Congresos UEx. Actas de Congresos. ISSN: 3101-7177,
vol. 2, 2026, XXI Congreso de la Asociación Española de Teledetección
DOI: https://doi.org/10.17398/3101-7177.2.270

Influencia del diseño muestral en la estimación del peligro de incendio forestal mediante XGBoost en el oeste de la Península Ibérica

Martínez García, Francisco Manuel [1] **Jiménez Romanillos, José** [2] **García Ramajo, Jorge Juan** [1] **Quirós, Elia** [1]

[1] Universidad de Extremadura, España

[2] Junta de Extremadura

ORCID: Martínez García 0000-0002-0862-9933 Jiménez Romanillos 0009-0009-0612-4390 García Ramajo 0009-0000-7296-3082 Quirós 0000-0002-8429-045X

Correspondencia: fmmgarcia@unex.es jose.jimenez@juntaex.es jjgarciar@unex.es equiros@unex.es

RESUMEN

Los incendios forestales han aumentado en frecuencia y severidad en el oeste de la Península Ibérica, en un contexto caracterizado por episodios térmicos extremos, sequías prolongadas y modificaciones estructurales del paisaje. Este trabajo analiza la capacidad del algoritmo XGBoost para estimar el peligro de incendio mediante la integración de variables meteorológicas de ERA5 land, índices espectrales de Sentinel-2 y factores territoriales estáticos. Se evalúa específicamente la influencia del diseño de muestreo, comparando tres configuraciones con distinta proporción de puntos negativos (50%, 75% y 100%) respecto a los positivos. Los modelos presentan un rendimiento elevado y estable (AUC entre 0,960 y 0,967; exactitud entre el 88,5 y el 90,2%), con ligeras mejoras al incrementar la representación de la clase no quemada. La humedad atmosférica, la cobertura del suelo y los indicadores del estado hídrico de la vegetación destacan entre los predictores más influyentes.

Palabras clave: *incendios forestales; XGBoost; peligro de incendio*

Fecha de recepción: 19 febrero 2026 · **Fecha de aceptación:** 19 febrero 2026

Influencia del diseño muestral en la estimación del peligro de incendio forestal mediante XGBoost en el oeste de la Península Ibérica

Martínez García, Francisco Manuel [(1)], Jiménez Romanillos, José [(2)], García Ramajo, Jorge Juan [(1)], Quirós, Elia [(1)]

[(1)] Universidad de Extremadura, España.

0000-0002-0862-9933, fmmgarcia@unex.es ; 0009-0000-7296-3082, jjgarciar@unex.es

0000-0002-8429-045X, equiros@unex.es.

[(2)] Junta de Extremadura.

0009-0009-0612-4390, jose.jimenez@juntaex.es

Resumen: Los incendios forestales han aumentado en frecuencia y severidad en el oeste de la Península Ibérica, en un contexto caracterizado por episodios térmicos extremos, sequías prolongadas y modificaciones estructurales del paisaje. Este trabajo analiza la capacidad del algoritmo XGBoost para estimar el peligro de incendio mediante la integración de variables meteorológicas de *ERA5 land*, índices espectrales de *Sentinel-2* y factores territoriales estáticos. Se evalúa específicamente la influencia del diseño de muestreo, comparando tres configuraciones con distinta proporción de puntos negativos (50%, 75% y 100%) respecto a los positivos. Los modelos presentan un rendimiento elevado y estable (AUC entre 0,960 y 0,967; exactitud entre el 88,5 y el 90,2%), con ligeras mejoras al incrementar la representación de la clase no quemada. La humedad atmosférica, la cobertura del suelo y los indicadores del estado hídrico de la vegetación destacan entre los predictores más influyentes.

Palabras clave: incendios forestales; XGBoost; peligro de incendio.

