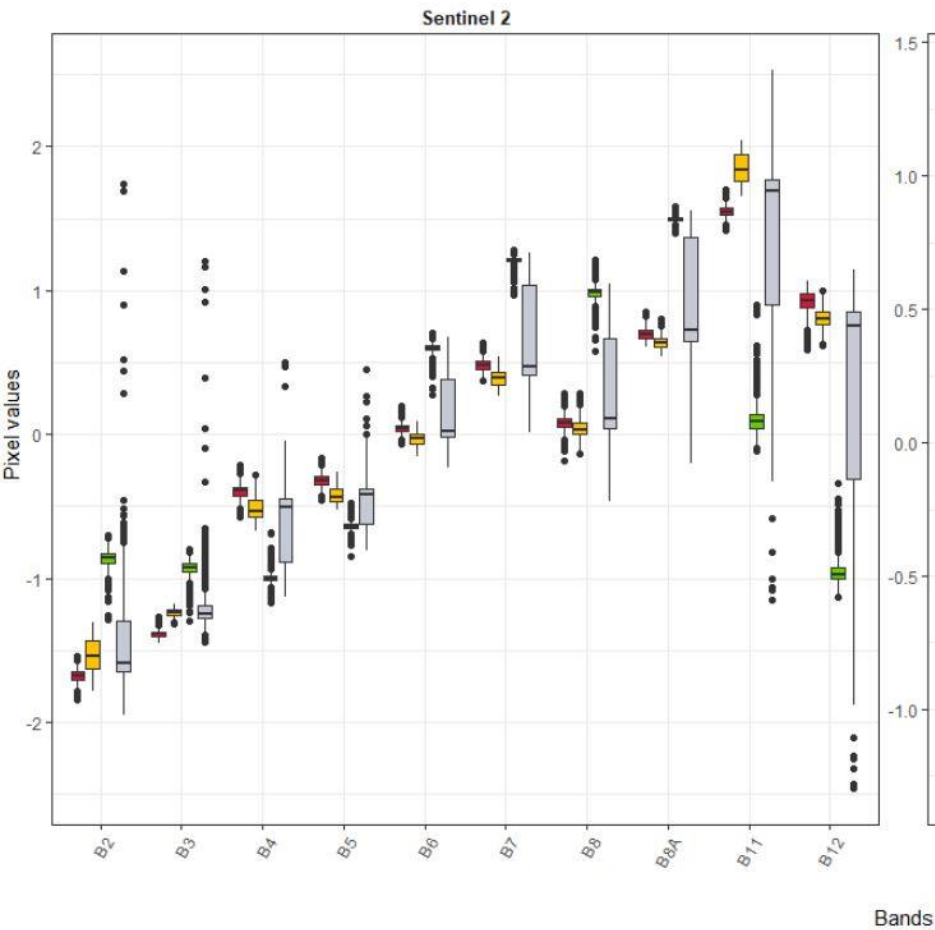




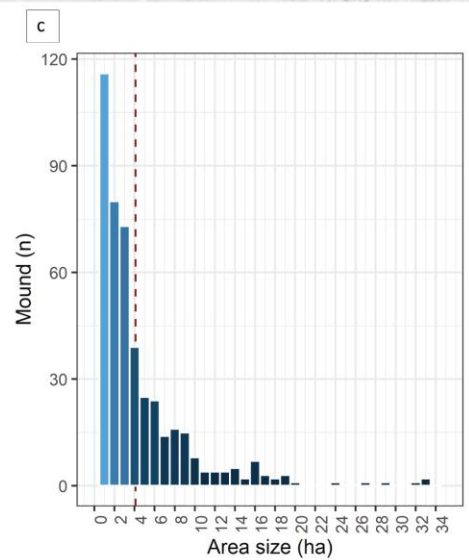
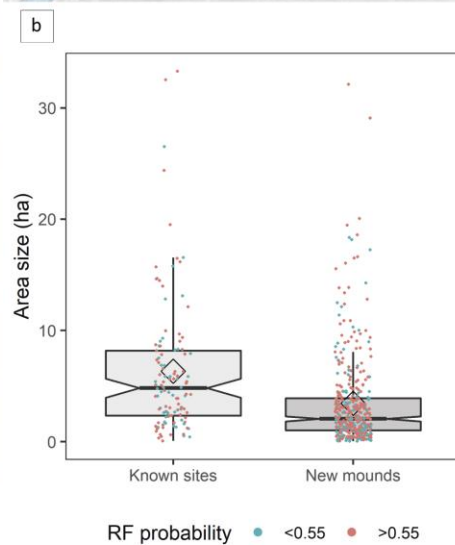
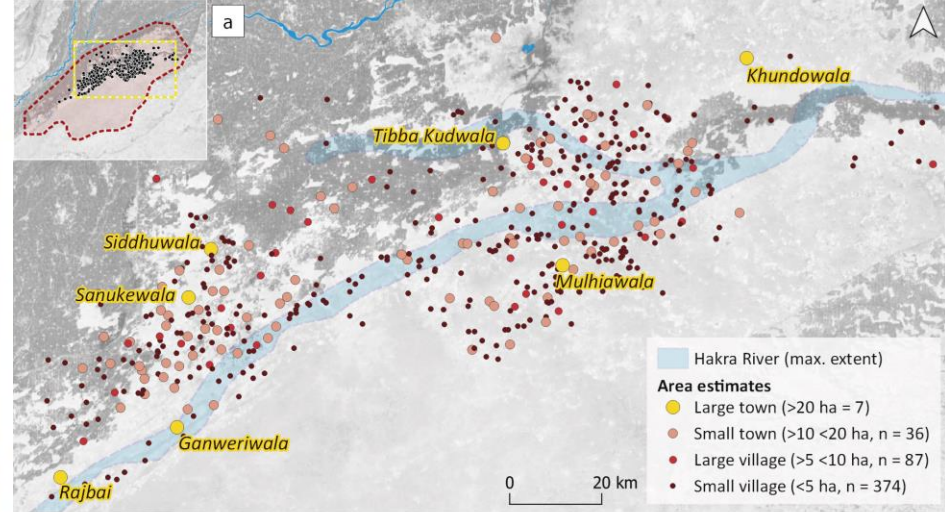
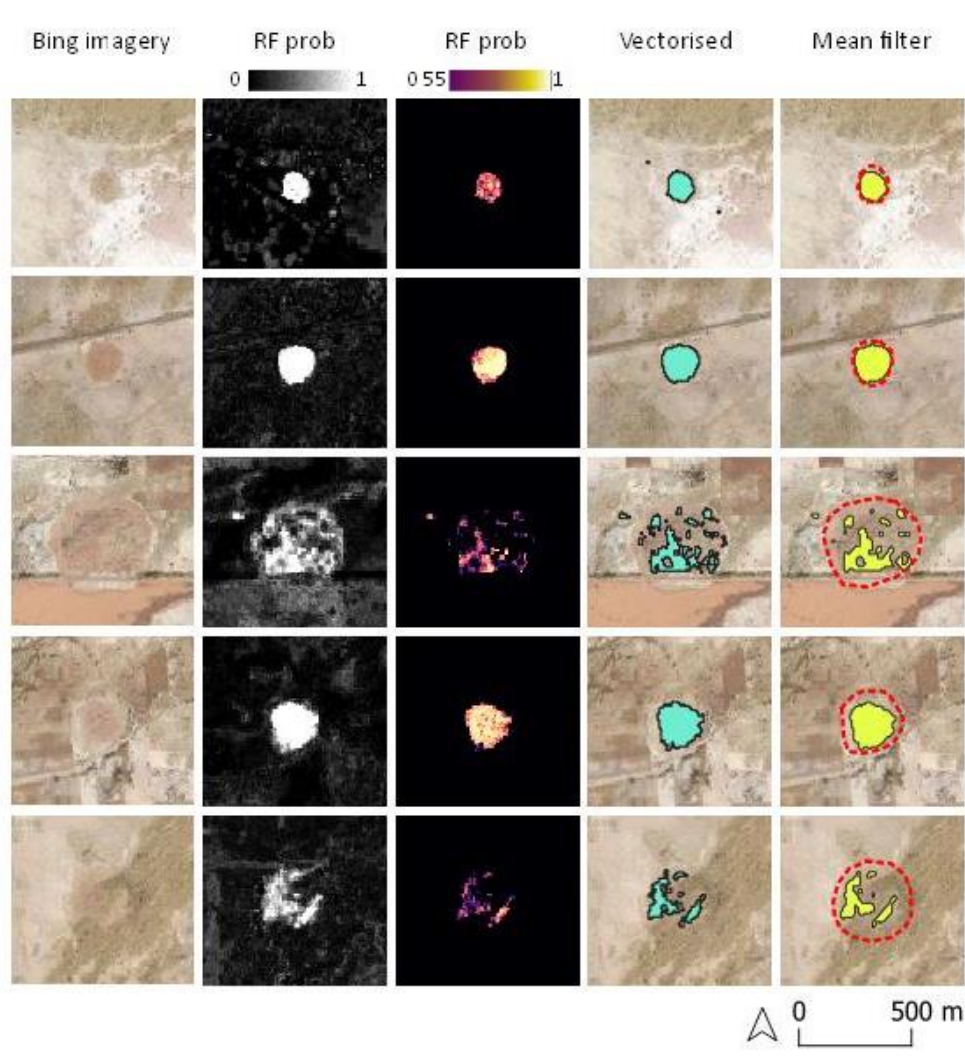
0 %

100





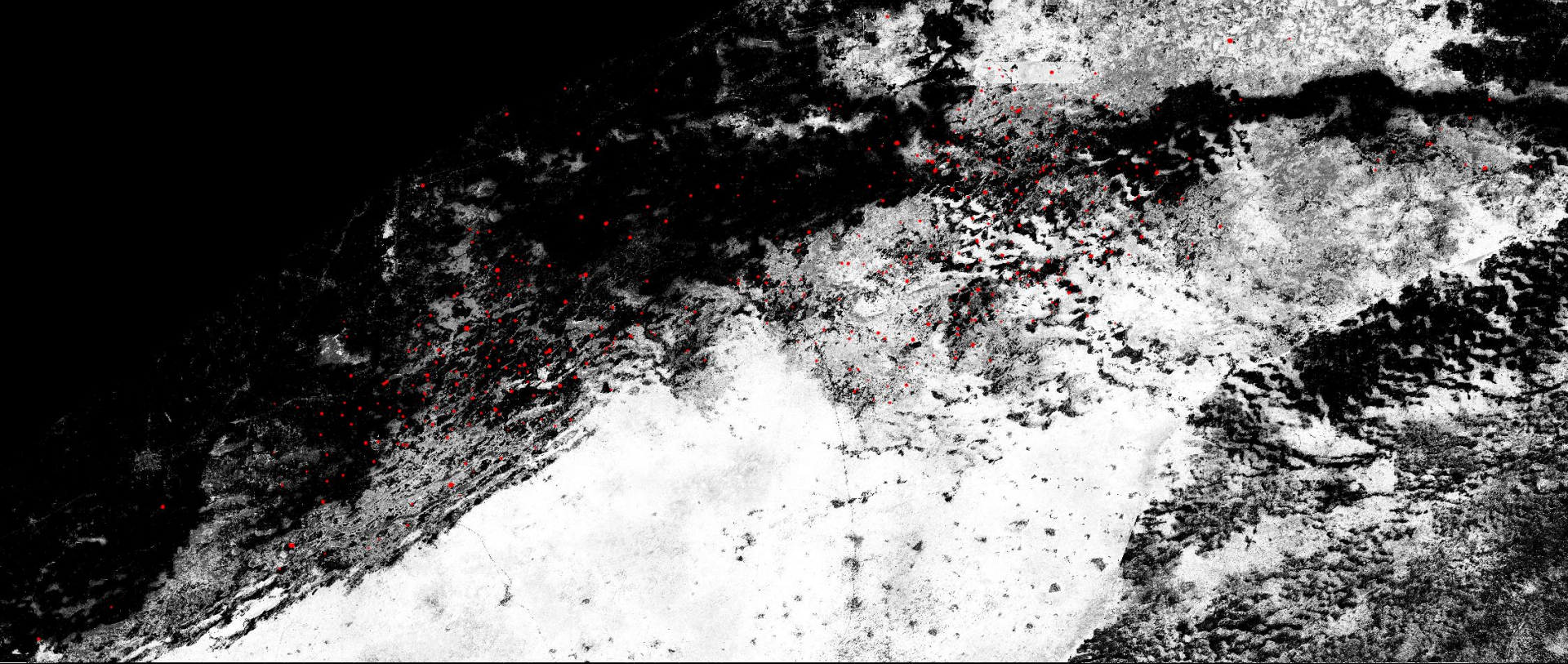






511 jaciments arqueològics detectats, només 135 es coneixien anteriorment  
4 noves ciutats de +30 ha (abans una)  
L'ocupació s'estén molt més enllà en el desert del que es pensava

Llocs atípics no detectats!! Només els que hem definit com a llocs: monticles  
Eficàcia de l'algoritme limitada a zones semidesèrtiques

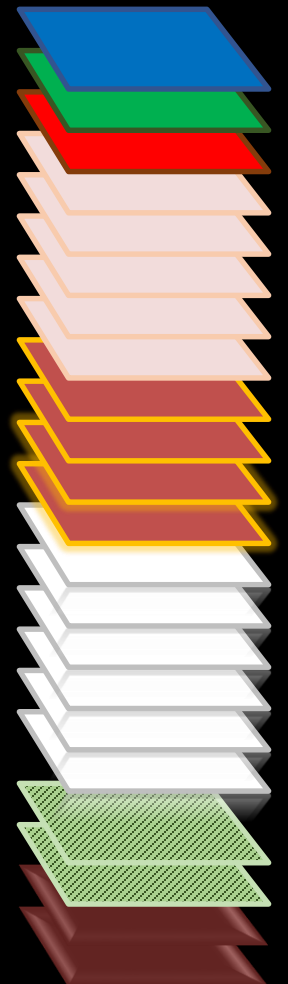


Distribution of sites + sketch of the signature →

Very good detection rate but not applicable to other environments of the Indus Civilisation

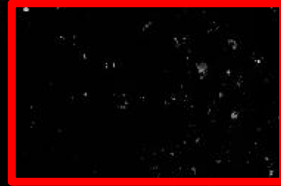
How to go about it?

Type	Source	Bands	Proxy	m/px			
Passive	Reflectance	B	Visible	Colour	10		
		G			10		
		R			10		
		Passive	Emissivity	RE1	Red Edge	Vegetation	20
				RE2			20
				RE3			20
				NIR			10
				RE4	Short Wave Infrared	Chemical composition	20
				SWIR1			20
				Active	Emissivity	SWIR2	Thermal Infrared
TIR1	30res						
TIR2	SAR					Physical/material Soil penetration	10
Cb A-VV							10
Cb A-VH		10					
Cb D-VV		10					
Cb D-VH		10					
Lb HH		25					
Lb HV	25						
Derived products		SMTVI	Multitemp			Seasonal variance	10
		MT-VIStD	Vegetation		10		
		MSRM	Topography	Prominence & size	12/30		
		MSRMStD			12/30		



# Cap al desenvolupament d'un algoritme de detecció de tells a escala continental

- Classificació no supervisada per dividir la variabilitat ambiental en zones ambientals coherents específiques
- Filtratge de taques per filtratge morfològic
- Segmentació
- Classificació zonal ML mitjançant dades d'entrenament específiques per a cada segment
- Unió de resultats i detecció d'objectes basada en aprenentatge profund



Estudi test d'aplicació global: de Síria al Golf Pèrsic

- Dades d'entrenament fàcils de seleccionar
- Resultats fàcils de provar

Moltes dades disponibles, però brutes, per exemple, Menze i els seus sòls antròpics del 2012  
Menze i Ur (2012) també ofereixen una comparació amb un enfocament de força bruta ML







Què passa amb els jaciments que no són visibles?

Aquells que no deixen rastre en les bandes que estem emprant

Com ho fem normalment?

