Un nuevo algoritmo de aprendizaje automático permite detectar cerca de 10.000 túmulos arqueológicos en Galicia

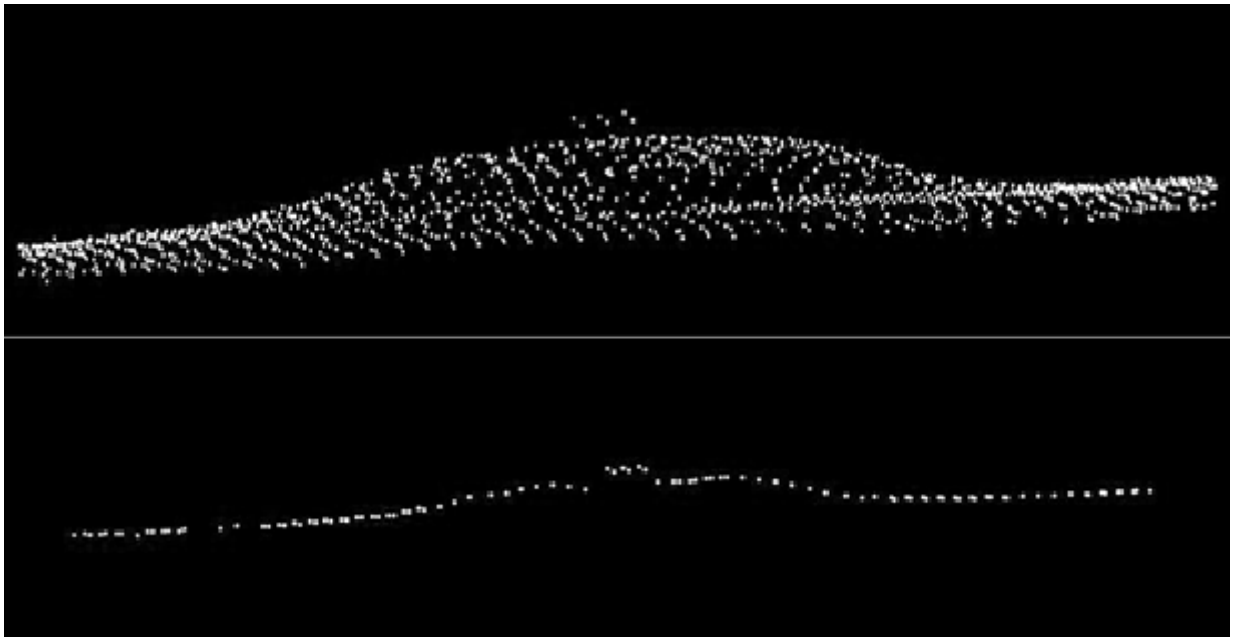
Los **túmulos arqueológicos** son uno de los tipos más comunes de sitios de interés arqueológico, y se pueden encontrar en todo el mundo. Esta es quizás la razón por la que muchos estudios han intentado desarrollar métodos para automatizar su detección.

Su **forma característica de montículo** ha sido el rasgo principal para su identificación en el terreno y mediante el uso de datos topográficos basados en LiDAR, que generalmente toman la forma de Modelos Digitales de Terreno (en inglés, *Digital Terrain Models* o DTM).

La forma simple de los túmulos es ideal para su detección mediante enfoques de aprendizaje profundo (*deep learning*). Los detectores de aprendizaje profundo generalmente requieren grandes cantidades de datos de entrenamiento (del orden de miles de ejemplos) para poder producir resultados significativos.

Sin embargo, la **forma homogénea semi-hemisférica** de los túmulos permite que los detectores funcionen óptimamente con una cantidad mucho menor de datos de entrenamiento, reduciendo considerablemente el esfuerzo requerido para obtenerlo y los importantes recursos computacionales necesarios para entrenar un detector de red neuronal convolucional (CNN).