Influence of Sampling Design on Wildfire Danger Estimation Using XGBoost in the Western Iberian Peninsula

***Abstract**: Forest fires have increased in frequency and severity in the west of the Iberian Peninsula, in a context characterized by extreme heat waves, prolonged droughts, and structural changes to the landscape. This study analyzes the capacity of the XGBoost algorithm to estimate fire risk by integrating ERA5 meteorological variables, Sentinel-2 spectral indices, and static territorial factors. The influence of the sampling design is specifically evaluated by comparing three configurations with different proportions of negative points (50%, 75%, and 100%) relative to positive points. The models show high and stable performance (AUC between 0.960 and 0.967; accuracy between 88.5 and 90.2%), with slight improvements when increasing the representation of the unburned class. Atmospheric humidity, land cover, and vegetation water status indicators stand out among the most influential predictors.*

***Keywords**: wildfire; XGBoost; fire danger.*

1. INTRODUCCIÓN

Los incendios forestales constituyen una de las principales amenazas ambientales en el sur de Europa, con especial incidencia en las regiones mediterráneas y atlánticas de la Península Ibérica, donde su frecuencia e intensidad han aumentado de forma notable en las últimas décadas. Este incremento supone una alteración significativa de los ecosistemas forestales cuando el fuego deja de responder a un régimen natural, fenómeno estrechamente vinculado al cambio climático y a la creciente presión antrópica sobre el territorio (Chuvieco *et al*., 2023). En este contexto, España y Portugal se sitúan entre los países europeos más afectados, tanto por el número de incendios como por la superficie quemada (Meira Castro *et al*., 2020).

El oeste de la Península Ibérica concentra una parte sustancial de esta problemática debido a la elevada continuidad de las masas forestales, la acumulación de combustible vegetal y la profunda transformación del paisaje asociada al abandono de usos agroforestales tradicionales (Meira Castro *et al*., 2020). Estas condiciones, combinadas con veranos cálidos y secos y episodios recurrentes de sequía están incrementando la peligrosidad de los incendios forestales (de Rigo *et al*., 2017; Chuvieco *et al*., 2023). La interacción entre factores meteorológicos de corto plazo y características estructurales del territorio genera escenarios de elevada

susceptibilidad a la ignición y propagación de incendios de gran extensión y severidad (Atalay *et al*., 2024).

En los últimos años, el avance de la teledetección ha permitido incorporar información detallada sobre el estado fisiológico de la vegetación, la humedad del combustible y las condiciones superficiales del suelo (Arcos *et al*., 2024). Paralelamente, la creciente disponibilidad de datos ambientales multifuente ha impulsado la aplicación de técnicas de aprendizaje automático capaces de modelizar relaciones no lineales complejas entre variables (Michael *et al*., 2021). Entre ellas, XGBoost se ha consolidado como una herramienta robusta y eficiente en problemas de clasificación ambiental de alta dimensionalidad.

El objetivo principal del trabajo es evaluar la capacidad del algoritmo XGBoost para estimar el peligro de incendio en el oeste de la Península Ibérica integrando variables meteorológicas de *ERA5*, índices espectrales de *Sentinel-2* y factores territoriales. Por tanto, se analiza específicamente cómo el diseño del muestreo afecta el rendimiento, estabilidad y robustez del modelo.

2. MATERIAL Y MÉTODOS

2.1. Área de estudio

El área de estudio comprende la zona limítrofe del noreste de Portugal con España (Figura 1), con una extensión total de 37 537,4 km^2. zona ha presentado alta recurrencia de incendios en las últimas décadas. Por ello, se seleccionaron y modelizaron tres incendios ocurridos entre el 7 y el 15 de julio de 2022, todos de origen natural (causados por rayos).