## Prospecció pedestre arqueològica intensiva / extensiva

Problemes amb la creació de grans conjunts de dades a través de les prospeccions pedestres:

- Intercomparabilitat de les prospeccions: entre prospeccions, dins de cada prospecció
- Avaluació dels factors que influeixen en la visibilitat del terreny
- Cost molt gran (fons i temps) per obtenir dades quantitatives d'alta qualitat
- Les dades resultants es basen en el mostreig i la interpolació



Prospecció amb drone  
automatitzada

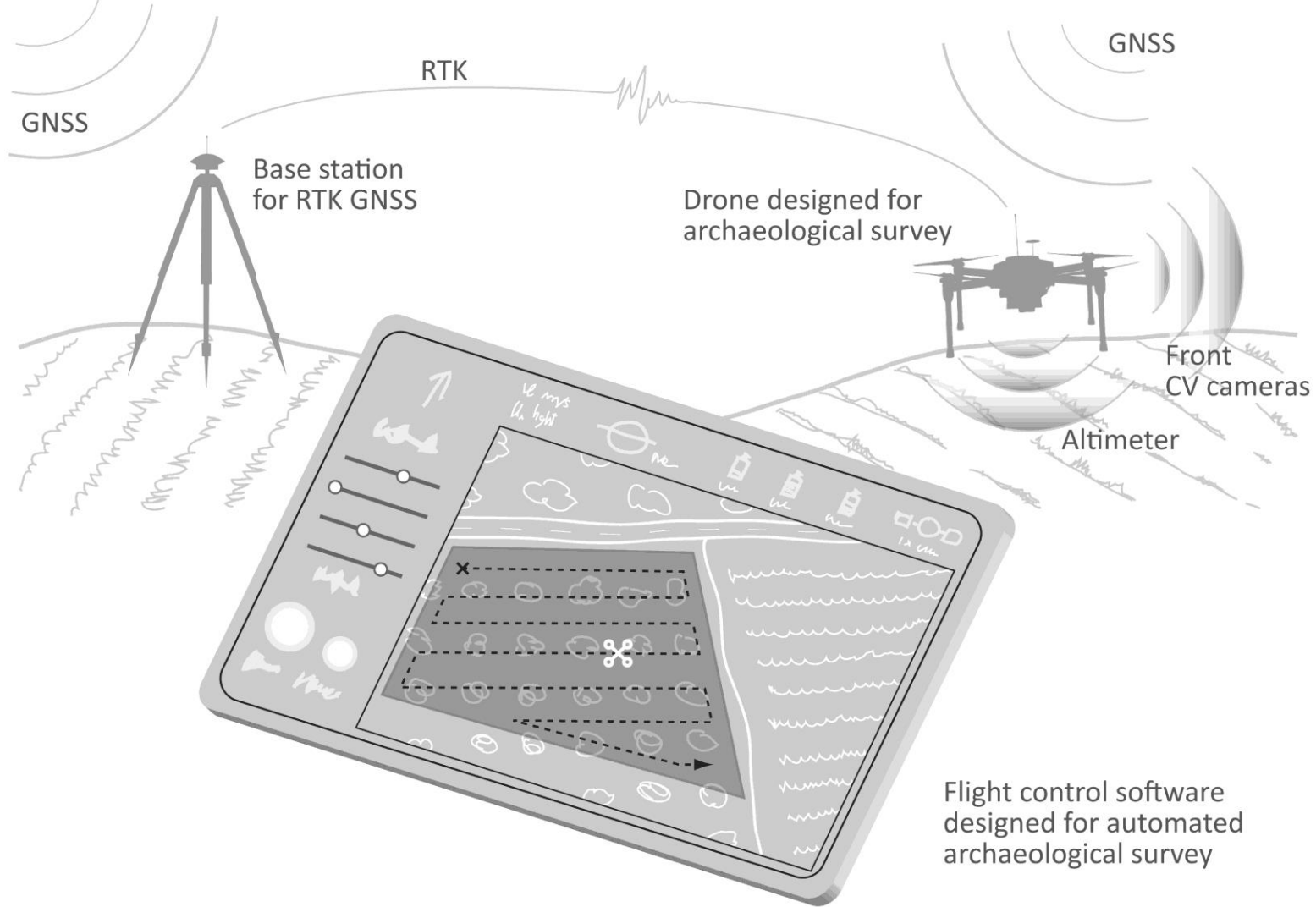
**VIII JORNADA  
DE PATRIMONI  
CULTURAL**

*LA INTEL·LIGÈNCIA  
ARTIFICIAL APLICADA  
AL PATRIMONI  
CULTURAL*



# Prospecció arqueològica automatitzada basada en drons

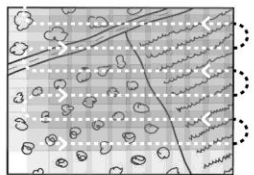




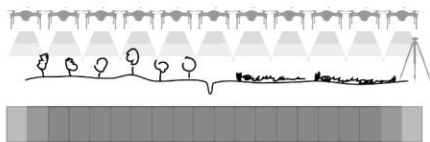


## Survey methods

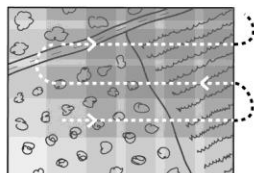
### Intensive total coverage survey with shape extraction of individual items



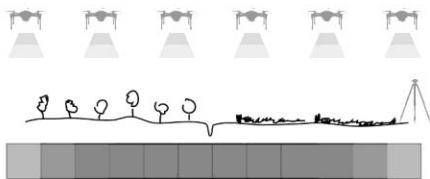
Flight path and coverage



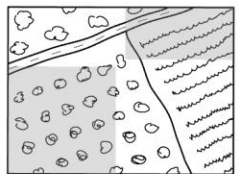
### Intensive total coverage survey with counting of individual items



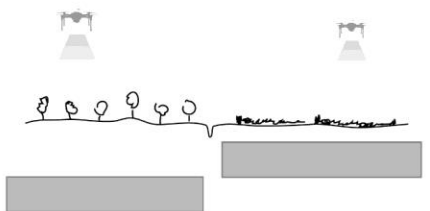
Flight path and coverage



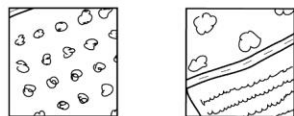
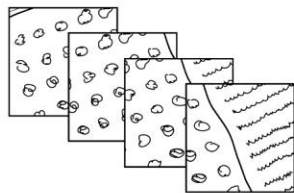
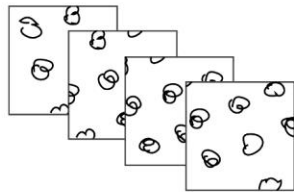
### Intensive sample-based survey



Coverage



## Photographs



## Computational processing stage

### Mosaicing



Monocular SLAM  
ORB-SLAM2

### Machine learning

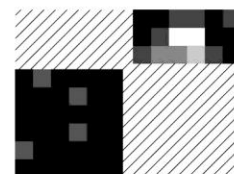
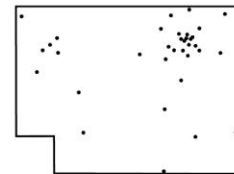
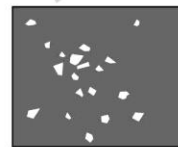
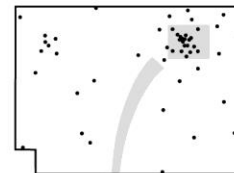


Deep learning-  
based segmentation

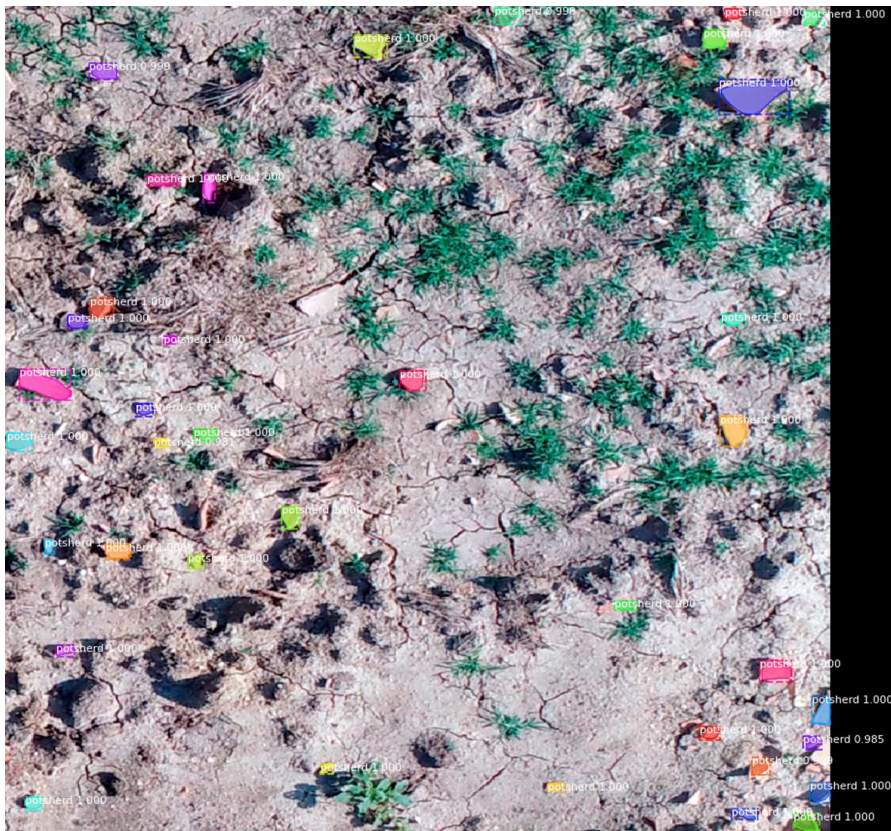


Random forest  
machine learning  
classifier (probabilistic)

## Outputs

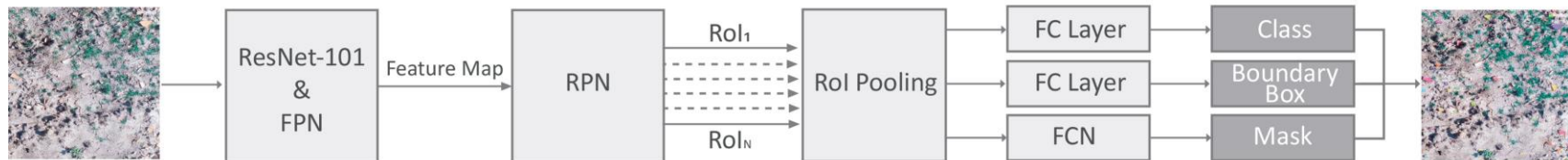






Orengo, H.A. & Garcia-Molsosa, A. 2019. A brave new world for archaeological survey: automated machine learning-based potsherd detection using high-resolution drone imagery. *Journal of Archaeological Science*, 112: 105013.

Orengo, H.A.; Garcia-Molsosa, A.; Berganzo-Besga, I.; Landauer, J.; Aliende, P. & Tres-Martínez, S. 2020. New developments in drone-based automated surface survey: towards a functional and effective survey system. *Archaeological Prospection*.







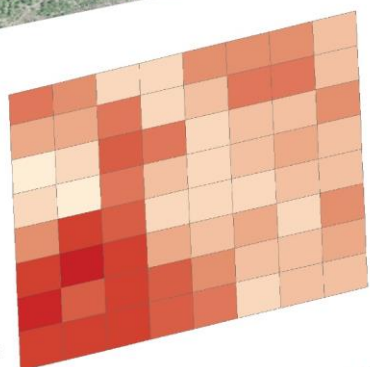
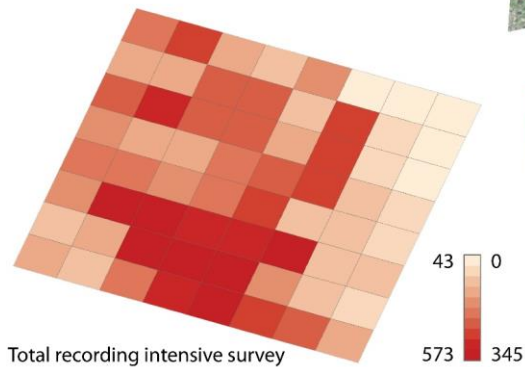
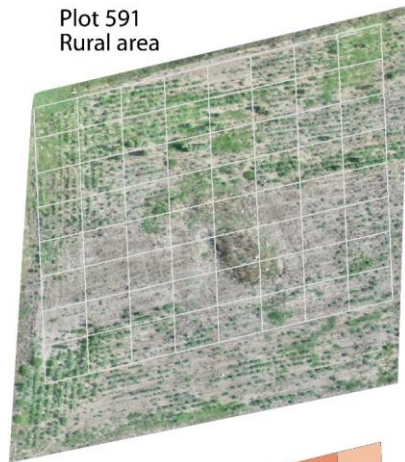
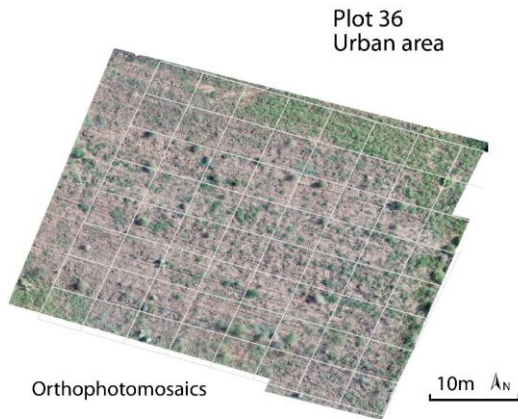


## Prospecció pedestre intensiva

Plots	Pedestrian total record		Pedestrian lines	
	Total time	Sherds	Total time	Sherds
36 - 5m lines	22 h	15779	3 h	1071
591 - 10m lines	18 h	2080	1:20 h	221

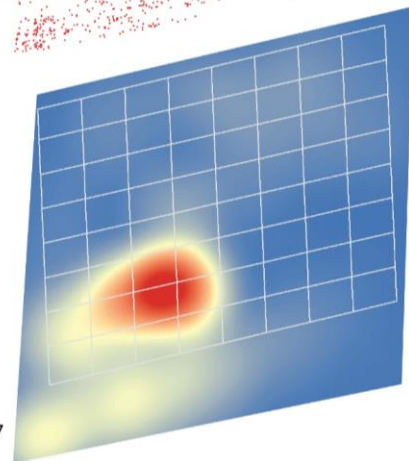
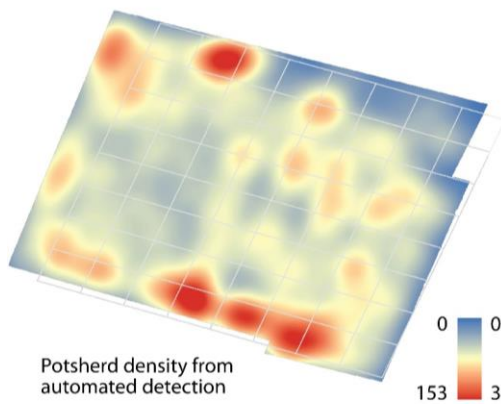
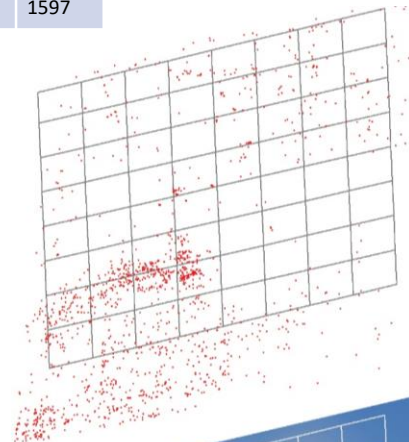
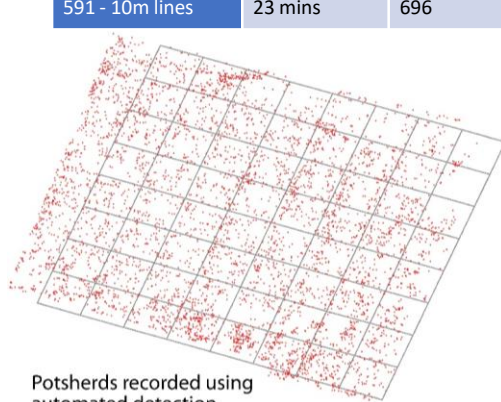
Plot 36  
Urban area