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



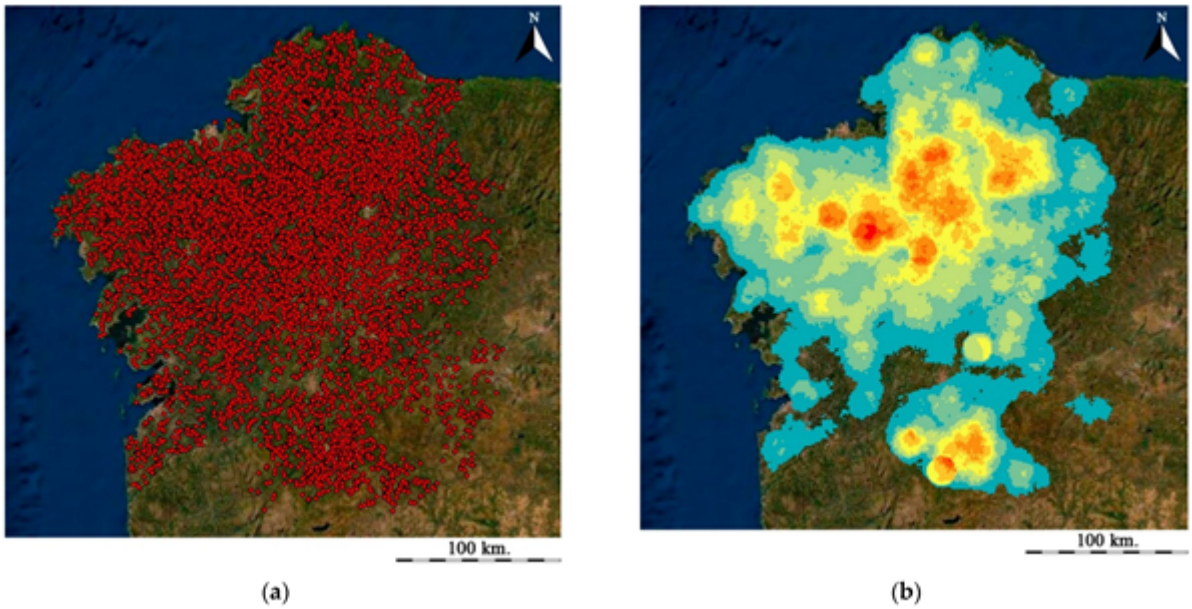
Datos topográficos basados en LiDAR. Imagen: Miguel Carrero-Pazos.

Al mismo tiempo, esta forma común, simple y regular de los túmulos puede representar también un inconveniente importante, y es que la forma es similar a muchas otras características no arqueológicas y, por lo tanto, los estudios que implementan métodos para la detección de montículos en DTM derivados de LiDAR y otros conjuntos de datos de alta resolución se caracterizan por **una gran presencia de falsos positivos** (objetos identificados incorrectamente como montículos).

El nuevo algortimo proporciona un camino a seguir para la detección de túmulos con una menor presencia de falsos positivos.

Los miembros del equipo GIAP del Instituto Catalán de Arqueología Clásica (ICAC) Iban Berganzo y Hèctor A. Orengo se encontraron con esta situación al llevar a cabo las primeras fases de un estudio sobre detección automática de túmulos en Galicia. Hace poco se publicó una breve nota sobre esta investigación inicial, en la que se localizaron casi 9.000 túmulos. Sin embargo, no todos eran túmulos reales, ya que los resultados de la detección

automatizada también incluían falsos positivos.



Tumuli detectados en Galicia: (a) distribución; (b) mapa de calor (*heatmap*). Autoría: Iban Berganzo (ICAC).

Tras la validación inicial de los datos, que se realizó en colaboración con los colegas del proyecto el Dr. Miguel Carrero (University College London / Universidad de Santiago de Compostela, GEPN-AAT), el Dr. João Fonte (Universidad de Exeter) y el Dr. Benito Vilas (Universidad de Vigo), los investigadores se dieron cuenta de que, del recuento inicial de cerca de 9.000 objetos detectados, solo cerca de 7.600 correspondían a montículos arqueológicos reales.

Aunque este era **un resultado excelente**, dado que la mayoría de los estudios similares presentaban porcentajes de falsos positivos aún más inferiores, el equipo pensó que podía mejorar la tasa de detección y disminuir el número de falsos positivos.

Durante el verano de 2021, Iban Berganzo y Hèctor A. Orengo, en colaboración con el Dr. Felipe Lumbreras del Centro de Visión por Computador (CVC), desarrollaron un

nuevo enfoque para reducir el número de falsos positivos y aumentar la tasa de detección.

Después de analizar la naturaleza de los falsos positivos detectados, este equipo de investigadores desarrolló **un enfoque híbrido que combina el aprendizaje automático clásico (*machine learning*) y el aprendizaje profundo (*deep learning*)**. El objetivo era obtener una definición más precisa de túmulos arqueológicos en la que no solo se tenga en cuenta la forma sino también las características multiespectrales de los objetos a la hora de buscar túmulos.

Los primeros resultados han visto ahora la luz en un artículo científico publicado en la revista ***Remote Sensing*** (en acceso abierto), una de las principales revistas científicas de la disciplina. En él, los investigadores amplían los datos analizados y la información sobre esta innovadora iniciativa de detección automática basada en la computación.

