



Proceeding Paper

Assessing Wildfire Susceptibility Utilizing Geospatial Datasets and Machine Learning Methods [†]

Suresh Babu KV ^{1,*} , Apostolos Sarris ¹ and Dimitris Stagonas ²

¹ Digital Humanities GeoInformatics Lab, Sylvia Ioannou Digital Humanities Chair, Department of History and Archaeology, University of Cyprus, Nicosia 1095, Cyprus; sarris.apostolos@ucy.ac.cy

² Department of Civil and Environmental Engineering, University of Cyprus, Nicosia 2109, Cyprus; stagonas.dimitris@ucy.ac.cy

* Correspondence: sureshbabu.iiith@gmail.com

[†] Presented at the 2nd International Workshop on Extreme Wildfire Events (X-Fire 2026), Prague, Czech Republic, 23–25 June 2026.

Abstract

This study investigates wildfire susceptibility in Cyprus, defined as the spatial probability of fire occurrence influenced by environmental factors and historical fire incidents. We integrated topographic variables, indicators of anthropogenic ignition (such as population density and proximity to infrastructure), and spectral vegetation indices (including NDVI, LAI, NDWI, and NDMI) using eight machine learning models trained on a historical wildfire dataset. Advanced ensemble models (XGBoost, Random Forest, and LightGBM) showed superior performance, achieving an accuracy of 0.82 and an AUC of 0.87. These non-linear classifiers successfully capture complex human–environmental interactions, delivering a high-fidelity spatial tool for targeted wildfire mitigation and forest management.

Keywords: wildfire; Random Forest; XGBoost; susceptibility; machine learning

1. Introduction

Wildfires are one of the most significant environmental threats in the Mediterranean basin, especially in Cyprus [1]. Recently, Cyprus has consistently ranked high among European countries regarding the area burned relative to its land mass [2]. The combination of a warming climate, land abandonment, and increasing human presence in the wildland–urban interface (WUI) has resulted in a high-risk environment [3]. Current fire management practices often depend on single-factor indices, which do not adequately capture the complex, non-linear interactions between environmental and human-caused factors [4].

This study tackles the challenges of wildfire management by developing a high-fidelity wildfire susceptibility map for Cyprus using a robust machine learning framework. We integrate historical wildfire data with various geospatial drivers and apply eight distinct machine learning algorithms. These range from traditional linear classifiers, such as logistic regression, to advanced ensemble algorithms such as Random Forest, XGBoost, and LightGBM. This approach establishes a clear performance hierarchy for regional risk modeling. The primary aim is to define wildfire susceptibility as a precise spatial probability of fire occurrence, explicitly illustrating how human infrastructure interacts with landscape vulnerability.



Academic Editor: Theodore M. Giannaros

Published: 2 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

2. Materials and Methods

In this study, wildfire susceptibility is defined as the localized spatial probability of fire occurrence, which includes both the initial eruption and establishment of fires. This probability is influenced by both static biophysical conditions and anthropogenic factors. Topographic parameters such as elevation, slope, aspect, plan curvature, and geomorphic landform classes are derived from the EU Digital Elevation Model. These factors illustrate how terrain influences the likelihood of fire spread. Additionally, anthropogenic factors including population density and the Euclidean distances to various features, such as roads, picnic areas, built-up regions, transmitter stations, and power lines, are also considered in this study. The input vector layers are sourced from the Cyprus Land and Survey Department. These characteristics help evaluate the likelihood of anthropogenic ignitions and identify critical infrastructure associated with ignition sources.

Moreover, vegetation and fuel predictors consist of various categorical vegetation types and indices that assess vegetation health and moisture levels. These indices are obtained from Landsat 8 and 9 satellite datasets [5], including the normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), normalized difference water index (NDWI), and normalized difference moisture index (NDMI). Together, these factors indicate the availability of fuel and its potential flammability. The predictors were compiled into static geospatial rasters and sampled to create a binary classification dataset that differentiated between fire and non-fire locations. Historical fire points from the VIIRS dataset, spanning the years 2015 to 2024, were acquired from the NASA FIRMS website [6] and utilized as the target variable.