Plot 591  
Rural area

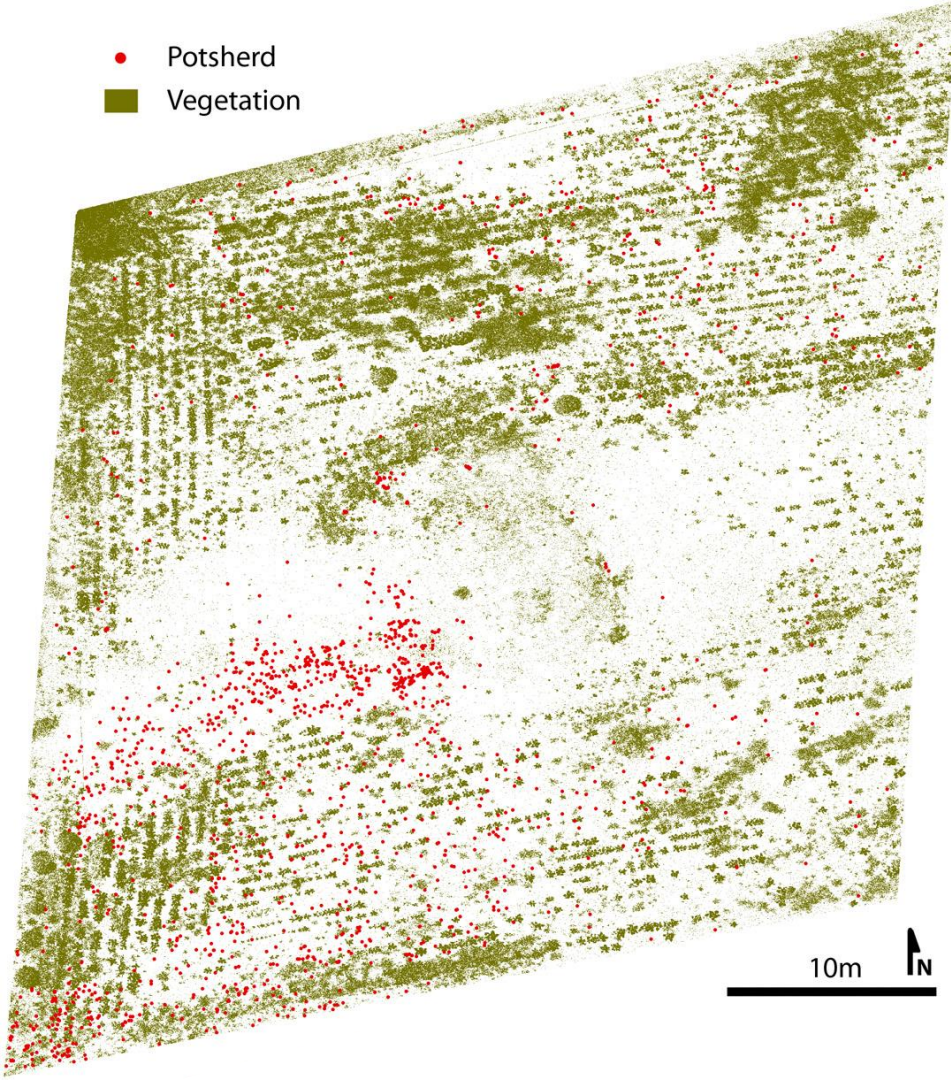


## Enregistrament automatitzat de fragments

Plots	Automatic recording		
	Time	Photographs	Sherds
36 - 5m lines	19 mins	564	5189
591 - 10m lines	23 mins	696	1597

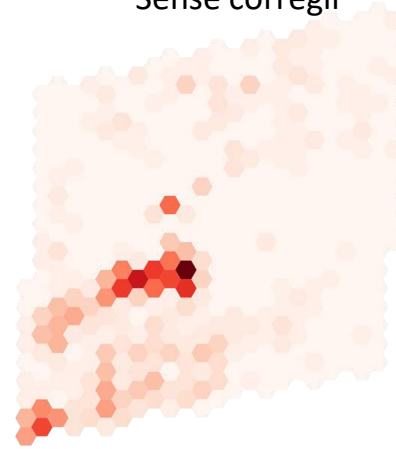


- Potsherd
- Vegetation

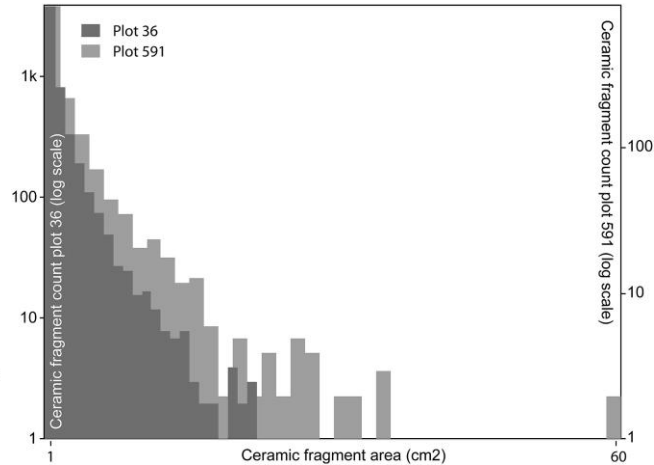
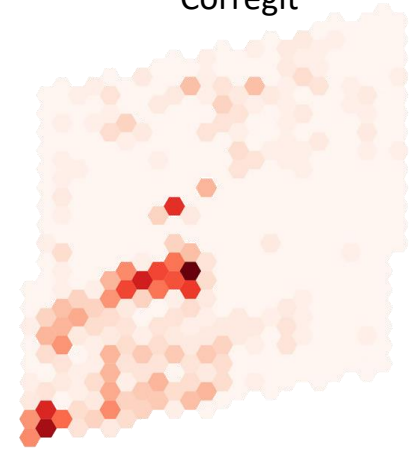


## Correcció basada en la densitat de la vegetació

Sense corregir



Corregit



Extracció de mida de fragments:

- Discriminació de jaciment
- Afectació
- Preservació del jaciment



{ Detecció de jaciments combinant  
mètodes d'aprenentatge automàtic }

**VIII JORNADA  
DE PATRIMONI  
CULTURAL**

*LA INTEL·LIGÈNCIA  
ARTIFICIAL APLICADA  
AL PATRIMONI  
CULTURAL*

Què passa amb aquests jaciments que no presenten distribucions ceràmiques?  
O s'amaguen sota la coberta forestal / vegetal?

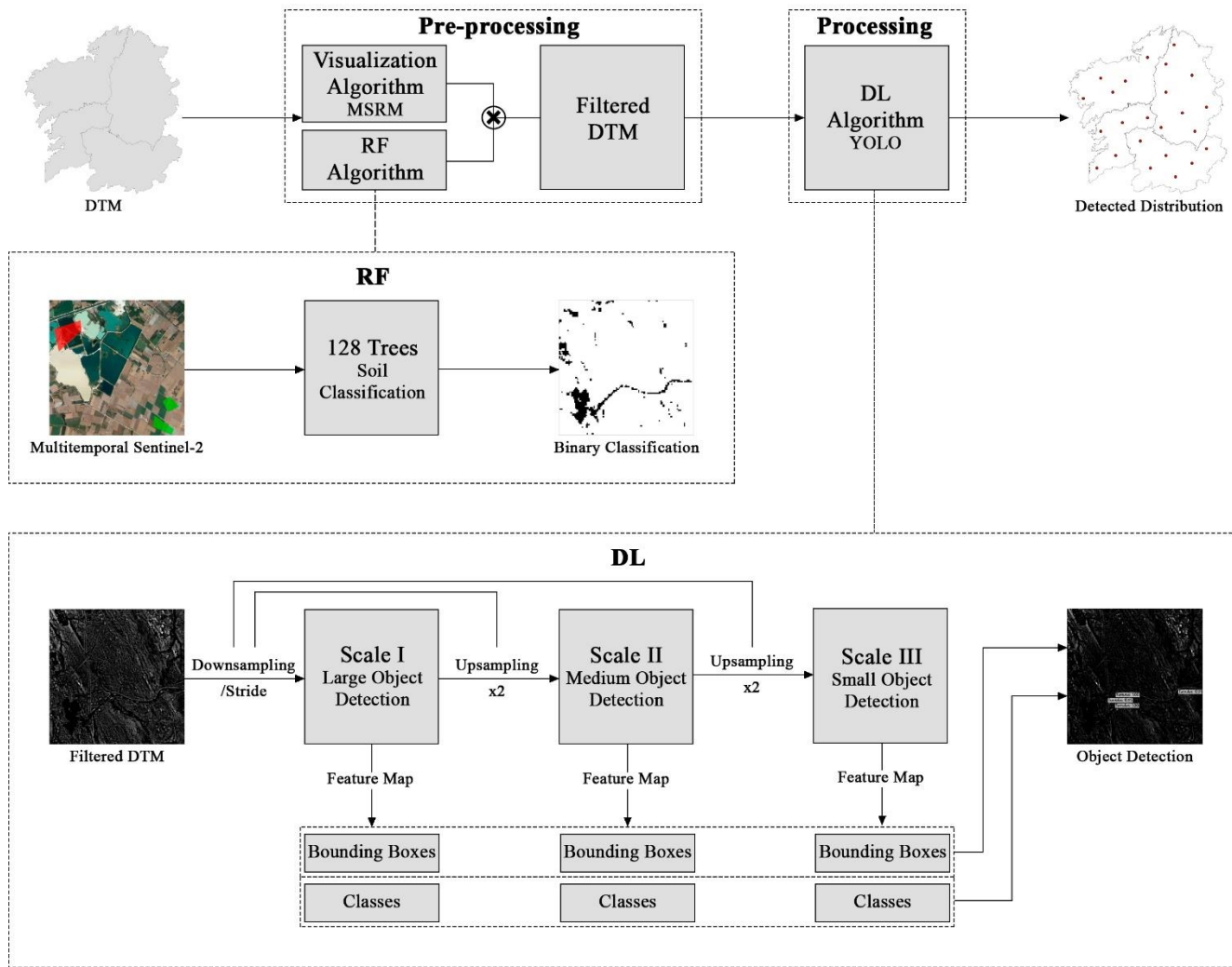
# Mètode de detecció híbrid ML/DL

## Estudi de cas sobre els túmuls gallecs:

Berganzo-Besga, I.; Orengo, H.A.; Lumbreras, F.; Carrero-Pazos, M.; Fonte, J. & Vilas-Estévez, B. 2021. Hybrid MSRM-Based Deep Learning and Multitemporal Sentinel 2-Based Machine Learning Algorithm Detects Near 10k Archaeological Tumuli in North-Western Iberia. *Remote Sensing*, 13(20): 4181.

- 30.000 km<sup>2</sup>
- Àrea complexa que inclou una gran varietat d'escenaris
- Els túmuls presenten una forma senzilla i constant
- Poques dades d'entrenament necessàries
- Gran nombre de falsos positius

Com augmentar el nombre de túmuls detectats evitant la inclusió de la majoria de falsos positius?



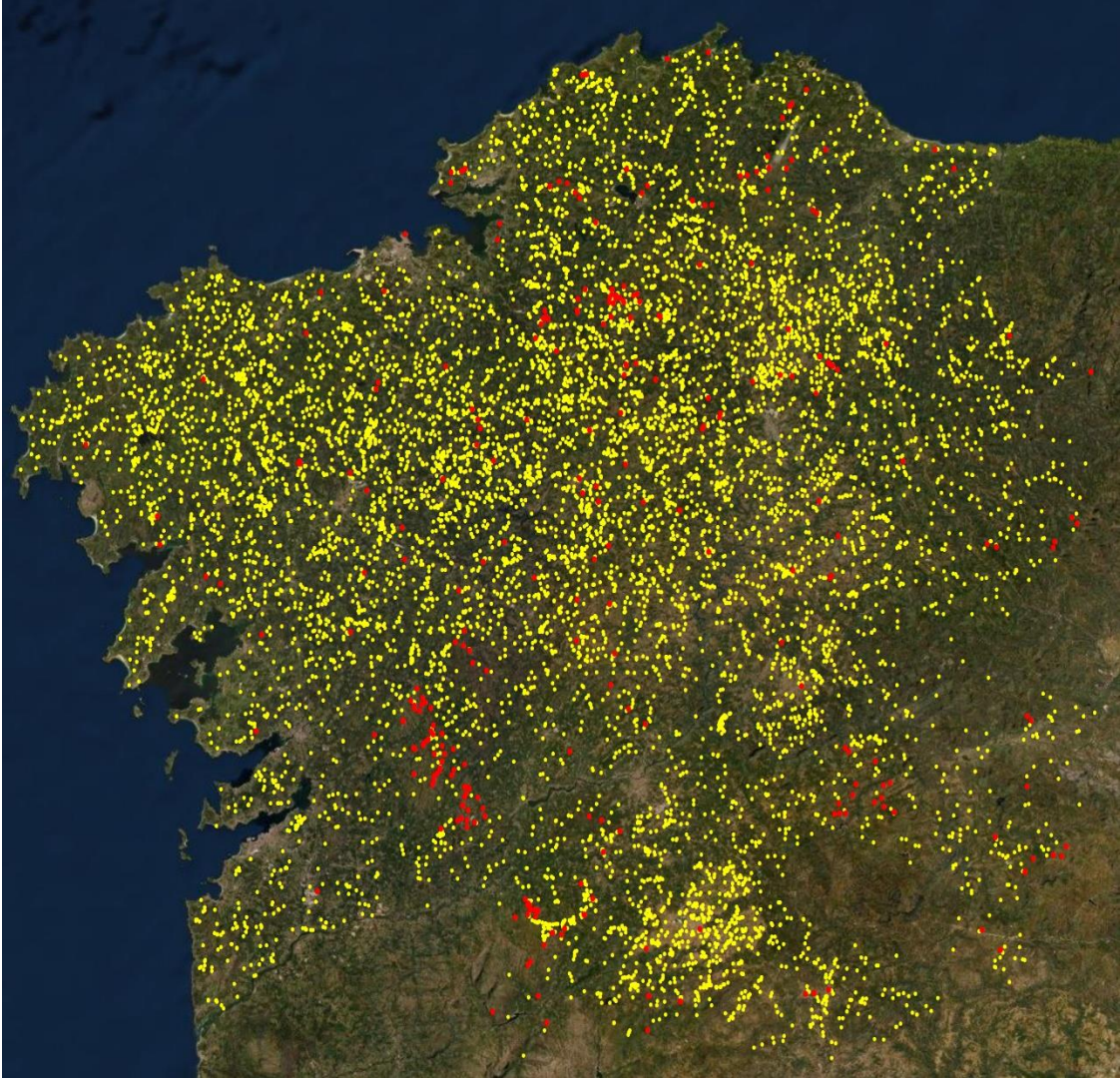


## Resultats sense precedents:

- Gairebé 30.000 km<sup>2</sup> analitzats, l'àrea més gran en què s'ha aplicat DL
- 10.527 objectes detectats, aproximadament 9.422 (89,5%) túmuls arqueològics
- Les imatges de codi obert de Sentinel 2 (10 m/px) emprades però amb imatges comercials la taxa d'èxit haurien augmentat per sobre del 97%.
- El codi, les fonts i els resultats estan disponibles gratuïtament i el codi està dissenyat per ser utilitzat en plataformes de computació en núvol d'accés lliure

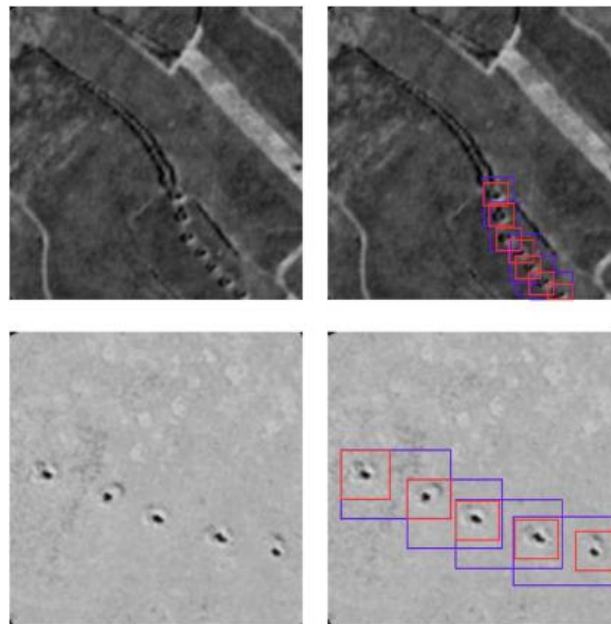
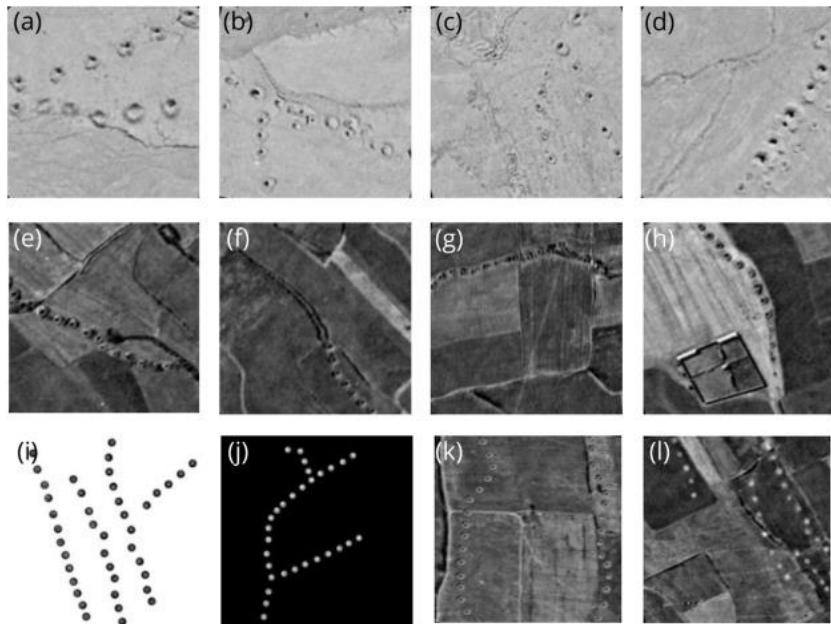
## Punts principals