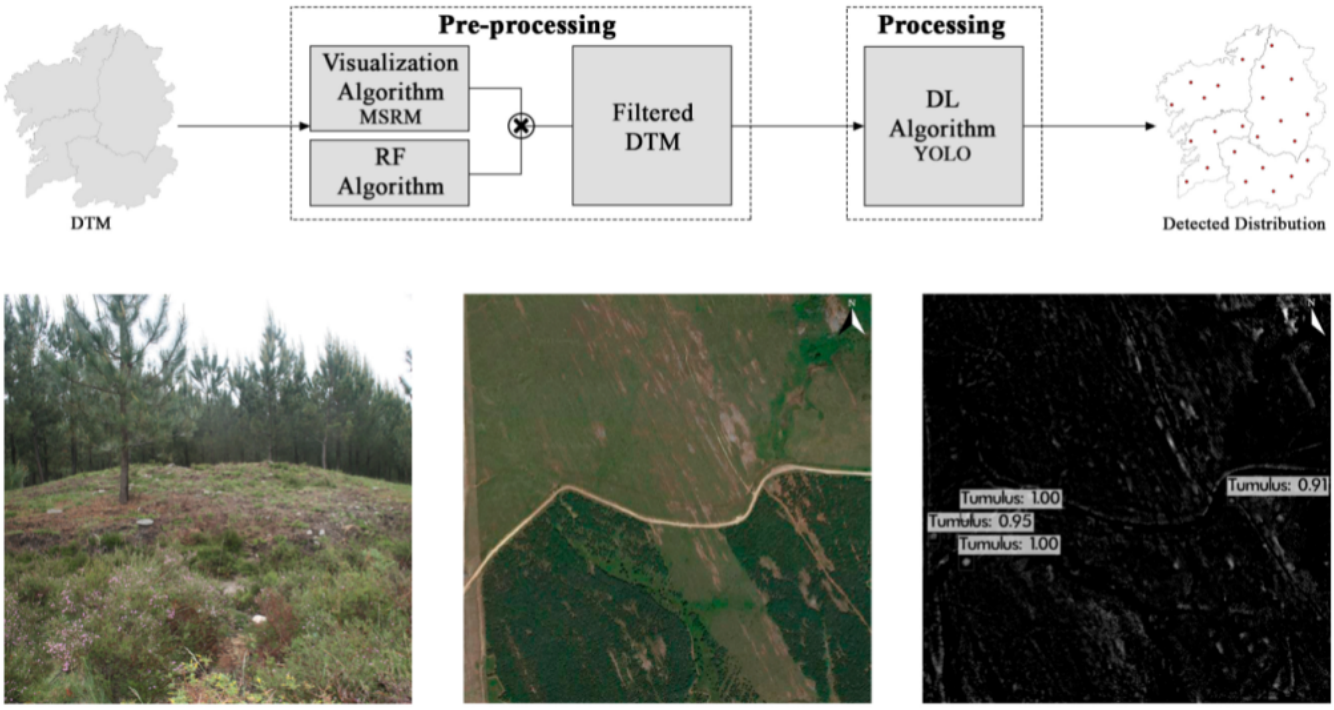


Imagen publicada en el artículo en la revista *Remote Sensing*, Berganzo et al., 2021: Graphical

En concreto, los resultados que ha producido este nuevo enfoque son nada menos que espectaculares:

- El **área cubierta es de casi 30.000 km^2**. Representa el área más grande (de la que tienen conocimiento los investigadores, por el momento) en la que se hayan aplicado enfoques arqueológicos.
- Se han detectado 10.527 sitos de interés, de los cuales aproximadamente 9.422 corresponden a túmulos arqueológicos (después de una cuidadosa validación visual con imágenes de alta resolución y pendiente de validación del terreno). Es decir, **un 89,5% de los túmulos detectados corresponden a verdaderos positivos**.
- Solo se han empleado **datos de código abierto** en esta investigación. Sin embargo, el uso de datos de mayor resolución, en particular imágenes por satélite de mayor resolución en lugar de las imágenes Sentinel 2 (10m/px) empleadas, disminuiría radicalmente el número de falsos positivos y se alcanzaría una tasa de éxito superior al 97%.
- El código, las fuentes y los resultados (incluida la validación) están disponibles de forma gratuita y el código está diseñado para ser utilizado en plataformas de computación en la nube de libre acceso (Google Colaboratory y Earth Engine), por lo que la falta de recursos computacionales no supondrá un problema para su aplicación a otras áreas de estudio.
Este **enfoque innovador** proporciona un camino a seguir para la detección de túmulos con una menor presencia de falsos positivos. Otro aspecto positivo es que **el algoritmo se puede aplicar en áreas del mundo donde se no dispone de datos topográficos de suficiente resolución**. En general, al proporcionar datos de entrenamiento específicos, este enfoque híbrido también se puede utilizar para detectar otros tipos de características en las que un gran número de falsos positivos son un problema.

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



Túmulo de Touro Morto (Oia, Galicia). Foto: Miguel Carrero-Pazos.

Financiación

Esta investigación ha recibido financiación de múltiples fuentes, a las que cabe agradecer su contribución:

- El doctorado de Iban Berganzo está financiado con una Ayuda a Equipos de Investigación Científica de la Fundación BBVA para el Proyecto DIASur;
- Hèctor A. Orengo es becario Ramón y Cajal (RYC-2016-19637) del Ministerio de Ciencia, Innovación y Universidades de España;
- El trabajo de Felipe Lumbreras está financiado en parte por el proyecto BOSSS TIN2017-89723-P del Ministerio de Ciencia e Innovación de España;

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

- Miguel Carrero y João Fonte son becarios Marie Skłodowska-Curie (acuerdos de subvención 886793 y 794048, respectivamente).
- Por otra parte, algunas de las GPU utilizadas en los experimentos son una donación del Programa de subvenciones de *hardware* de Nvidia.

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