Un nou algoritme d'aprenentatge automàtic permet detectar prop de 10.000 túmuls arqueològics a Galícia

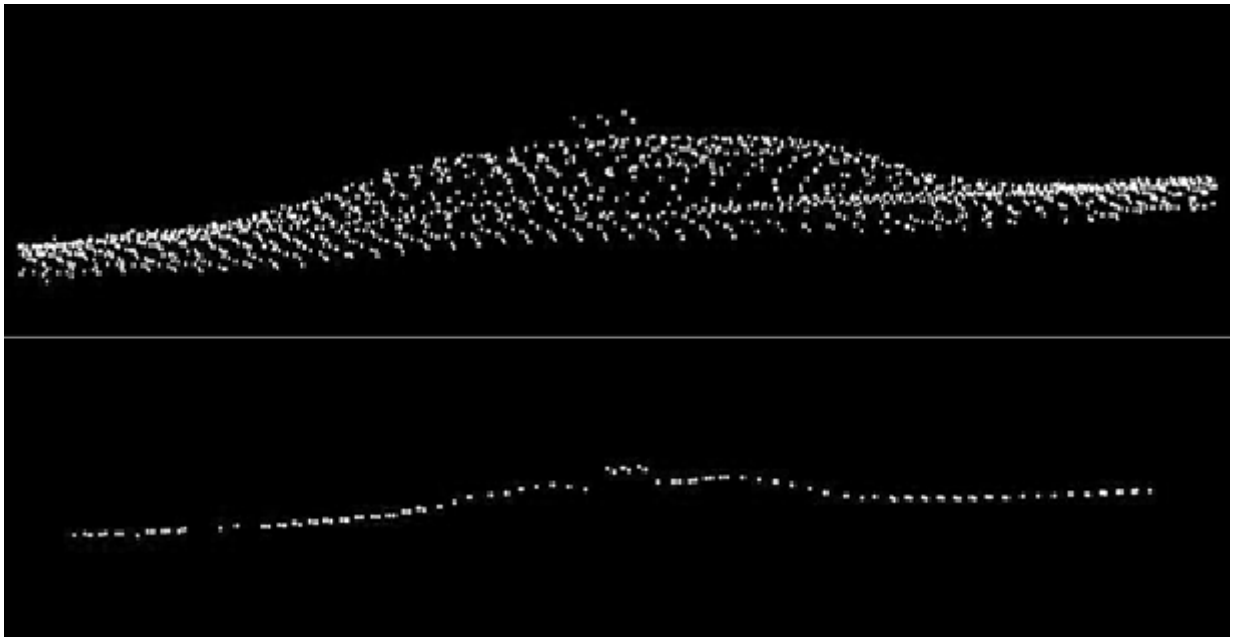
Els **túmuls arqueològics** són un dels tipus més comuns de llocs d'interès arqueològic, i es poden trobar arreu del món. Aquesta és potser la raó per la qual molts estudis han provat de desenvolupar mètodes per a automatitzar la seva detecció.

La seva **forma característica de monticle** ha estat el tret principal per a la seva identificació en el terreny i mitjançant l'ús de dades topogràfiques basades en LiDAR, que generalment prenen la forma de Models Digitals de Terreny (en anglès, *Digital Terrain Models* o DTM).

La forma simple dels túmuls és ideal per a la seva detecció mitjançant enfocaments d'aprenentatge profund (*deep learning*). Els detectors d'aprenentatge profund generalment requereixen grans quantitats de dades d'entrenament (de l'ordre de milers d'exemples) per a poder produir resultats significatius.

No obstant, la **forma homogènia semi-hemisfèrica** dels túmuls permet que els detectors funcionin òptimament amb una quantitat molt menor de dades d'entrenament, reduint considerablement l'esforç requerit per a obtenir-lo i els importants recursos computacionals necessaris per a entrenar un detector de xarxa neuronal convolucional (CNN).

Contingut extret de la web de l'ICAC (www.icac.cat)