- Ús de MSRM
- Tècniques d'augment de dades
- Combinació de ML càssica i DL

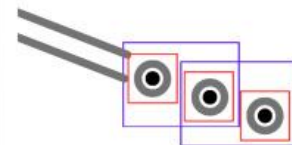






Detecció d'estructures d'irrigació  
fent servir imatges satèl·lites

# Imatges de satèl·lit espia Corona i Hexàgon (1970-80)



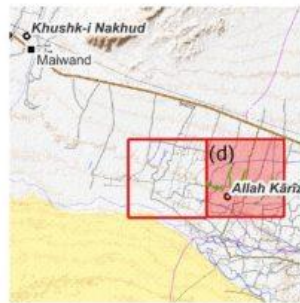
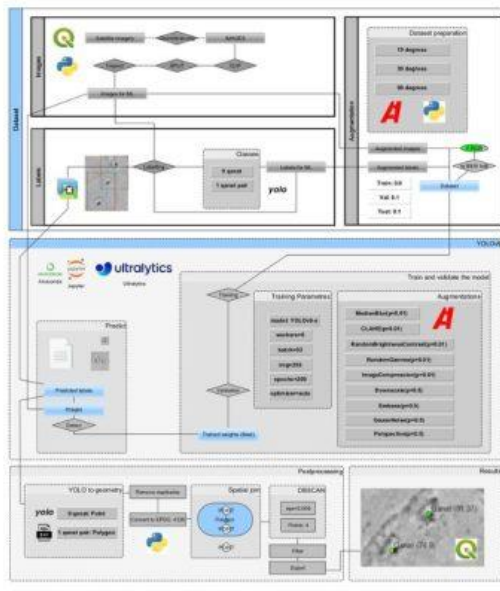
## Used labelling



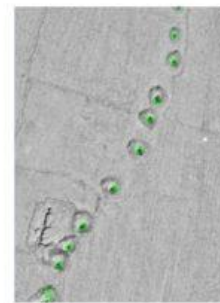
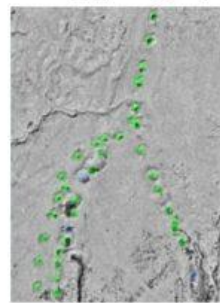
-  Qanat shaft (Class 0)
-  Pair of shafts (Class 1)
-  Canal (not detected in the study)
-  Other features



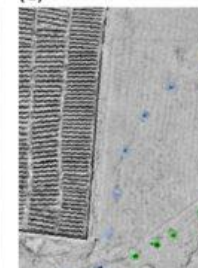
# A universal Object Detection model for mapping qanats



(d)



(e)



(f)

Test	Qanat shafts in pairs [class 0]					
	TP	FP	FN	Precision	Recall	F1
G2_D3C1216	1456	476	320	0.754	0.82	0.785
G2_D3C1219	693	314	562	0.688	0.552	0.613
G3_DS1110*	479	9	370	0.982	0.564	0.717
AFG2_D3C1209	351	56	262	0.862	0.573	0.688
AFG2_D3C1219	607	4	214	0.993	0.739	0.848
MR1_D3C1211	4581	339	3592	0.931	0.561	0.7
MR2_D3C1218	4600	214	3334	0.956	0.58	0.722
Average				0.881	0.627	0.725

Bulawka, Nazarij, Hector A. Orengo, and Iban Berganzo-Besga. 2024. 'Deep Learning-Based Detection of Qanat Underground Water Distribution Systems Using HEXAGON Spy Satellite Imagery'. Journal of Archaeological Science 171:106053.

<https://doi.org/10.1016/j.jas.2024.106053>

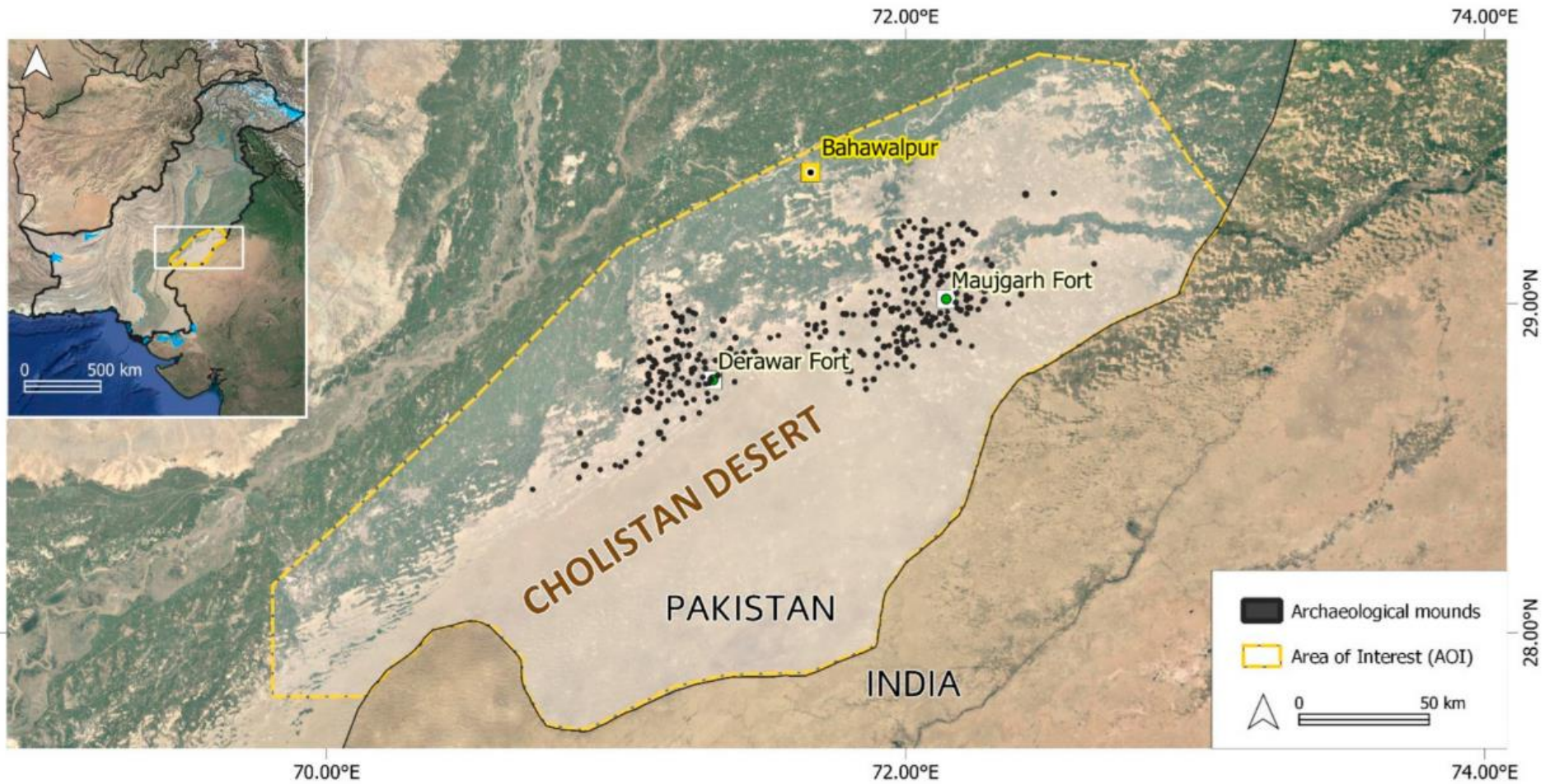


Monitorització de jaciment  
detectats en temps real

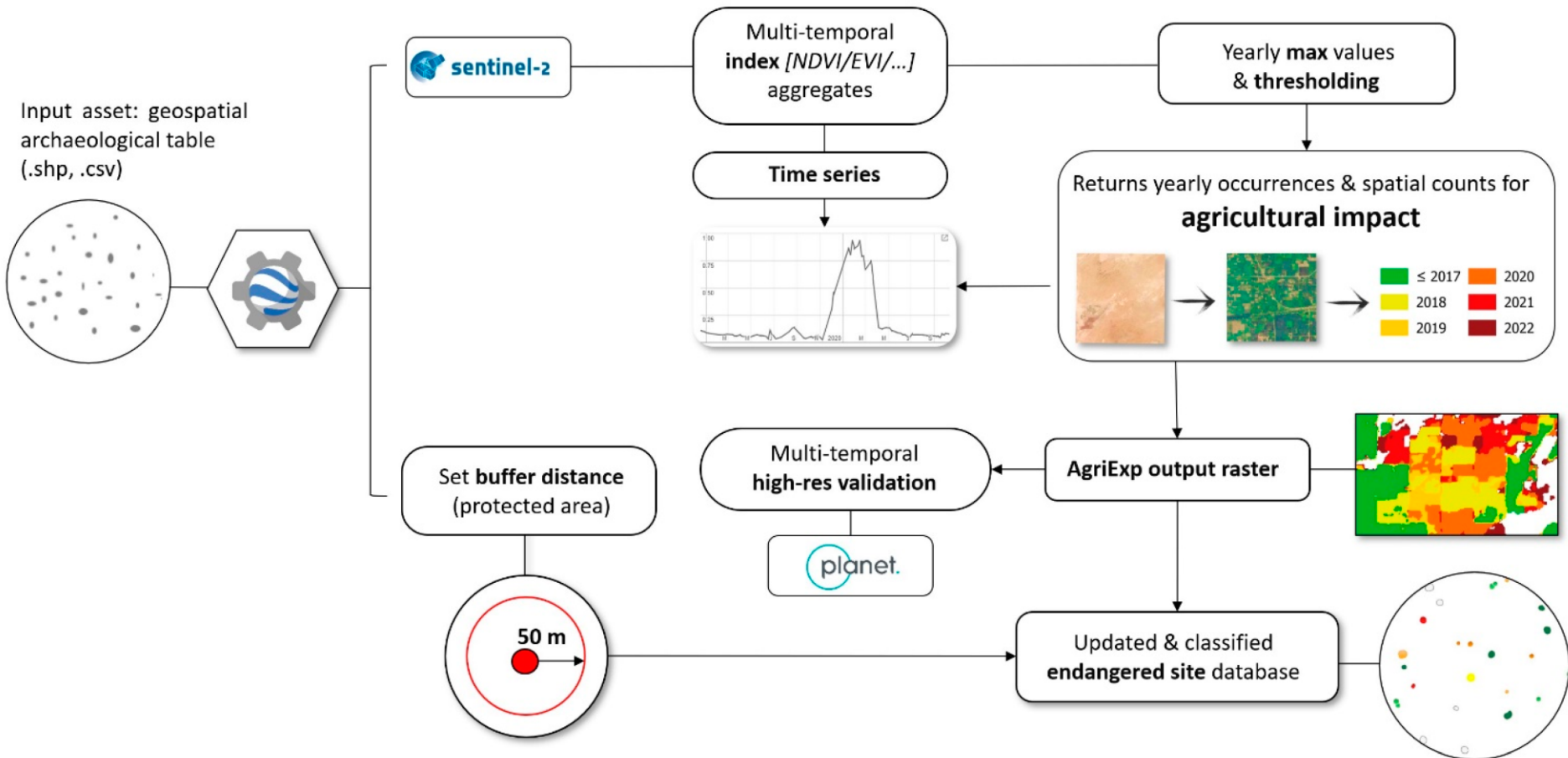
**VIII JORNADA  
DE PATRIMONI  
CULTURAL**

*LA INTEL·LIGÈNCIA  
ARTIFICIAL APLICADA  
AL PATRIMONI  
CULTURAL*

# Detectant l'impacte de l'expansió agrícola en el patrimoni

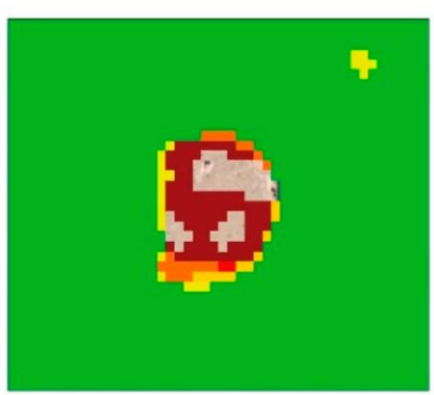
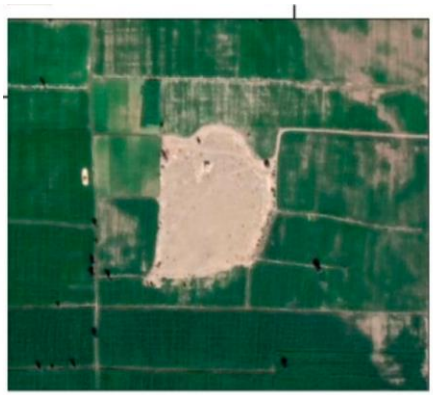
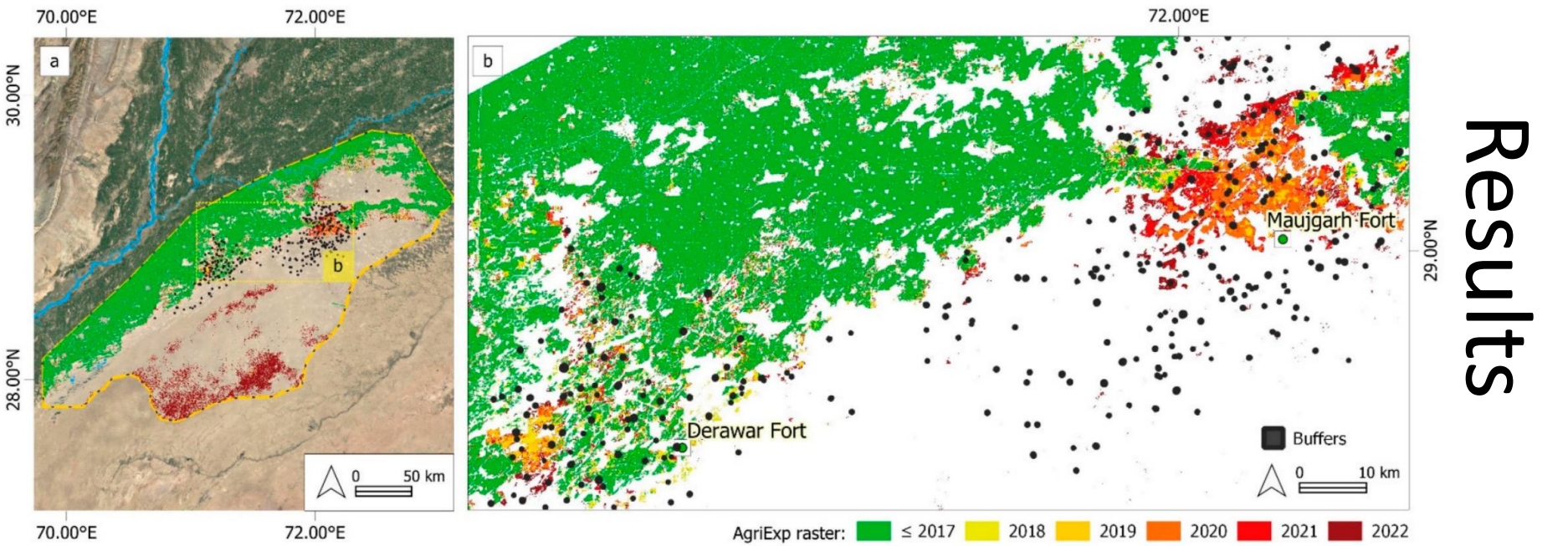