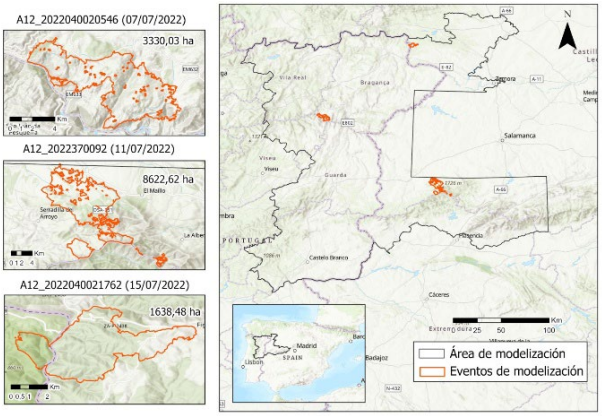

Figura 1. Área de trabajo.

2.2. Fuentes de datos

El análisis integró datos meteorológicos, variables biofísicas derivadas de teledetección y variables estáticas del territorio, con el objetivo de caracterizar los factores ambientales asociados al peligro de incendios forestales. La información meteorológica procedió del reanálisis *ERA5 land*, mientras que las variables espectrales y biofísicas se obtuvieron a partir de imágenes de *Copernicus Sentinel-2*. Las variables estáticas se recopilaron de fuentes europeas armonizadas, principalmente el *Copernicus Land Monitoring Service* (CLMS), complementadas con información de usos del suelo de la Agencia Espacial Europea (ESA) y variables topográficas derivadas de la misión *Shuttle Radar Topography Mission* (STRM) de la *National Aeronautics and Space Administration* (NASA). Este enfoque garantiza coherencia y comparabilidad en el territorio de estudio.

Para caracterizar las condiciones ambientales asociadas al peligro de incendios forestales se empleó un amplio conjunto de variables meteorológicas que incluyen parámetros atmosféricos y de superficie, radiativos y energéticos, y variables hidrológicas, así como indicadores relacionados con nieve y el suelo, incluyendo contenido de agua y temperatura. Esto permitió describir tanto las condiciones instantáneas como la acumulación de sequedad y la disponibilidad hídrica, fundamentales para la ignición y propagación de incendios (Vitolo *et al*., 2020).

La información de teledetección se utilizó para derivar diferentes índices espectrales orientados a caracterizar el estado fisiológico, estructural y hídrico de la vegetación y del combustible (Arcos *et al*., 2024). Entre ellos se incluyeron índices de vigor y actividad fotosintética, como el NDVI (Índice de Vegetación de Diferencia Normalizada); GNDVI (Índice de Vegetación de Diferencia Normalizada Verde); SAVI (Índice de Vegetación Ajustado al Suelo); MSAVI2 (Índice de Vegetación Ajustado al Suelo Modificado 2) y S2REP (Posición del Borde Rojo de Sentinel-2), así como indicadores de humedad del combustible (NDMI (Índice de Humedad de Diferencia Normalizada); NDWI (Índice de Agua de Diferencia Normalizada); MSI (Índice de Estrés Hídrico). Además, se incorporaron variables estructurales biofísicas como LAI (Índice de Área Foliar); fcover (Fracción de Cobertura Vegetal); FAPAR (Fracción de Radiación Fotosintéticamente Activa Absorbida); CI (Índice de Clorofila) y BI (Índice de brillo).

Además, se incorporaron variables estáticas del territorio, como topografía (elevación, pendiente y orientación), la cubierta vegetal (fracciones de arbolado, matorral y pastizal), el tipo de suelo, los usos del suelo y la influencia antrópica, representada mediante variables de distancia a infraestructuras y núcleos habitados, con el fin de capturar factores territoriales que condicionan la ocurrencia y propagación de incendios forestales.

2.3. Preprocesado de datos y diseño muestral

El preprocesado incluyó una estrategia de muestreo espacial con tres conjuntos de puntos, con el objetivo de representar de forma equilibrada áreas incendiadas y no incendiadas.