A rigorous comparative modeling framework was established, employing eight distinct supervised machine learning classifiers that represented a variety of mathematical architectures. The ensemble and gradient-boosting models include Random Forest, extreme gradient boosting, and light gradient boosting machine. Additionally, statistical, instance-based, and tree-based classifiers are represented by support vector machines with a radial basis function kernel, logistic regression, K-nearest neighbors, and standard decision trees. Furthermore, deep learning capabilities were assessed using a multi-layer perceptron artificial neural network, which was optimized through backpropagation. The unified binary dataset was randomly divided into a 70% training subset for model calibration and hyperparameter tuning and a 30% independent testing subset for validation. Hyperparameters for each of the eight algorithms were fine-tuned using a grid search cross-validation strategy conducted solely on the training partition.

3. Results and Discussion

The predictive performance, discriminative capacity, and generalization ability of the static-feature machine learning models were thoroughly evaluated using two primary benchmark metrics derived from the confusion matrix of the independent testing dataset. The overall classification accuracy was calculated to determine the global proportion of correctly classified pixels, including both the presence and absence of fire, in relation to the total size of the validation sample. the area under the receiver operating characteristic curve (ROC-AUC) was employed to evaluate each model's ability to distinguish between high-susceptibility and low-susceptibility zones across all classification thresholds.

The comparative performance analysis of eight machine learning models reveals a distinct architectural hierarchy for mapping wildfire susceptibility across Cyprus, as illustrated in Table 1.

Table 1. Overall accuracy and AUC values for machine learning models.

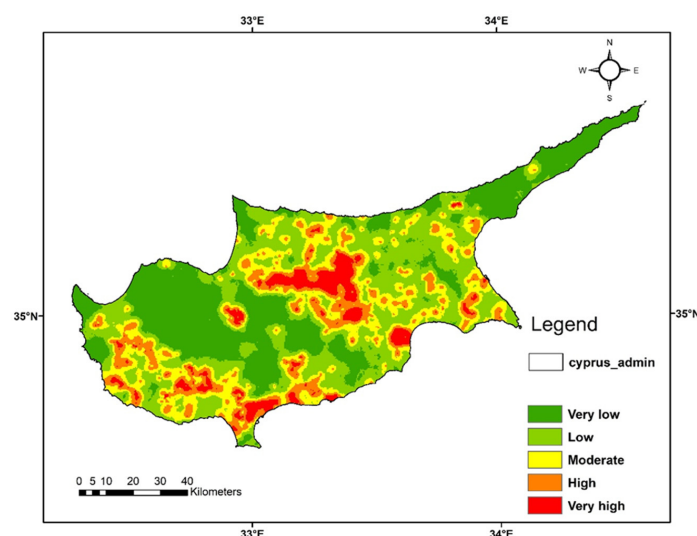
ML Model	Overall Accuracy	ROC-AUC
RF	0.82	0.89
ANN	0.77	0.85
KNN	0.80	0.86
DT	0.76	0.78
XGBoost	0.82	0.89
LGBM	0.82	0.89
SVM	0.80	0.85
LR	0.74	0.78

Advanced ensemble and gradient-boosting frameworks specifically XGBoost, LGBM, and RF, demonstrated superior predictive capabilities, each achieving an overall classification accuracy of 82% and an ROC-AUC score of 0.89. This top-tier performance emphasizes the robustness of tree-based ensembles in handling multi-dimensional geospatial data, as their internal boosting and bagging mechanisms effectively minimize variance and capture complex, non-linear interactions among topographic variables, human ignition proxies, and dynamic vegetative indices without overfitting.

In contrast, intermediate classifiers such as K-nearest neighbors (KNN) and support vector machines (SVM) exhibited moderate predictive performance, achieving an accuracy of 80% and AUC scores of 0.86 and 0.85, respectively. The artificial neural network (ANN) followed closely behind, with a 77% accuracy and a 0.85 AUC. This indicates that instance-based and deep learning architectures lack the precise feature-weighting flexibility needed to handle sharp geospatial transitions, such as localized wildland–urban interface drop-offs.

The lowest predictive capabilities were recorded by simpler baseline models, namely decision trees (DT) and logistic regression (LR), which yielded accuracies of 76% and 74%, along with identical AUC scores of 0.78. This performance drop highlights the limitations of linear boundaries and rigid, orthogonal decision structures in effectively capturing the dynamics of Mediterranean wildfires. Complex, exponential environmental thresholds fundamentally influence these dynamics, rather than linear gradients.

To leverage the collective predictive power of these superior models, the final wildfire susceptibility map was generated using a soft-voting ensemble approach that averaged the predicted probabilities of RF, XGBoost, and LGBM [7,8], which was subsequently categorized into five distinct fire risk levels: very low (<0.2), low (0.2–0.4), moderate (0.4–0.6), high (0.6–0.8), and very high (>0.8) (Figure 1).