# Flux de treball



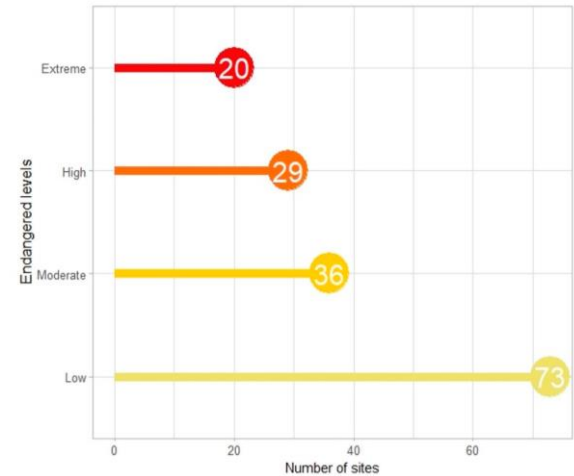
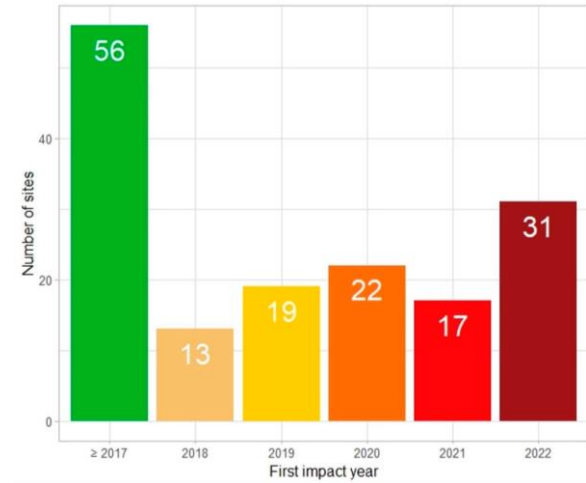
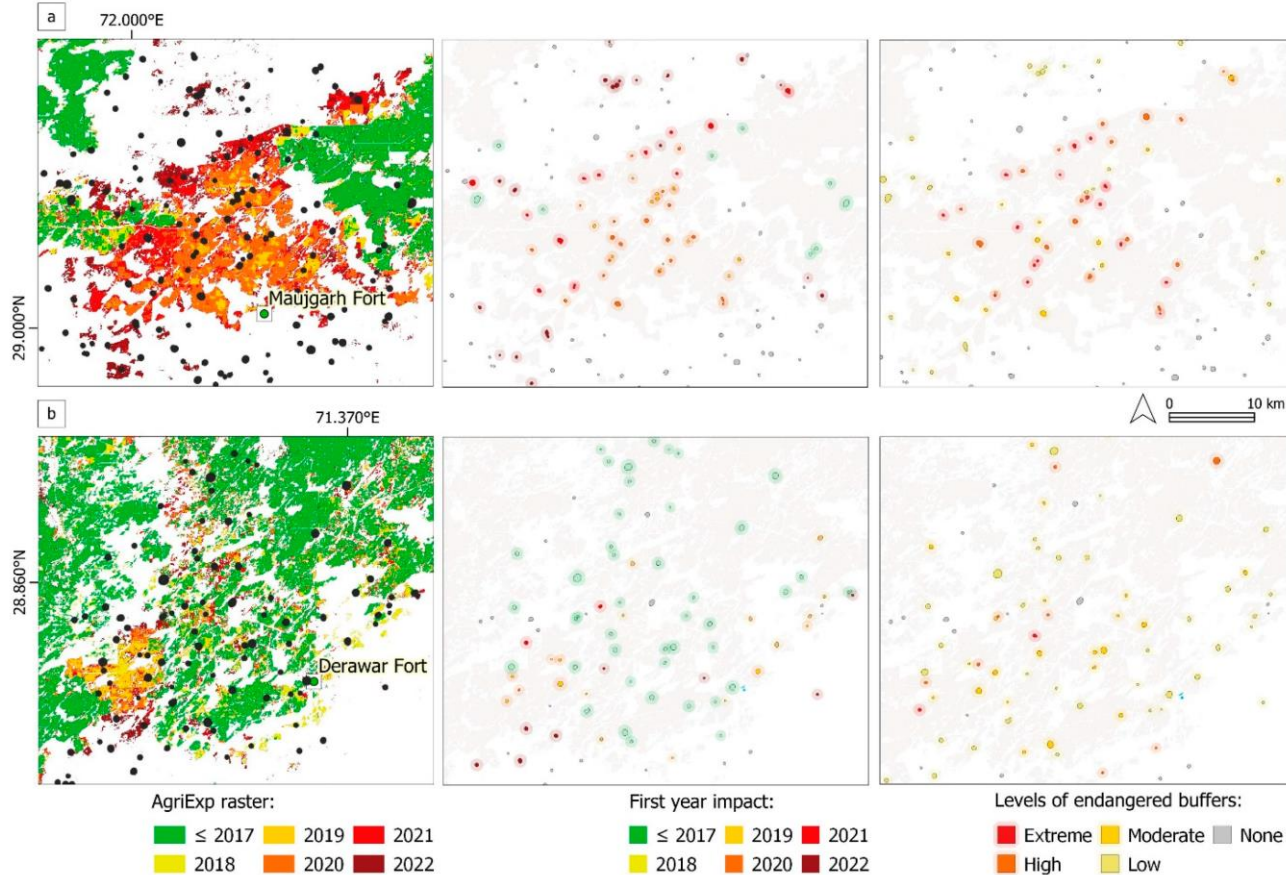


# Results





# Resultats



# Detecting the impact of urban expansion on the heritage

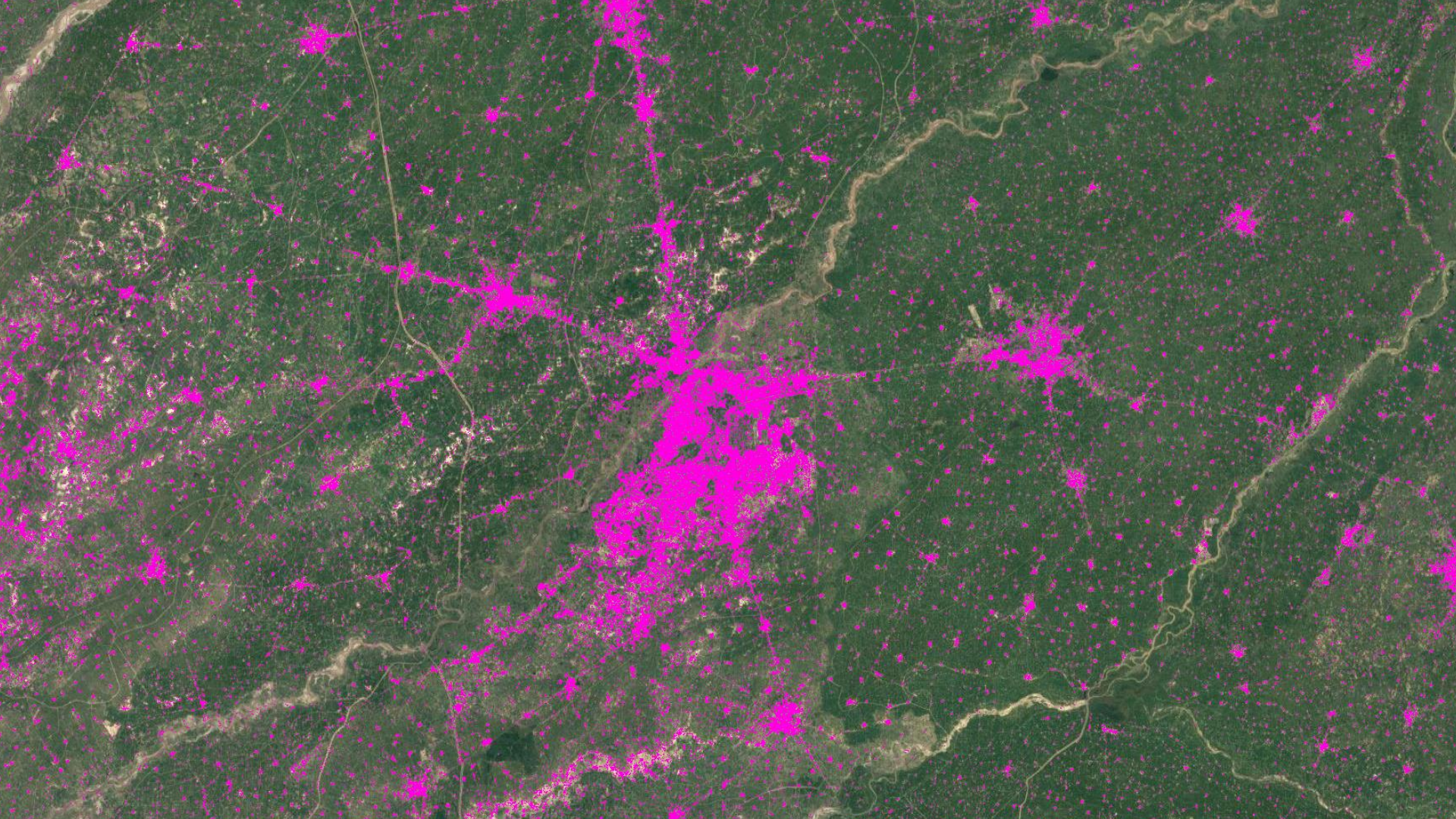
- Tipus d'anàlisi complexa
- Els tipus habituals d'aprenentatge automàtic / IA són difícils d'aplicar a causa de:
  - La diversitat específica dels assentaments urbans
  - La necessitat de detecció a alta resolució espacial
  - La necessitat de detecció a alta resolució temporal
  - La necessitat de reduir les capes d'imatges per a una producció ràpida i eficient de resultats

Estratègia específica desenvolupada per a la detecció d'expansió urbana gairebé en temps real:

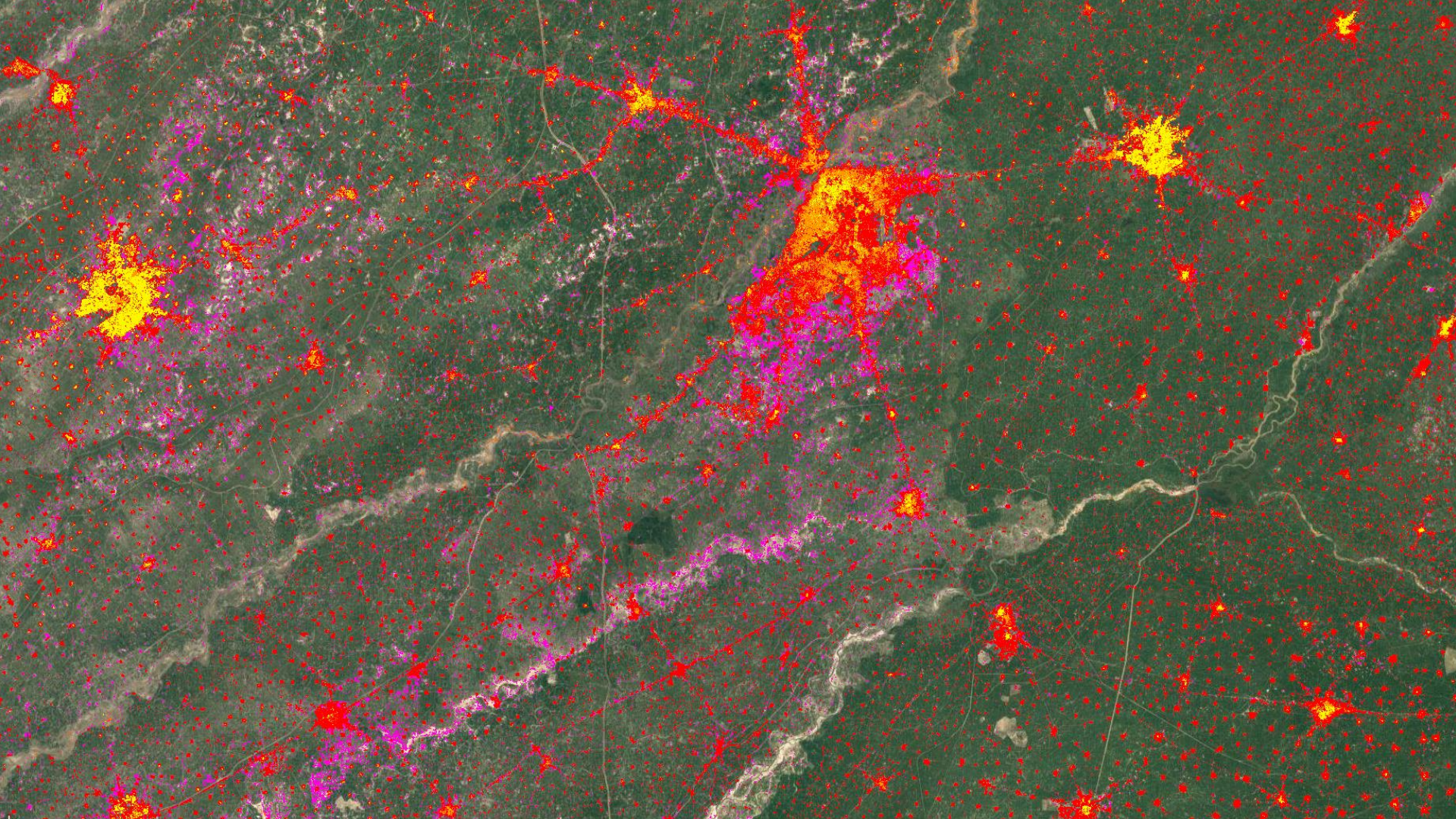
- Ús d'un enfocament multitemporal (mitjanes anuals per a l'expansió anterior, setmanals per a l'expansió actual)
- Ús d'un enfocament multisensor (SAR d'orientació múltiple i imatges multispectrals)
- Ús de fonts desenvolupades específicament (índexs de vegetació multitemporal i mapes de pendents convolucionals)
- Capacitat de detecció setmanal de llocs patrimonials afectats
- Detecció d'altíssima resolució (10m/px)



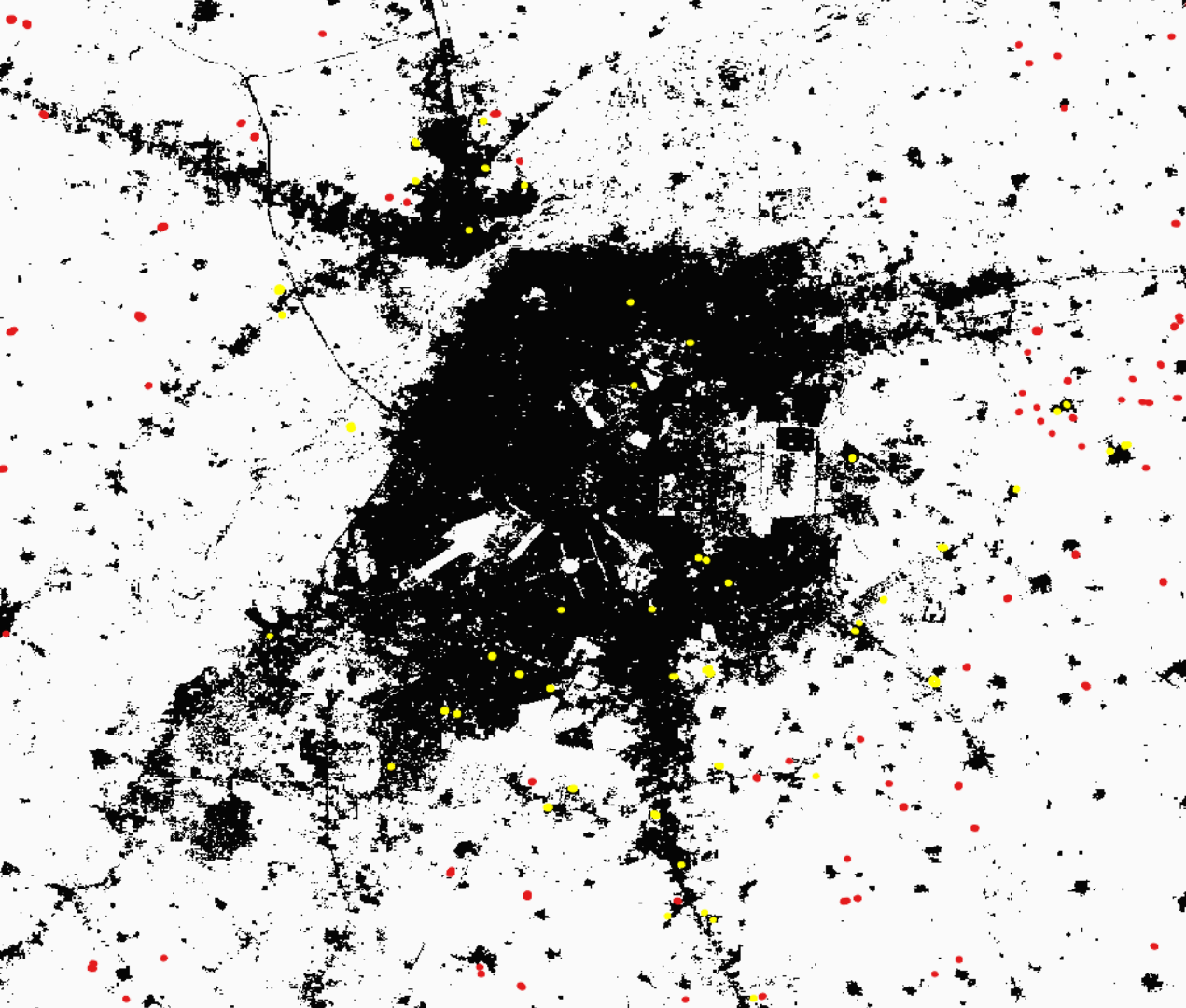












At least 63 potential sites have disappeared with the expansion of Lahore in the last 48 years (probably many more)