Se definieron puntos negativos generales distribuidos homogéneamente cada 1 km con el objetivo de caracterizar las condiciones ambientales de fondo del área de estudio. Asimismo, se establecieron puntos positivos sobre las superficies incendiadas (135,64 km²), con una separación regular de 100 m entre ellos, con el fin de capturar la variabilidad espacial directamente asociada a la ocurrencia del incendio. Finalmente, se incluyeron puntos negativos próximos, situados en el entorno inmediato de las áreas afectadas y también espaciados cada 100 m, para representar zonas no

quemadas bajo condiciones ambientales similares a las de las superficies incendiadas.

A cada punto de muestreo se asignaron variables meteorológicas de *ERA5*, índices de *Sentinel-2* y factores territoriales, calculando medias de 7 (med7), 14 (med14) y 30 (med30) días para las variables dinámicas, las cuales fueron armonizadas espacial y temporalmente para garantizar consistencia y comparabilidad en los datos.

Se evaluaron tres configuraciones (Figura 2) según la proporción de negativos respecto a positivos (50% (V1), 75% (V2) y 100% (V3)). Aunque los incendios representan una fracción reducida del territorio, el uso de conjuntos parcialmente equilibrados permite reducir sesgos hacia la clase mayoritaria y analizar comparativamente su efecto en el rendimiento e interpretación del modelo.

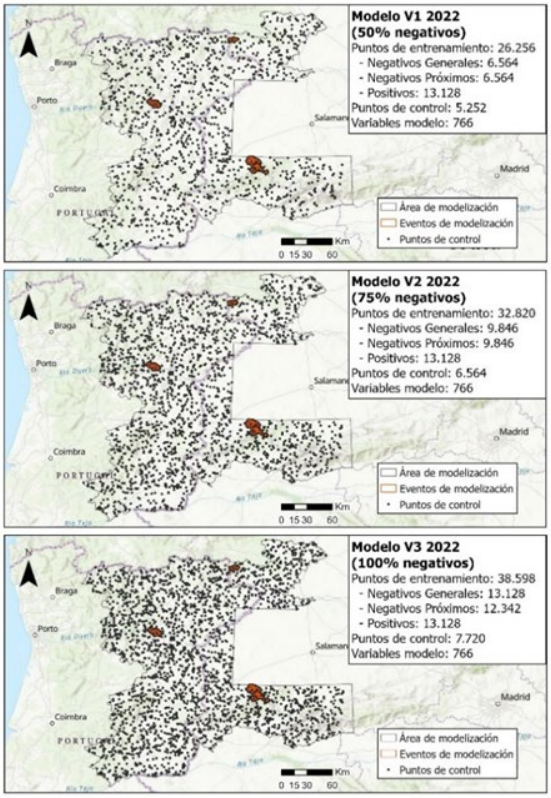

Figura 2. Distribución de puntos por selección de negativos.

Se depuró y seleccionó el conjunto de predictores para mejorar la generalización, equilibrando información y complejidad, revisando puntos negativos y ajustando variables para reducir ruido, redundancias y colinealidad.

2.4. Modelización

La modelización del peligro de incendio se realizó mediante XGBoost (*Extreme Gradient Boosting*), una técnica basada en árboles de decisión ampliamente utilizada en estudios ambientales por su solidez y capacidad predictiva (Michael *et al*., 2021). Se construyeron modelos para cada configuración muestral, corrigiendo errores secuencialmente y optimizando el ajuste. Para reducir la autocorrelación espacial, la partición entrenamiento-validación aseguró independencia geográfica entre conjuntos. Se evaluaron métricas como AUC (*Area Under the Curve*), exactitud global y tasas de falsos positivos y negativos, y la importancia de variables se analizó mediante valores SHAP (*Shapley Additive Explanations*).

3. RESULTADOS

Los resultados (Figura 3) muestran un rendimiento elevado y consistente en las tres configuraciones, con valores de AUC de 0,960 (V1), 0,965 (V2) y 0,967 (V3), y una exactitud global de 88,5%, 89,8% y del 90,2% para los verdaderos negativos (VN) y verdaderos positivos (VP).