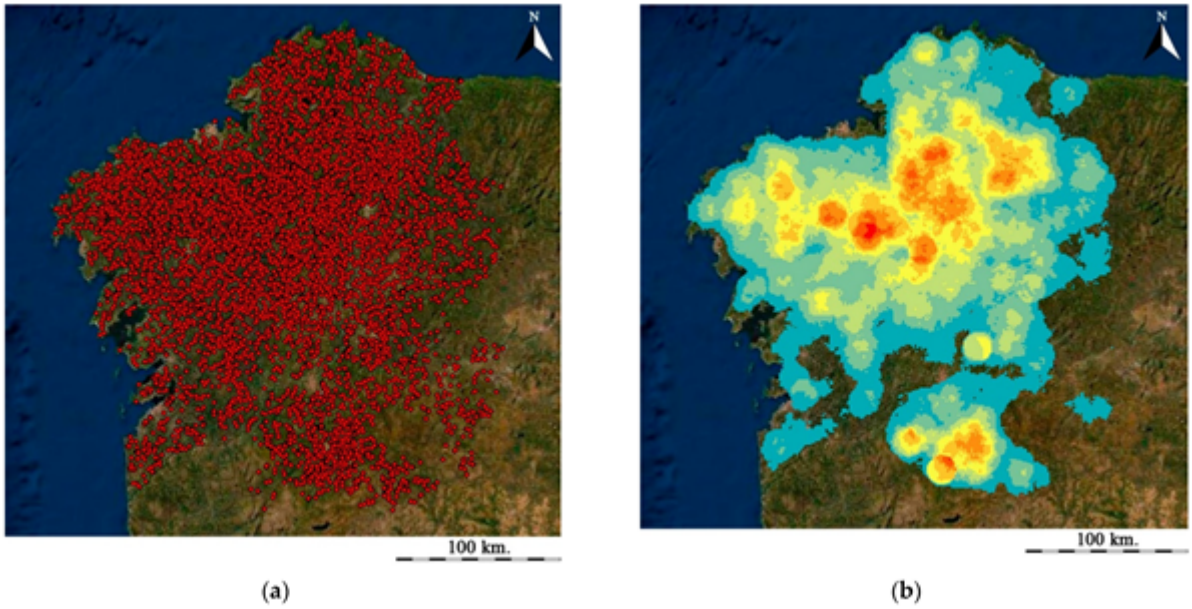
Dades topogràfiques basades en LiDAR. Imatge: Miguel Carrero-Pazos.

Però, alhora, aquesta forma comuna, simple i regular dels túmuls pot representar també un inconvenient important, i és que la forma és similar a moltes altres característiques no arqueològiques i, per tant, els estudis que implementen mètodes per a la detecció de monticles en DTM derivats de LiDAR i altres conjunts de dades d'alta resolució es caracteritzen per **una gran presència de falsos positius** (objectes identificats incorrectament com a monticles).

El nou algortime proporciona un camí a seguir per a la detecció de túmuls amb una menor presència de falsos positius.

Els membres de l'equip GIAP de l'Institut Català d'Arqueologia Clàssica (ICAC) Iban Berganzo i Hèctor A. Orengo es van trobar amb aquesta situació en portar a terme les primeres fases d'un estudi sobre detecció automàtica de túmuls a Galícia. Fa poc es va publicar una breu nota sobre aquesta recerca inicial, en què es van localitzar gairebé 9.000 túmuls. No obstant, no tots eren túmuls reals, ja que els resultats de la detecció automatitzada també incloïen falsos positius.

Contingut extret de la web de l'ICAC (www.icac.cat)



Tumuli detectats a Galícia: (a) distribució; (b) mapa de calor ('heatmap'). Autoria: Iban Berganzo (ICAC).

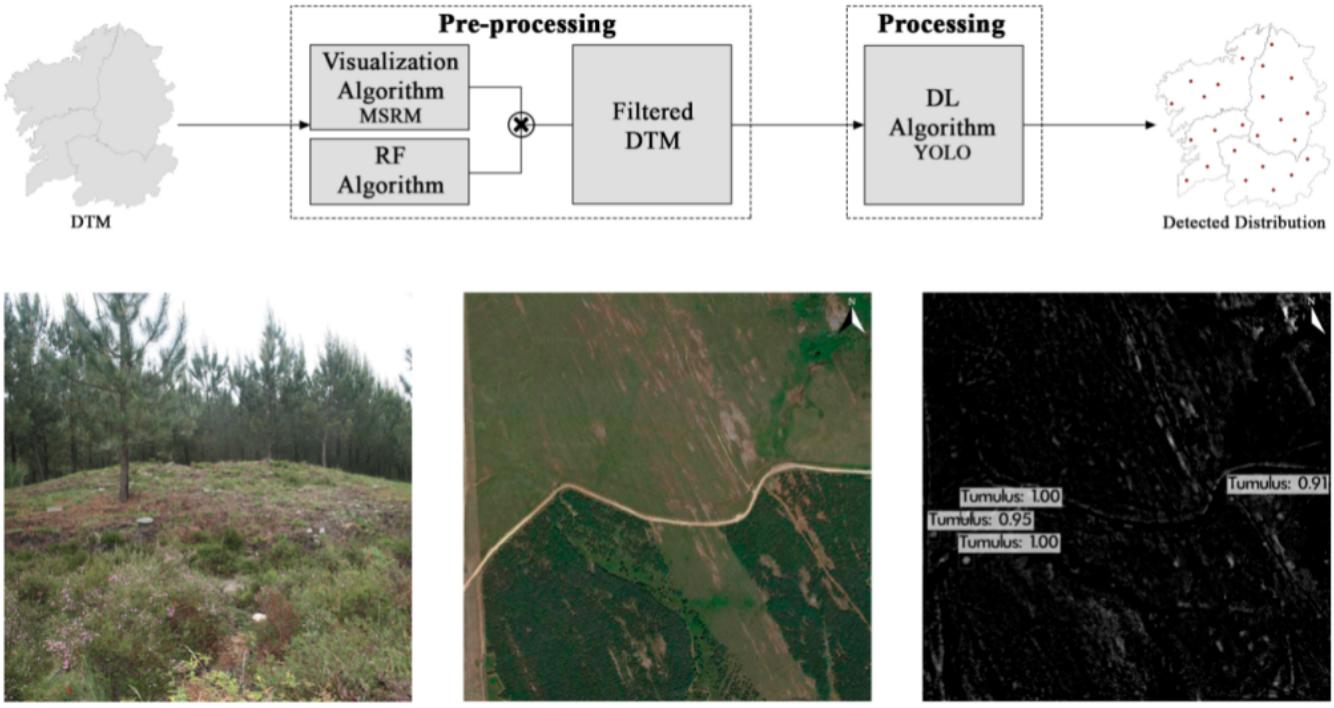
Després de la validació inicial de les dades, que es va fer en col·laboració amb els col·legues del projecte el Dr. Miguel Carrero (University College London / Universitat de Santiago de Compostel·la, GEPN-AAT), el Dr. João Fonte (Universidad de Exeter) i el Dr. Benito Vilas (Universidad de Vigo), els investigadors es van adonar que, del recompte inicial de prop de 9.000 objectes detectats, només prop de 7.600 corresponien a monticles arqueològics reals.