**Figure 1.** Wildfire susceptibility map for the Cyprus.

Ultimately, these findings demonstrate the clear advantages of advanced tree-based ensemble frameworks over traditional parametric methods. They serve as a high-fidelity spatial decision-support tool, crucial for targeted forest management and effective wildfire mitigation in high-risk areas.

4. Conclusions

This study develops a high-resolution wildfire susceptibility map for Cyprus by incorporating various topographic, anthropogenic, and spectral vegetation factors within a sophisticated machine learning model. A systematic comparative evaluation of eight classifiers revealed that tree-based ensemble and gradient-boosting architectures, specifically XGBoost, Random Forest, and LightGBM, exhibited a superior performance hierarchy, with each achieving an overall accuracy of 82% and an AUC of 0.89.

By utilizing a soft-voting ensemble approach to average the predictive probabilities from these top-performing models, the final susceptibility map successfully categorizes the regional landscape into five standardized risk levels, ranging from very low to very high risk. This methodology effectively captures the intricate, non-linear interactions between humans and the environment, particularly emphasizing high-risk areas where critical human infrastructure directly interacts with vulnerable, drought-stressed fuels.

In conclusion, this approach highlights the importance of moving away from traditional single-factor indices and embracing data-driven, soft-voting ensembles. This map is essential for strategic resource allocation, zoning in wildland–urban interfaces, and targeted wildfire mitigation in high-risk Mediterranean ecosystems.

5. Future Work

Future work will focus on transforming this static susceptibility framework into a dynamic real-time forecasting system by integrating hourly meteorological data streams, such as live wind velocity vectors, fluctuations in relative humidity, and anomalies in ambient temperature, directly into the soft-voting ensemble architecture. Additionally, incorporating physics-based algorithms for fire spread enables the model to extend beyond initial ignition probabilities, actively simulating how a fire can grow, spread, and breach containment barriers across various terrain features.

Author Contributions: Conceptualization, methodology, software, validation, S.B.K., A.S. and D.S.; writing, original draft preparation, S.B.K.; review and editing, A.S. and D.S.; supervision, A.S. and D.S. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 101034403 (ONISILOS MSCA Co-fund project). The first author is grateful to ONISILOS and the University of Cyprus for providing the opportunity for and support in conducting this research.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. Satellite datasets can be downloaded freely from the NASA Earth Explorer.

Acknowledgments: The authors gratefully acknowledge the National Aeronautics and Space Administration (NASA) Earthdata platform and the Fire Information for Resource Management System (FIRMS) for providing free access to the visible infrared imaging radiometer suite (VIIRS) active fire data utilized in this study. The authors also express their appreciation to the U.S. Geological Survey (USGS) for providing the Landsat 8 and Landsat 9 Operational Land Imager (OLI) spectral datasets

essential for the vegetative and fuel moisture analysis. The authors also acknowledge the Lands and Surveys Department of Cyprus for providing the datasets.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Elia, M.; D'Este, M.; Ascoli, D.; Giannico, V.; Spano, G.; Ganga, A.; Sanesi, G. Estimating the probability of wildfire occurrence in Mediterranean landscapes using Artificial Neural Networks. *Environ. Impact Assess. Rev.* **2020**, *85*, 106474. [CrossRef]
2. Department of Forests, Cyprus. Available online: https://www.moa.gov.cy/moa/fd/fd.nsf/fd93_en/fd93_en?OpenDocument (accessed on 15 April 2026).
3. Georgiou, K.E. Forest Fires in Cyprus. *Maldives Natl. J. Res.* **2023**, *11*, 7–17. [CrossRef]
4. Wang, C.; Liu, H.; Xu, Y.; Zhang, F. A forest fire prediction framework based on multiple machine learning models. *Forests* **2025**, *16*, 329. [CrossRef]
5. Earth Explorer. Available online: <https://earthexplorer.usgs.gov/> (accessed on 13 March 2026).
6. NASA FIRMS. Available online: <https://firms.modaps.eosdis.nasa.gov> (accessed on 21 March 2026).
7. Rezaei Barzani, A.; Pahlavani, P.; Ghorbanzadeh, O. Ensembling of decision trees, KNN, and logistic regression with soft-voting method for wildfire susceptibility mapping. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *10*, 647–652. [CrossRef]
8. Karlos, S.; Kostopoulos, G.; Kotsiantis, S. A soft-voting ensemble-based co-training scheme using static selection for binary classification problems. *Algorithms* **2020**, *13*, 26.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.