# Resultats

- De 6000 jaciments, 4956 es troben dins de l'àrea
- 608 s'han vist afectats per l'expansió urbana abans de 1975
- 348 s'han vist afectats entre 1975 i 2015
- 143 s'han vist afectats entre el 2015 i el 2023
- 
- Tenint en compte les capacitats de detecció de l'algoritme, els seus valors de recuperació i precisió i el percentatge de monticles de mapes que corresponen a llocs reals, el nombre de llocs afectats podria ser significativament més gran.
- L'expansió urbana afecta fortament els jaciments arqueològics
- Aquest impacte està augmentant exponencialment al llarg del temps

# Resultats

- El sistema proporciona una manera eficient de monitoritzar contínuament l'afectació dels jaciments arqueològics per l'expansió urbana
- Pot localitzar fins i tot edificis petits mentre es construeixen setmanalment
- Permet l'establiment de zones d'amortiment perquè es pugui donar una alerta abans de l'afectació
- És fàcil d'utilitzar, gratuït i una sola persona pot executar-lo amb molt poca formació i sense inversió de temps
- Es pot combinar amb el sistema de monitorització d'expansió agrícola
- Els jaciments es poden integrar immediatament a mesura que es cataloguen



Automatització d'analítiques

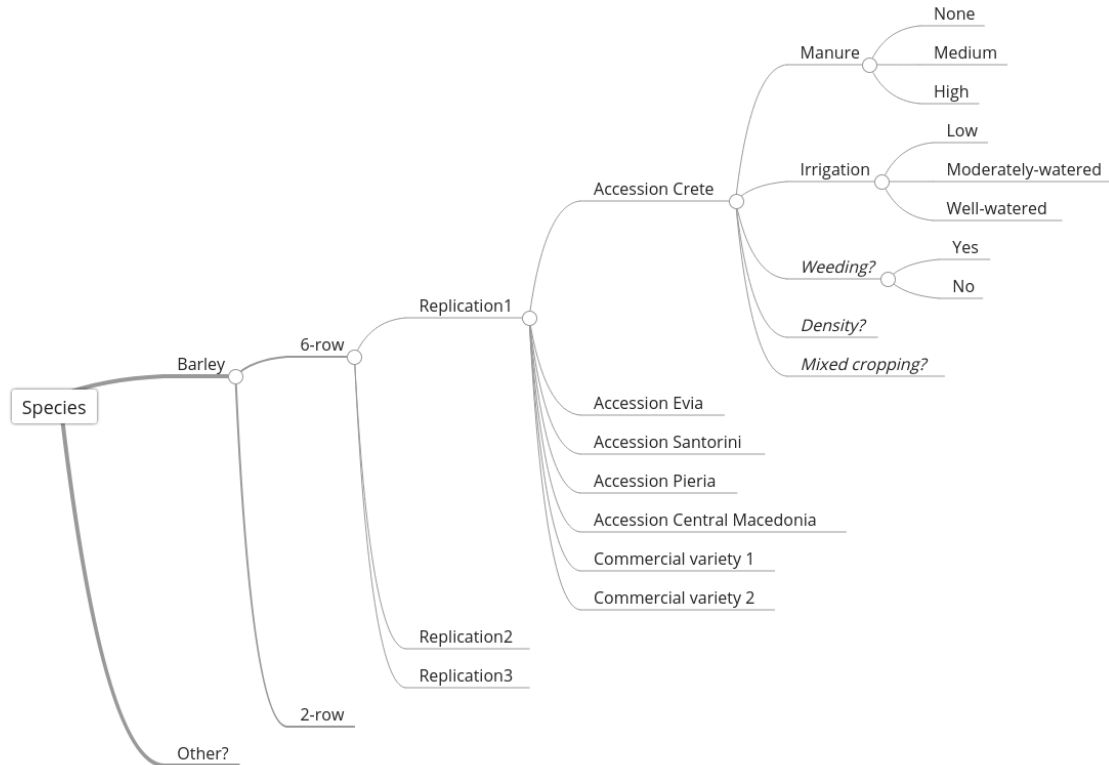
**VIII JORNADA  
DE PATRIMONI  
CULTURAL**

*LA INTEL·LIGÈNCIA  
ARTIFICIAL APLICADA  
AL PATRIMONI  
CULTURAL*

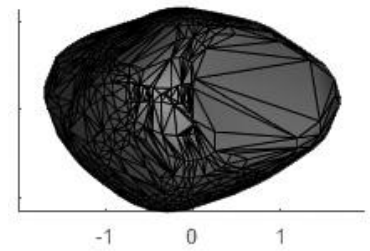
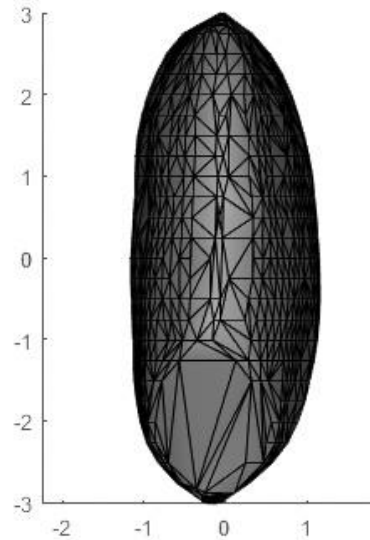
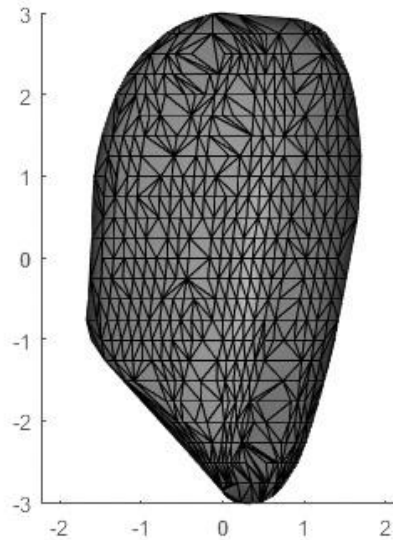
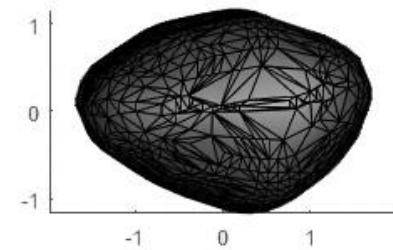
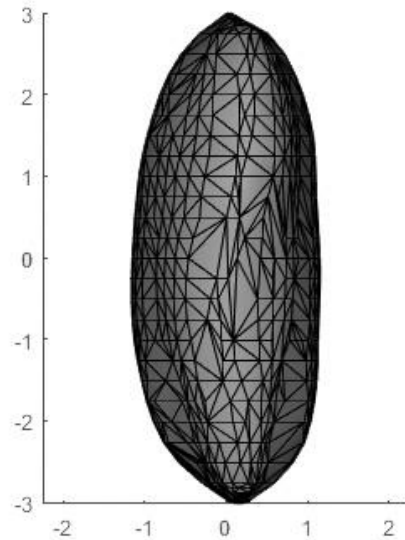
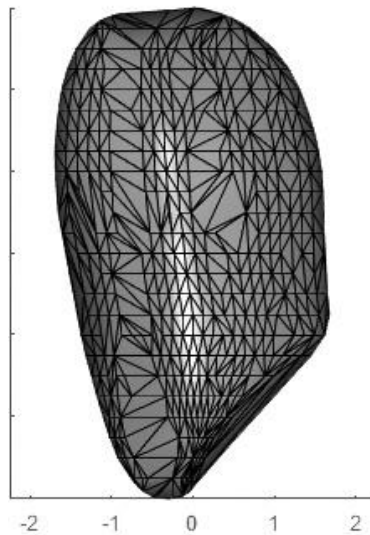
# Identificació d'espècies de llavors, varietats autòctones i règims agrícoles passats

L'objectiu final és desenvolupar una eina per a la identificació automatitzada de règims agrícoles passats mitjançant la forma de les llavors

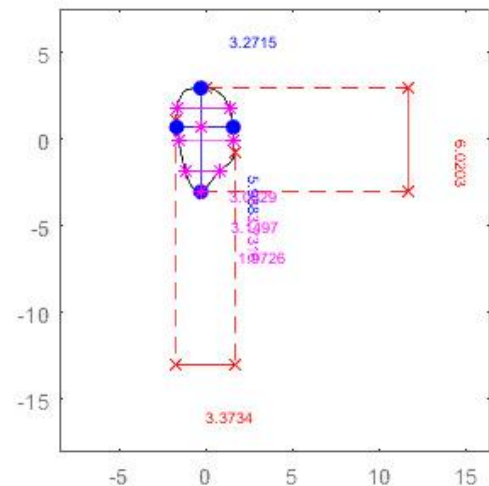
El primer pas va ser el desenvolupament de dades d'entrenament i per això es va desenvolupar un experiment de cultiu de dos anys