Si bé aquest era **un resultat excel·lent**, tenint en compte que la majoria dels estudis similars presenten percentatges de falsos positius encara més baixos, l'equip va pensar que podia millorar la taxa de detecció i disminuir el nombre de falsos positius.

Durant l'estiu de 2021, Iban Berganzo i Hèctor A. Orengo, en col·laboració amb el Dr. Felipe Lumbreras del Centre de Visió per Computador (CVC), han desenvolupat un nou algoritme per a reduir el nombre de falsos positius i augmentar la taxa de detecció.

Després d'analitzar la naturalesa dels falsos positius detectats, aquest equip d'investigadors ha desenvolupat **un enfocament híbrid que combina l'aprenentatge automàtic clàssic (*machine learning*) i l'aprenentatge profund (*deep learning*)**. L'objectiu era obtenir una definició més precisa de túmuls arqueològics en què no només es tingui en compte la forma sinó també les característiques multiespectrals dels objectes a l'hora de buscar túmuls.

Els primers resultats han vist ara la llum en un article científic publicat a la revista ***Remote Sensing*** (en accés obert), una de les principals revistes científiques de la disciplina. En aquest article, els investigadors amplien les dades analitzades i la informació sobre aquesta innovadora iniciativa de detecció automàtica basada en la computació.



Imatge publicada a l'article de 'Remote Sensing', Berganzo et al., 2021: Graphical Abstract.

En concret, els resultats que ha produït aquesta nova aproximació són força espectaculars:

- L'**àrea coberta és de quasi 30.000 km^2**. Representa l'àrea més gran (de què tinguin

coneixement els investigadors) en què s'hagin aplicat enfocaments arqueològics.

- S'han detectat 10.527 llocs d'interès, dels quals aproximadament 9.422 corresponen a túmuls arqueològics (havent fet una curosa validació visual amb imatges d'alta resolució i pendent de validació del terreny). És a dir, **un 89,5% dels túmuls detectats corresponen a veritables positius**.
- Només s'han emprat **dades de codi obert** en aquesta investigació. No obstant, l'ús de dades de major resolució, en particular imatges per satèl·lit de major resolució en lloc de les imatges Sentinel 2 (10m/px) emprades, disminuiria radicalment el nombre de falsos positiu i s'arribaria a una taxa d'èxit superior al 97%.
- El codi, les fonts i els resultats (inclosa la validació) estan disponibles de forma gratuïta i el codi està dissenyat per a ser utilitzat en plataformes de computació al núvol de lliure accés (Google Colaboratory i Earth Engine), per la qual cosa la falta de recursos computacionals no suposarà un problema per a la seva aplicació a altres àrees d'estudi.

  Una aproximació innovadora que proporciona un camí a seguir per a la detecció de túmuls amb una menor presència de falsos positiu. Un altre aspecte positiu és que **l'algoritme es pot aplicar en àrees del món on no es disposa de dades topogràfiques de suficient resolució**. En general, al proporcionar dades d'entrenament específiques, aquesta aproximació híbrida també es pot utilitzar per a detectar altres tipus de característiques en què un gran nombre de falsos positius són un problema.



Túmul de Touro Morto (Oia, Galícia). Foto: Miguel Carrero-Pazos.

Finançament

Aquesta recerca ha rebut finançament de múltiples fonts, a les que cal agrair la seva contribució:

- El doctorat de l'Iban Berganzo està finançat amb una Ajuda a Equips d'Investigació Científica de la Fundació BBVA per al Projecte DIASur;
- Hèctor A. Orengo és becari Ramón y Cajal (RYC-2016-19637) del Ministeri de Ciència, Innovació i Universitats;
- El treball de Felipe Lumbreras està finançat en part pel projeceo BOSSS TIN2017-89723-P del Ministeri de Ciència i Innovació;
- Miguel Carrero i João Fonte són becaris Marie Skłodowska-Curie (acords de subvenció

Contingut extret de la web de l'ICAC (www.icac.cat)

886793 i 794048, respectivament).

- D'altra banda, algunes de les GPU emprades en els experiments són una donació del Programa de subvencions de *hardware* de Nvidia.