Unveiling Hidden Patterns in Wildfire ignitions: An Explainable Machine Learning Approach for Peninsular Spain.

Understanding the drivers behind wildfires ignitions is essential for effective fire prevention and management, yet a large proportion of recorded wildfires across Mediterranean Europe remain classified with unknown causes. This Study develops a Machine Leaning framework method to infer ignition causes from environmental an anthropogenic variables in Peninsular Spain. The model leverages a comprehensive database of fire events from 1985 to 2021 compiled using IberFire burned area perimeters dataset and the national wildfire dataset EGIF, focusing on four ignition categories: Lightning, Negligence, Intentional and Restart of fires.

We applied a Random Forest Classifier with data balancing via SMOTE to address class imbalance and used SHAP (Shapley Additive Explanations) to interpret feature contributions. The model achieved an overall global accuracy and F1-score of 85%. Intentional ignitions were classified with the highest precision (F1-score = 0.91), while natural and restart of fire ignitions showed lower accuracy (F1-Score = 0.52, 0.09). Key predictive features included climatic variables and human variables, reflecting the distinctive ignition dynamics between natural and human-caused fires.

Our results highlight that explainable machine learning can support the identification of ignition causes in large fire databases with high rates of unknown origin, thereby improving the evidence base for targeted fire prevention strategies. This approach offers valuable insights for fire managers and policy-makers aiming to strengthen risk mitigation in fire-prone Mediterranean landscapes.