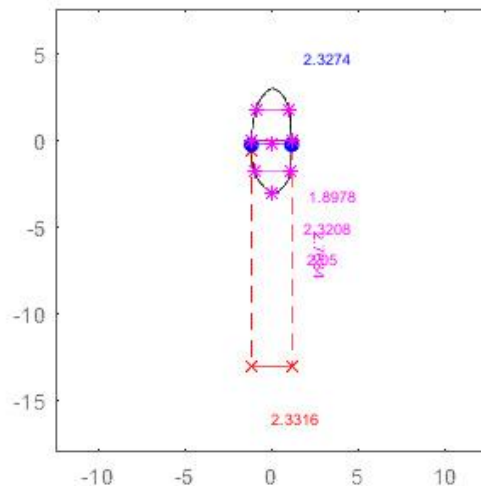
Barcode 3D surface distribution



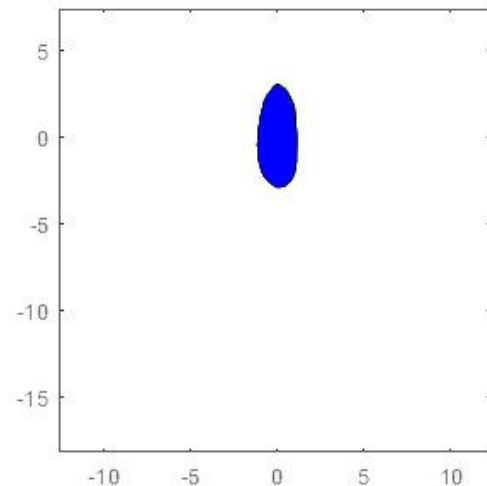
Front view



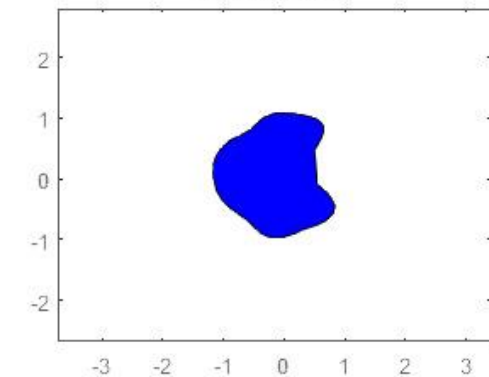
Side view



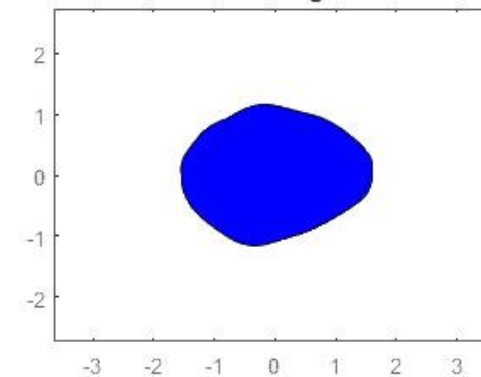
Longitudinal section at half width



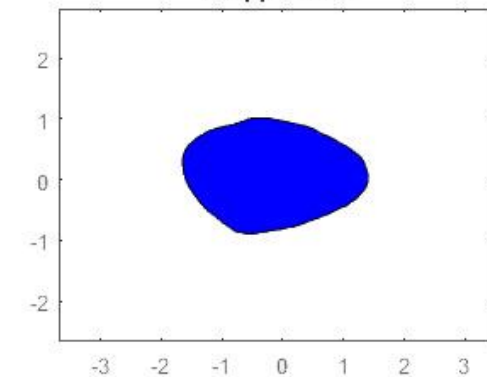
Transversal section at lower fifth



Transversal section at half length

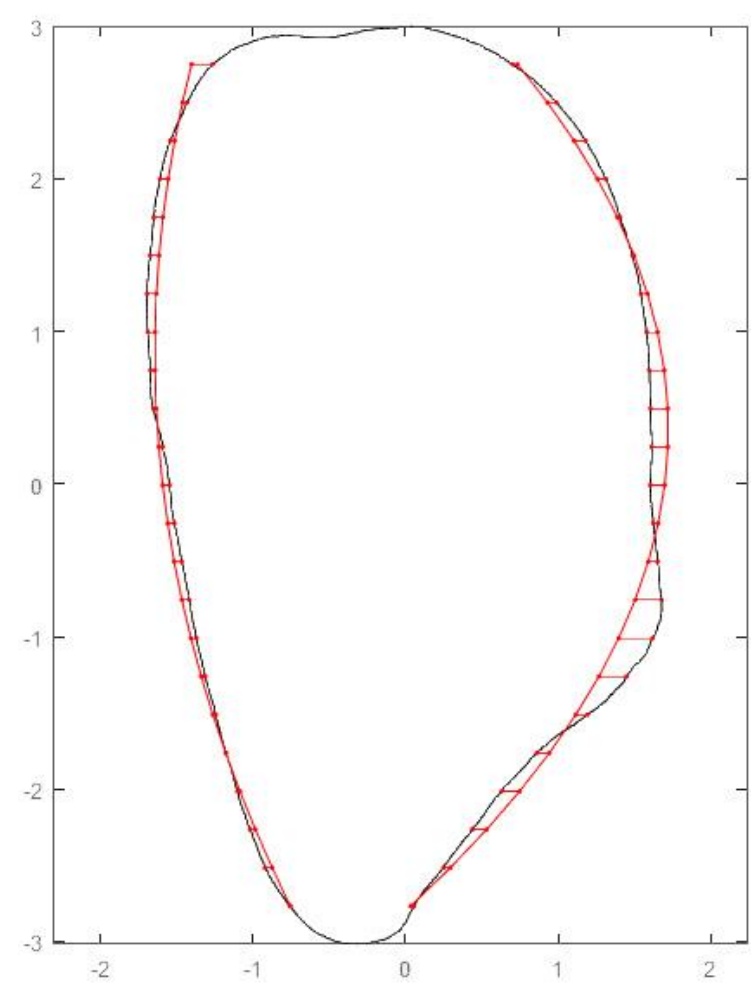


Transversal section at upper fifth

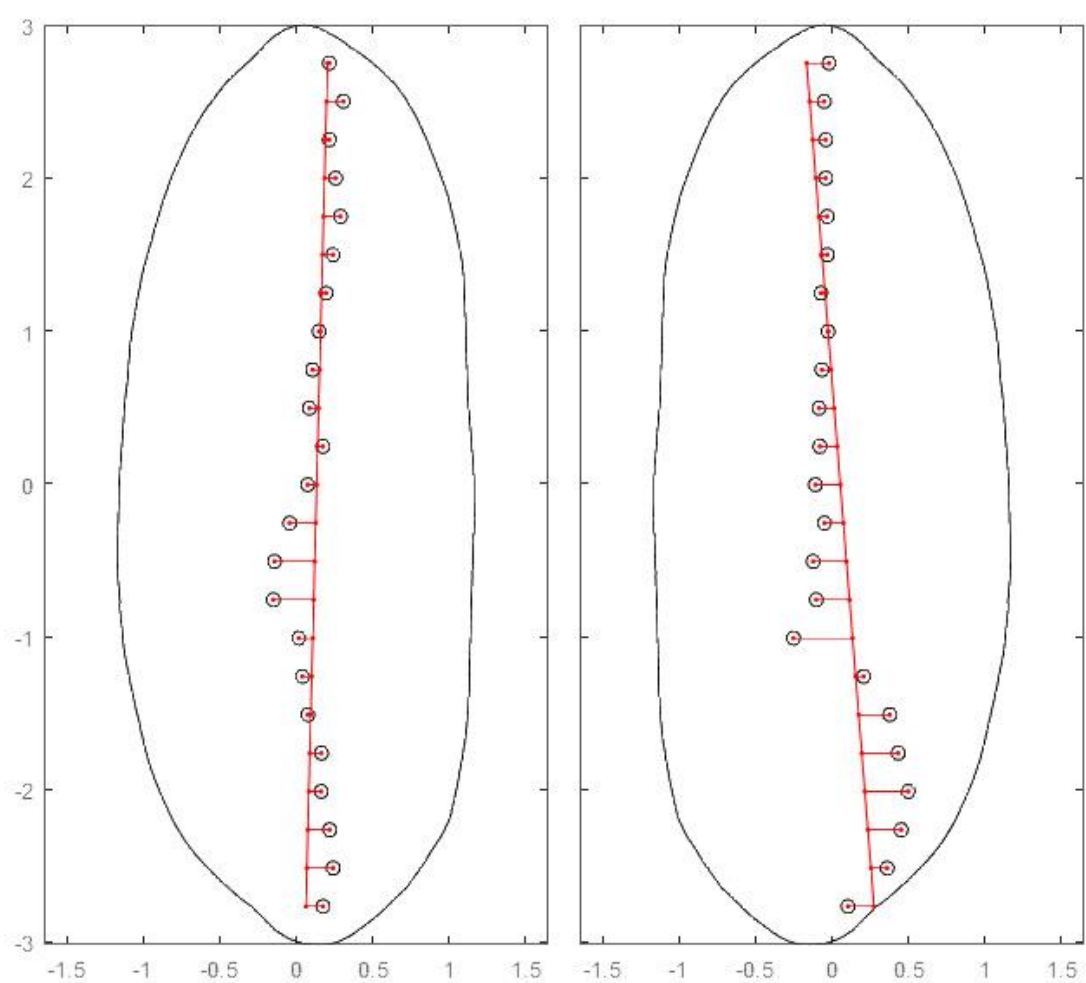


Caliper Grid at specified locations all values are in mm





Edges curvature



Edges regularity

# Propietats utilitzades per a la classificació ML

## Atributs inherents

Volum  
Superfície  
Diferències de variabilitat de forma  
Diferència de forma mitjana  
Diferències de volum  
Diferències d'àrea  
Desviació de les diferències de simetria bilateral perfectes  
Desviació de les diferències de simetria bifrontal perfectes  
Diferències de curvatura de la vora esquerra  
Diferències de curvatura de la vora dreta  
Diferències d'irregularitat de la forma del pla de la vora esquerra  
Diferències d'irregularitat de la forma del pla de la vora dreta  
Diferències d'irregularitat de la secció de la vora esquerra  
Diferències d'irregularitat de la forma del pla de la vora dreta

## Mètriques

Longitud màxima  
Amplada màxima  
Gruix màxim  
Longitud de la quadrícula  
Amplada de la quadrícula  
Gruix de la quadrícula  
Longitud a l'amplada de la quadrícula  
Longitud al gruix de la quadrícula  
Amplada a la cinquena part inferior  
Amplada a mitja longitud  
Amplada a la cinquena part superior  
Gruix a la cinquena part inferior  
Gruix a mitja longitud  
Gruix a la cinquena part superior

## Índexs i ràtios propis...

Activities Firefox Web Browser ML\_workflow\_3D\_seeds x Home Page - Select or cr x ML\_workflow\_3D\_seeds x 127.0.0.1:8888/notebooks/ML\_workflow\_3D\_seeds.ipynb jupyter ML\_workflow\_3D\_seeds Last Checkpoint: 6 minutes ago (autosaved) Logout Python 3 O

## Machine learning workflow for the analysis of 3D seeds

### Import CSV as an ML-ready array

```
In [7]: import pandas as pd
import numpy as np

trData = pd.read_csv('seedsData.txt')

x_train = np.asarray(trData.iloc[:, :-1])

y_train = np.asarray(trData.iloc[:, -1]) # Last column storing classes
```

```
In [8]: x_train
Out[8]: array([[ 5.17,  3.45,  2.81,  5.12,  3.43,  2.79,  3.21,  2.85,  1.62,
                3.31,  3.16,  1.4 ,  2.77,  2.32],
               [10.57,  6.65,  5.11, 10.56,  6.63,  5.08,  6.13,  5.7 ,  4.8 ,
                6.45,  6.24,  3.89,  5.05,  4.58],
               [12. ,  8.79,  5.25, 11.99,  8.77,  5.15,  7.92,  5.16,  6.92,
                8.53,  8.17,  4.75,  5.13,  4.66],
               [ 6.02,  3.37,  2.33,  5.99,  3.27,  2.33,  3.73,  2.77,  1.97,
                3.15,  3.03,  2.05,  2.32,  1.9 ]])
```

```
In [9]: y_train
Out[9]: array(['Landrace1', 'Landrace2', 'Landrace3', 'Landrace4'], dtype=object)
```

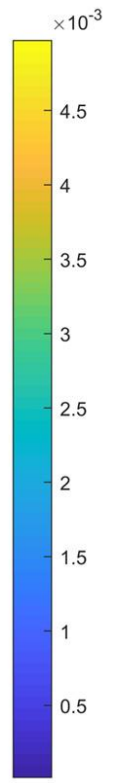
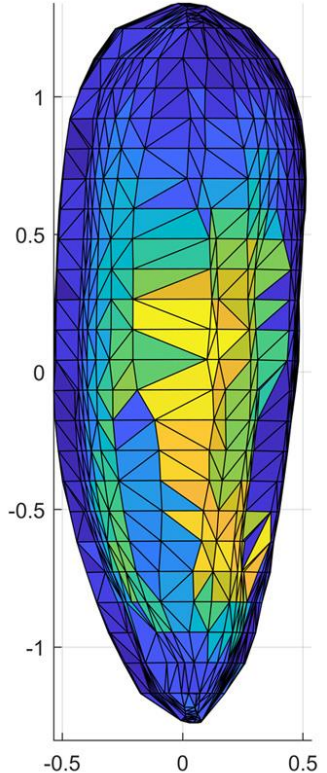
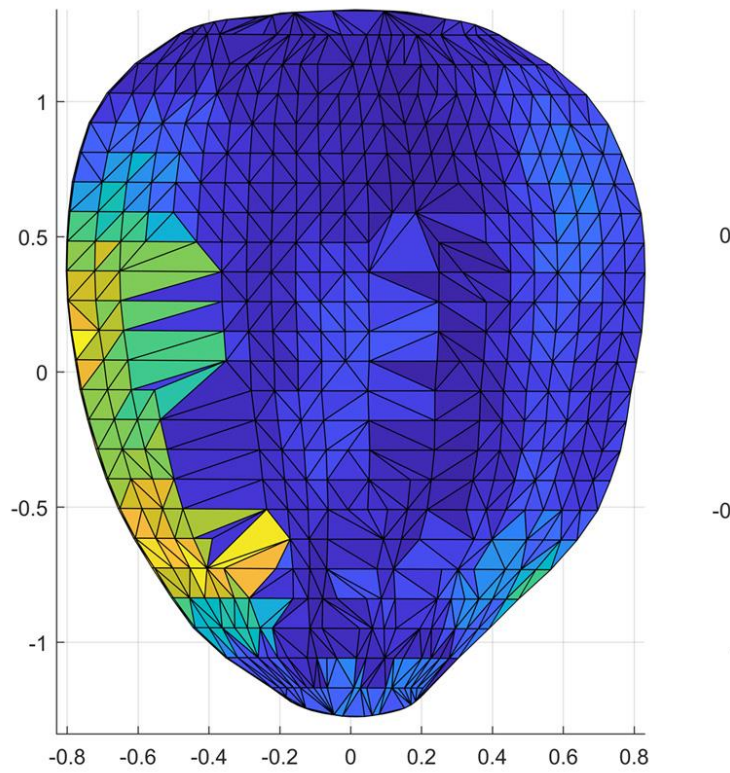
### Train the classifier

```
In [10]: from sklearn.ensemble import RandomForestClassifier
```

```
In [11]: clf = RandomForestClassifier(n_estimators=10)
clf = clf.fit(x_train, y_train)
```

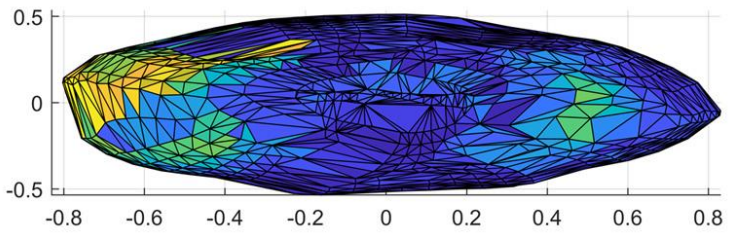
### Apply the classifier to unknown seeds

```
In [12]: print(clf.predict([[8.12,4.6,3.53,7.99,4.55,3.52,5.02,6.24,3.33,4.4,4.12,2.1,3.01,3.48]]))
['Landrace2']
```



← 5 microns

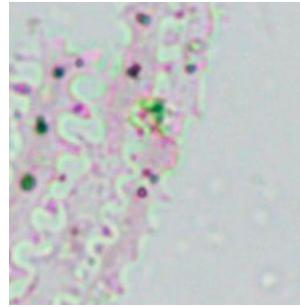
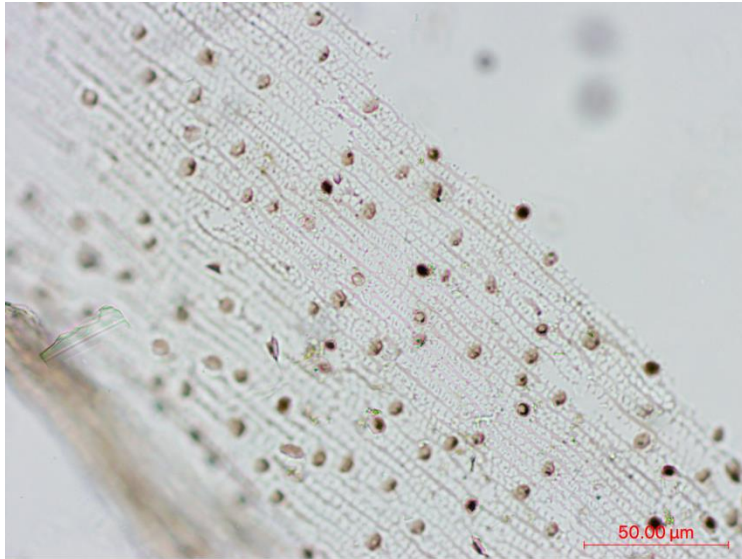
← 0.5 microns



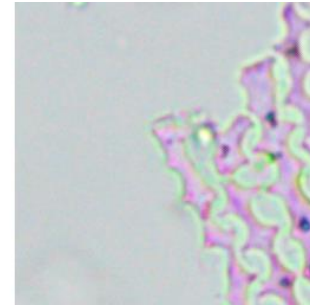
**Analitzant les diferències de forma  
entre grups de llavors**



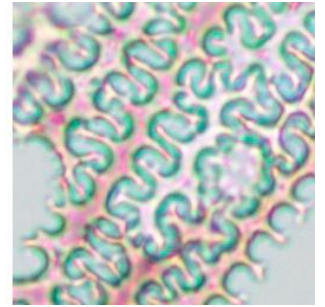
# Phytolith identification from microscope images using CNNs



*Avena*



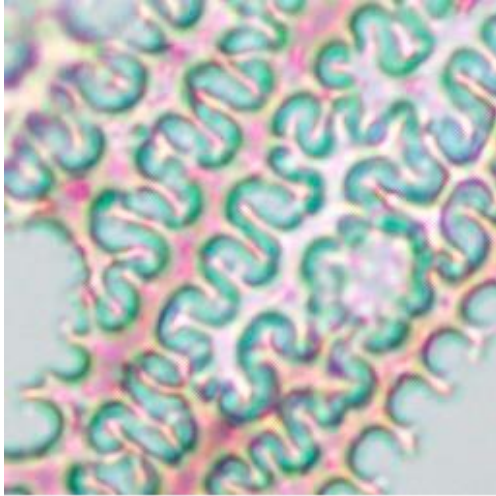
*Hordeum*



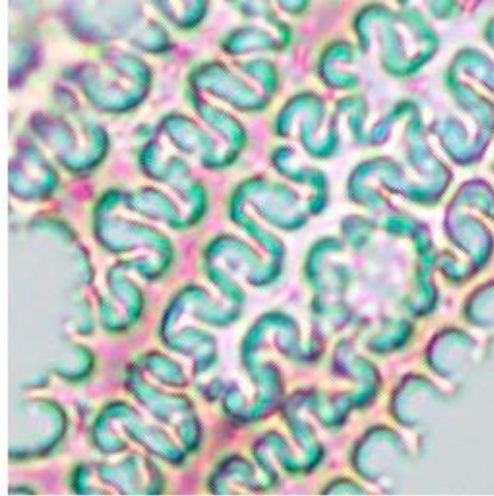
*Triticum*

*Avena* phytolith

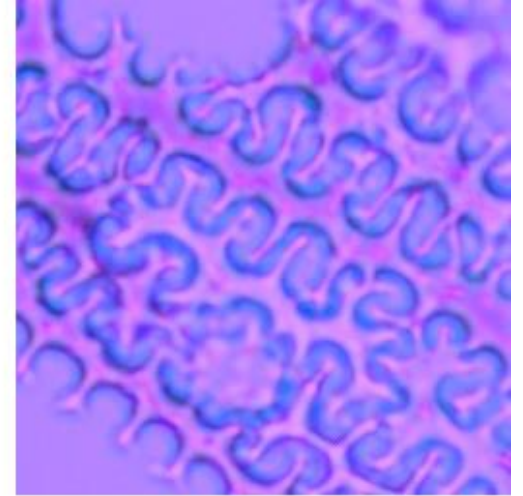
# Tècniques d'augment de dades



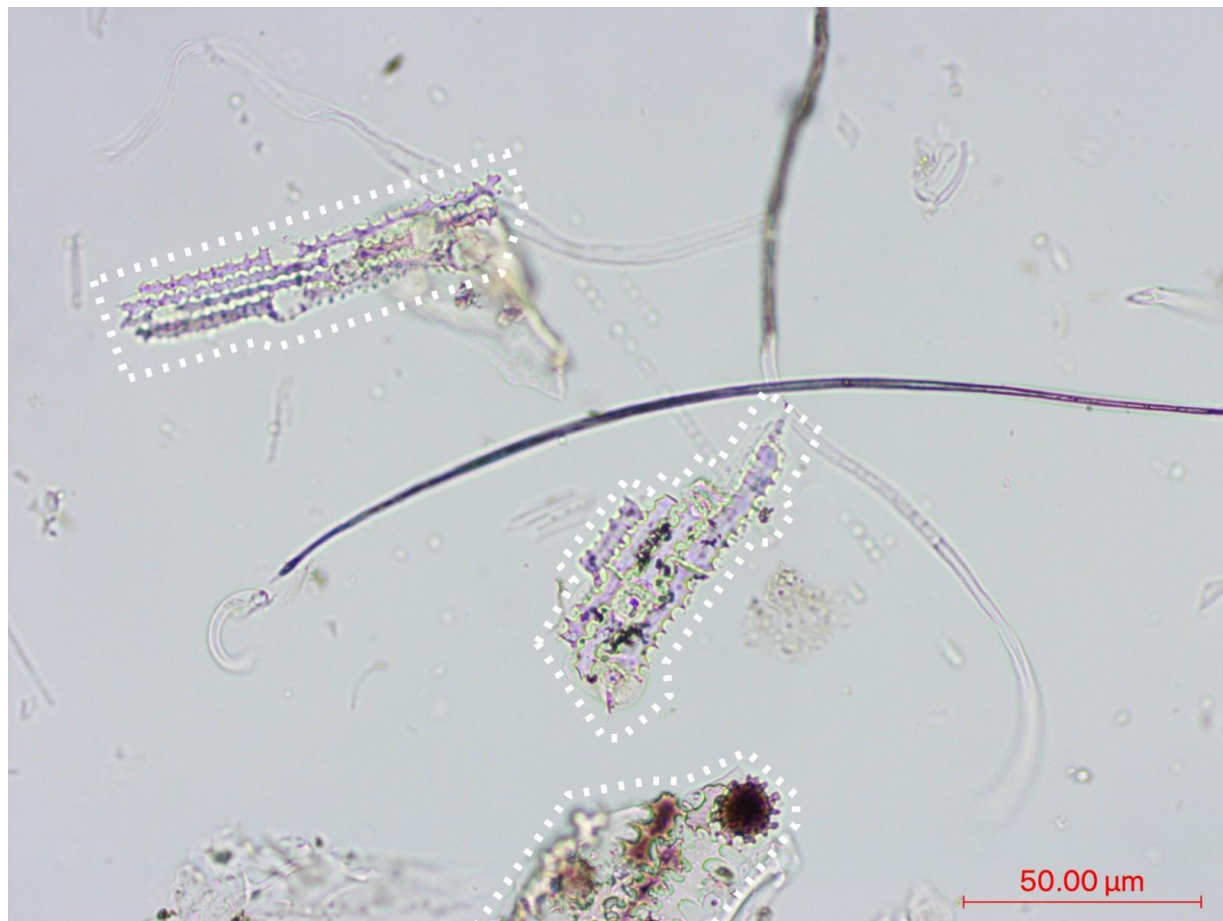
Original



La nitidesa i el desenfocament s'implementen per imitar els canvis d'enfocament en les imatges del microscopi