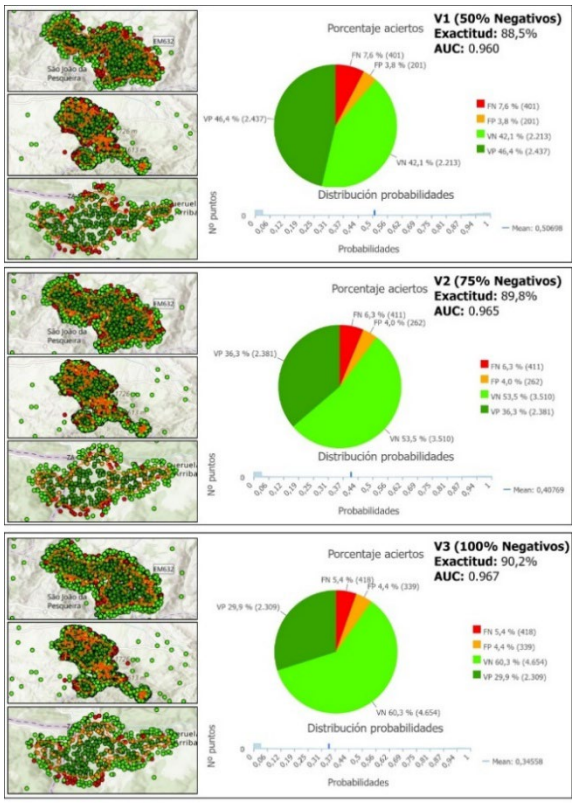

Figura 3. Resultados de modelización por versiones.

Por tanto, los resultados comparativos evidencian que estas decisiones metodológicas influyen de forma apreciable en los valores de AUC y en la precisión predictiva alcanzada por los distintos modelos evaluados.

A nivel de distribución de exactitud de los puntos, los mayores errores se concentran en zonas de transición entre áreas incendiadas y superficies negativas próximas, debido a su similitud. No obstante, el aumento de puntos negativos mejora progresivamente la capacidad discriminativa del modelo, evidenciando que un muestreo más amplio y equilibrado refuerza los resultados sin comprometer su estabilidad general. Esta mejora se refleja especialmente en la reducción de los falsos negativos (FN), que pasan del 7,6% al 5,4% a medida que se incrementa el número de puntos negativos, destacando la importancia de incluir

suficientes observaciones negativas para minimizar los errores de omisión. En cuanto a los falsos positivos (FP) se encuentran en torno al 4% de los puntos modelizados, alcanzando su valor máximo en la V3 (4,4%), probablemente debido al mayor peso de los puntos negativos en esta versión. Así, una mayor proporción de puntos negativos mejora el rendimiento global y optimiza la modelización en zonas limítrofes de incendios.

El análisis de importancia de variables (Figura 4), apoyado en los gráficos SHAP del modelo XGBoost, permitió identificar las variables con mayor contribución en las tres versiones.

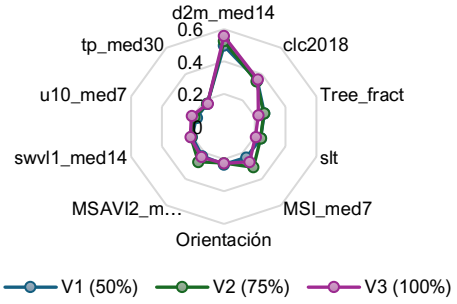

Figura 4. Importancia de variables por versión.

La variable *d2m_med14* (humedad media a 2 m) es el predictor más influyente en todas las versiones, incrementando su importancia. Le siguen *clc2018* (usos del suelo) y *Tree_fract*,(porcentaje de cobertura arbórea) con aportaciones altas y estables, lo que resalta la importancia conjunta de las características territoriales y la estructura de la vegetación. También destacan *MSI_med7* y *MSAVI2_med14*, especialmente en la V2, junto con la orientación y *swvl1_med14* (Capa volumétrica de agua del suelo 1), confirmando el carácter multifactorial del peligro de incendio.

4. CONCLUSIONES

El diseño muestral ejerce un impacto significativo en la evaluación de la exactitud de los modelos XGBoost para la estimación del peligro de incendio. El ajuste de la proporción de puntos negativos mejora la capacidad discriminativa, incrementando métricas clave como el AUC y la exactitud global, y reduciendo los falsos negativos. Esto evidencia que un muestreo equilibrado durante la fase de entrenamiento optimiza la robustez predictiva y la discriminación entre áreas incendiadas y no incendiadas.

El análisis de importancia de variables indica que el riesgo de incendio depende de la interacción entre factores meteorológicos (temperatura, humedad relativa y velocidad del viento) y características estructurales de la vegetación, como densidad, tipo y cobertura, cuya influencia se mantiene estable en distintas configuraciones.

Los resultados confirman que la combinación de variables meteorológicas, biofísicas y territoriales capta eficazmente los factores que condicionan la susceptibilidad a la ignición en el oeste de la Península Ibérica, manteniendo una influencia estable en las distintas configuraciones analizadas.

5. AGRADECIMIENTOS

Cofinanciado por la Unión Europea a través del Programa Interreg VI-A España-Portugal (POCTEP) 2021-2027.

"Redes de alertas tempranas, para la teledetección de riesgos derivados del cambio climático, por satélites de observación de la tierra para respuesta de protección civil (RAT_EOS_PC)".

6. REFERENCIAS

Arcos, M. A., Balaguer-Beser, Á., & Ruiz, L. Á. (2024). Evaluating the performance of spectral indices and meteorological variables as indicators of live fuel moisture content in Mediterranean shrublands. *Ecological Indicators*, *169*(March). https://doi.org/10.1016/j.ecolind.2024.112894

Atalay, H., Dervisoglu, A., & Sunar, A. F. (2024). Exploring Forest Fire Dynamics: Fire Danger Mapping in Antalya Region, Türkiye. *ISPRS International Journal of Geo-Information*, *13*(3). https://doi.org/10.3390/ijgi13030074

Chuvieco, E., Yebra, M., Martino, S., Thonicke, K., Gómez-Giménez, M., San-Miguel, J., Oom, D., Velea, R., Mouillot, F., Molina, J. R., Miranda, A. I., Lopes, D., Salis, M., Bugaric, M., Sofiev, M., Kadantsev, E., Gitas, I. Z., Stavrakoudis, D., Eftychidis, G., … Viegas, D. (2023). Towards an Integrated Approach to Wildfire Risk Assessment: When, Where, What and How May the Landscapes Burn. *Fire*, *6*(5), 1-60. https://doi.org/10.3390/fire6050215

de Rigo, D., Libertà, G., Houston Durrant, T., Artés Vivancos, T., & San-Miguel-Ayanz, J. (2017). Forest fire danger extremes in Europe under climate change: variability and uncertainty. En *Publication* Office *of the European Union*. https://publications.jrc.ec.europa.eu/repository/handle/JRC108974

Meira Castro, A. C., Nunes, A., Sousa, A., & Lourenço, L. (2020). Mapping the causes of forest fires in Portugal by clustering analysis. *Geosciences (Switzerland)*, *10*(2), 7-11. https://doi.org/10.3390/geosciences10020053

Michael, Y., Helman, D., Glickman, O., Gabay, D., Brenner, S., & Lensky, I. M. (2021). Forecasting fire risk with machine learning and dynamic information derived from satellite vegetation index time-series. *Science of the Total Environment*, *764*, 142844. https://doi.org/10.1016/j.scitotenv.2020.142844

Vitolo, C., Di Giuseppe, F., Barnard, C., Coughlan, R., San-Miguel-Ayanz, J., Libertá, G., & Krzeminski, B. (2020). ERA5-based global meteorological wildfire danger maps. *Scientific Data*, *7*(1), 1-11. https://doi.org/10.1038/s41597-020-0554-z