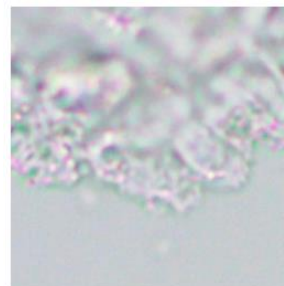
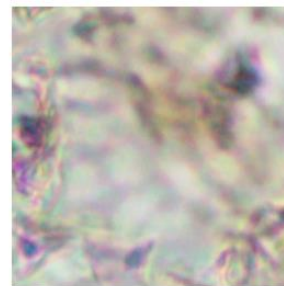


La fluctuació del color s'implementa per ignorar la coloració variable a causa del tractament químic i les variacions del microscopi / càmera  
La rotació s'implementa per multiplicar dades



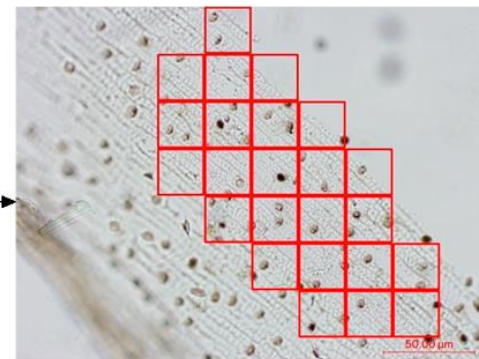
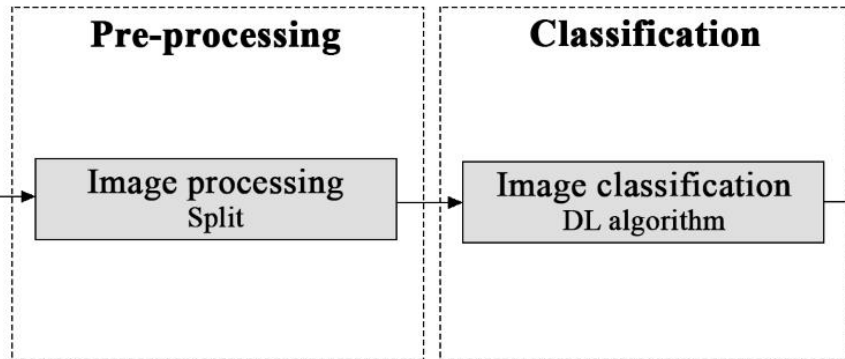
Presència de diversos objectes a la imatge que no són fitòlit

També es van incloure com a dades d'entrenament juntament amb imatges de fons





Microscope image



Identified phytoliths

Predicted classes

	<i>Avena</i>	<i>Hordeum</i>	<i>Triticum</i>	Artifacts	Back.
<i>Avena</i>	95.84%	0.00%	0.00%	2.08%	2.08%
<i>Hordeum</i>	8.33%	79.17%	6.25%	2.08%	4.17%
<i>Triticum</i>	2.08%	0.00%	93.75%	4.17%	0.00%
Artifacts	0.00%	0.00%	2.08%	81.25%	16.67%
Back.	0.00%	0.00%	0.00%	0.00%	100.00%

Refinement of the results and combination of detection and identification algorithms

#	Refinement	Overall Acc.	<i>Avena</i> Acc.	<i>Hordeum</i> Acc.	<i>Triticum</i> Acc.
1	Yes	93.68%	98.44%	93.55%	89.06%
2	No	88.95%	93.75%	91.94%	81.25%

Algoritme llest per ser implementat en fluxos de treball de microscopi digital

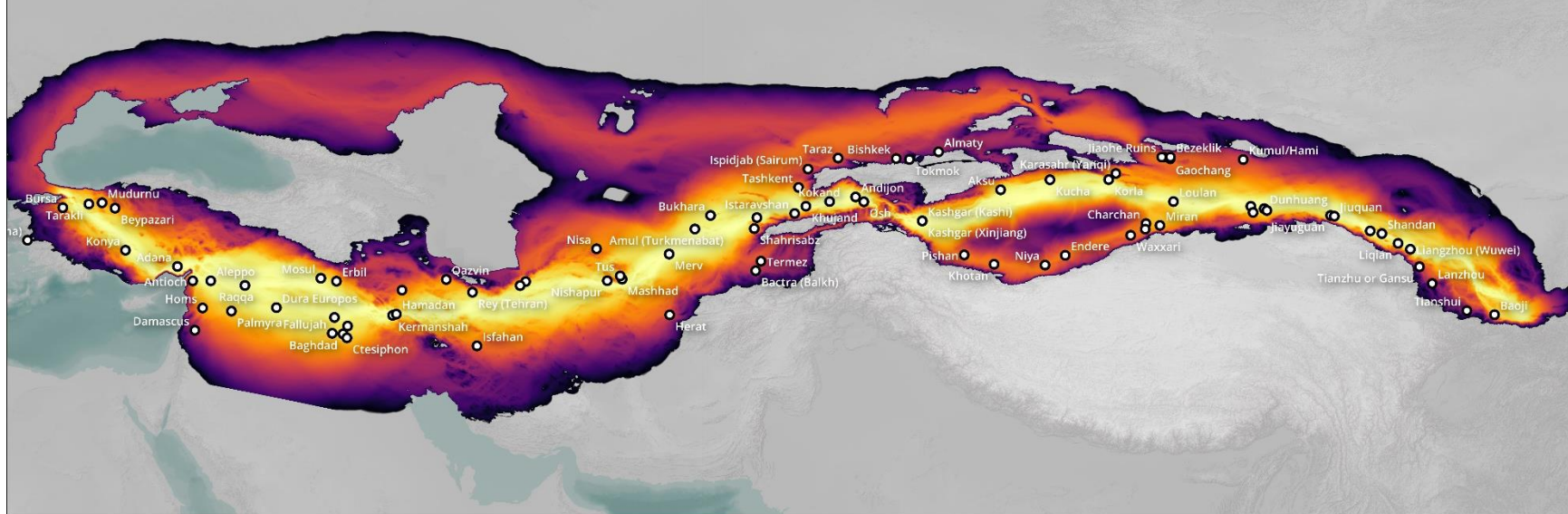
Resultats de la classificació automatitzada de fitòlits en llàmines de microscopi



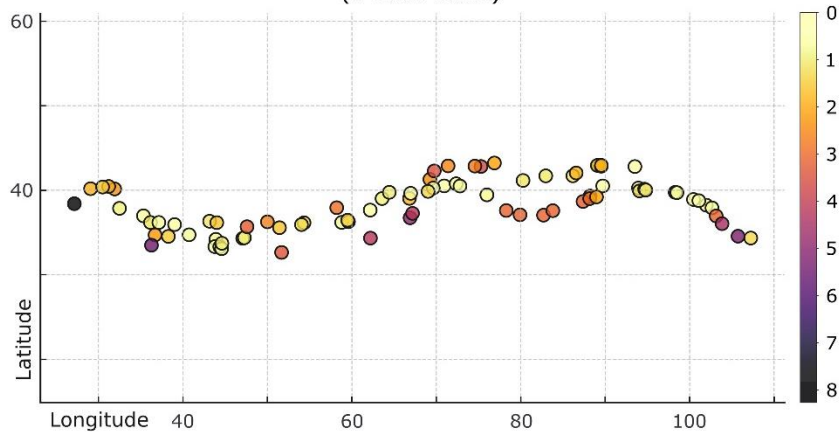
Modelització de rutes  
transcontinentals

**VIII JORNADA  
DE PATRIMONI  
CULTURAL**

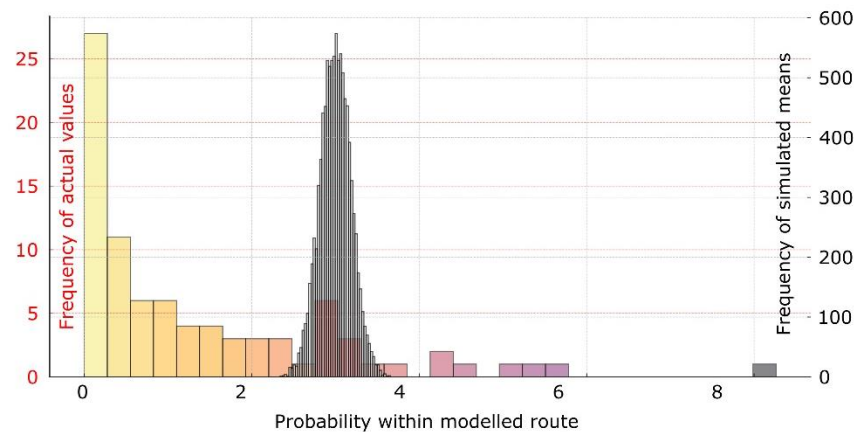
*LA INTEL·LIGÈNCIA  
ARTIFICIAL APLICADA  
AL PATRIMONI  
CULTURAL*



Location of sites according to probability within modelled Silk Road route (0-1000 scale)



Comparison between actual probability values and Monte Carlo simulation



Anàlisi de dades de  
Percepció de l'entorn

**VIII JORNADA  
DE PATRIMONI  
CULTURAL**

*LA INTEL·LIGÈNCIA  
ARTIFICIAL APLICADA  
AL PATRIMONI  
CULTURAL*

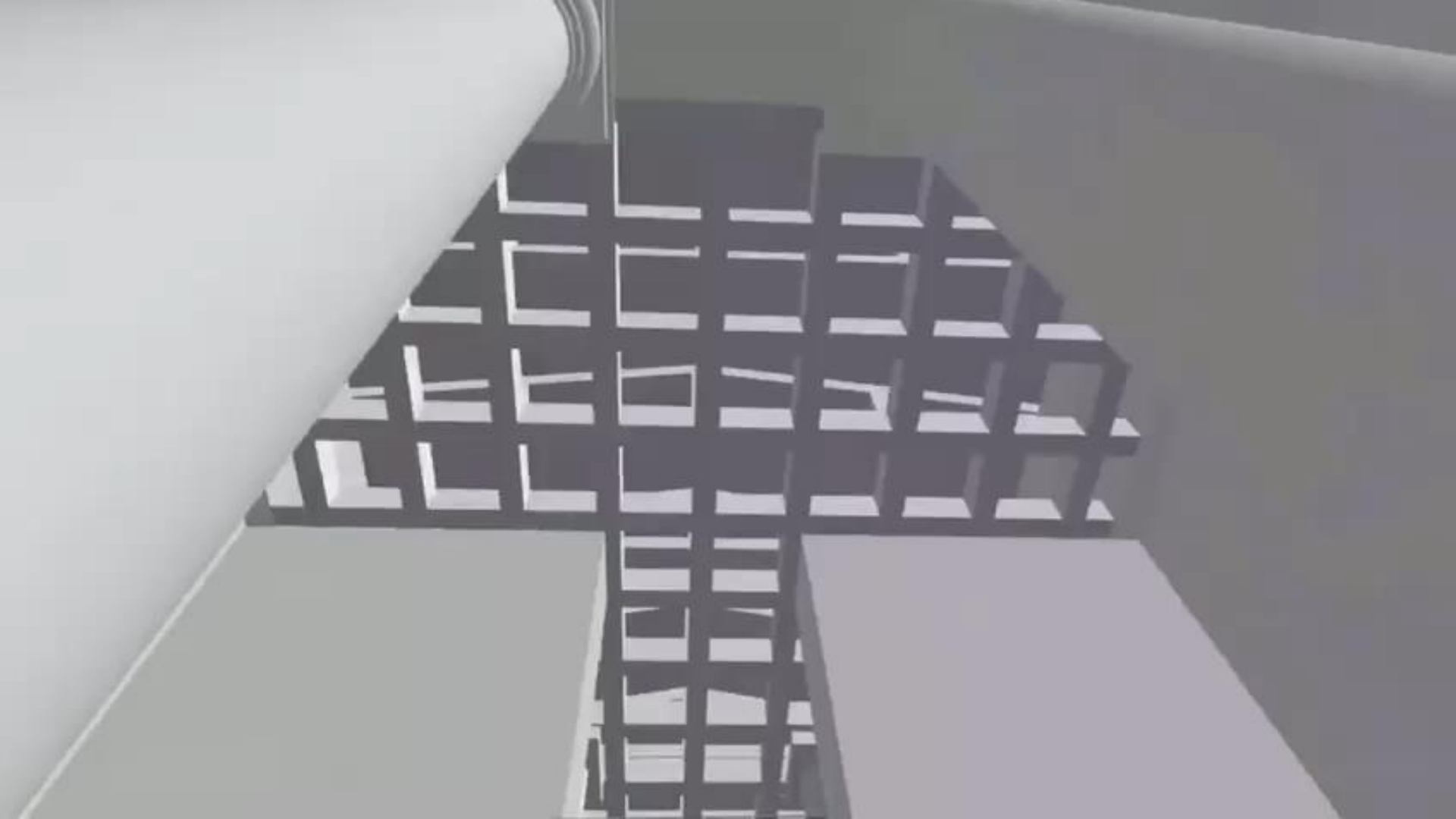








STAR-AGESS MSCA project co-directed with Prof. E. Boutsikas to create georeferenced exploratory virtual environments contextualised in their larger space (horizon) and the skyscape



# Conclusions

## Vivim un moment apassionant per a l'aplicació de l'aprenentatge automàtic en arqueologia

- Increment disponibilitat de fonts i repositoris accessibles, incloses constel·lacions virtuals de dades multifont i multitemporals
- Augment de l'accés a grans recursos informàtics
- Mètodes de captura de dades cada vegada més barats i eficients
- L'aprenentatge automàtic, la ciència de xarxes i les estadístiques ofereixen una manera única de donar sentit a aquests conjunts de dades complexes i donar-les una aplicació pràctica
- Es poden combinar utilitzant diferents escales amb diferents objectius per crear fluxos de treball únics per resoldre problemes aplicats
- El seu potencial encara està, en gran part, sense explotar
- Proporcionen una gran quantitat d'informació d'alta qualitat, que és essencial per planificar i executar la gestió del patrimoni

Tot i això:

- Cal incrementar els repositoris de dades amb dades específiques pel patrimoni
- Es necessari fer servir combinacions de fonts i algoritmes
- S'han de desenvolupar processos específics pel tractament i anàlisi de dades patrimonials
- Cal desenvolupar formes de d'augment de dades, incloent-hi dades sintètiques
- Cal implicar els professionals del patrimoni en desenvolupament dels entrenaments del algoritmes
- Es necessari realitzar validacions extensives

Gràcies per la vostra atenció



GOBIERNO  
DE ESPAÑA

MINISTERIO  
DE CIENCIA  
E INNOVACIÓN



Els treballs presentats han estat realitzats en col·laboració amb:



Iban Berganzo-Besga



Felipe Lumbreras



Efrosyni Boutsikas



Nazarij Buławka



Arnau Garcia-Molsosa



Francesc C. Conesa



Paloma Aliende

Molts altres col·laboradors han participat en el desenvolupament d'aquests treballs