Wildfire Susceptibility Mapping in Karabuk Province, Türkiye Using Machine Learning Algorithms

Sohaib K M Abujayyab[1], İsmail Rakip Karaş[2], Hacer Kübra Sevinç[3],
Şüheda Semih AÇMALI[4], Melih Yılmaz[5], Ahmet Serdar Dönmez[6], Osama Afana[7]

[1] Dept. of Geography, Karabuk University, Karabuk, Türkiye – sjayyab@karabuk.edu.tr
[2,5,6,7] Dept. of Computer Engineering, Karabuk University, Karabuk, Türkiye,
[2]ismail.karas@karabuk.edu.tr, [5]2010205069@ogrenci.karabuk.edu.tr,
[6]2010205024@ogrenci.karabuk.edu.tr, [7]2110213578@ogrenci.karabuk.edu.tr
[3,4] Vocational School of Information Technologies, Karabuk University, Karabuk, Türkiye,
[3]kubrasevinc@karabuk.edu.tr, [4]semihacmali@karabuk.edu.tr

Keywords: Wildfire, Susceptibility Mapping, Machine Learning, Karabuk, Türkiye

Abstract

Wildfires are among the most devastating natural hazards, increasingly intensified by climate change and anthropogenic pressures. Accurate susceptibility mapping is essential for disaster preparedness, risk mitigation, and sustainable land management. This study investigates the performance of five boosting-based machine learning algorithms—Gradient Boosting Machine (GBM), Adaptive Boosting (AdaBoost), Light Gradient Boosting Machine (LightGBM), Categorical Boosting (CatBoost), and Extreme Gradient Boosting (XGBoost)—in wildfire susceptibility assessment. Thirteen conditioning factors representing topographic, vegetational, climatic, and anthropogenic drivers were integrated into the models after preprocessing and multicollinearity checks.
The results show that XGBoost, CatBoost, and LightGBM significantly outperformed GBM and AdaBoost, with XGBoost achieving the highest predictive accuracy (94.5%), AUC (0.939), and Kappa index (0.890). Feature importance analysis revealed that land cover, NDVI and temperature were the most significant factors, followed by slope, wind speed and proximity to human settlements. The susceptibility maps produced by the best-performing models provided spatially consistent and interpretable patterns, successfully delineating high-risk areas.
This research confirms the effectiveness of advanced ensemble learning techniques, particularly XGBoost, in improving the accuracy and interpretability of wildfire susceptibility mapping. The findings provide actionable insights for forest management, land-use planning, and the development of early warning systems, contributing to more resilient strategies against escalating wildfire threats in a changing climate.

1. Introduction

Wildfires represent one of the most destructive natural hazards, with profound ecological, economic, and social impacts. Their frequency and severity have been increasing in recent years due to climate change, land-use pressures, and anthropogenic activities. These dynamics underline the urgent need for accurate wildfire susceptibility assessment to support disaster preparedness, risk mitigation, and sustainable land management.

Conventional approaches to wildfire risk analysis—based largely on field surveys and statistical methods—often suffer from limited spatial resolution and predictive power. With the advancement of geospatial technologies and machine learning, it has become possible to integrate diverse datasets such as topography, vegetation, climate, and human activities into more robust predictive frameworks. Nevertheless, challenges remain regarding the optimal selection of variables, model accuracy, and practical applicability in real-world decision-making.

This study proposes a systematic framework for wildfire susceptibility mapping by employing geospatial datasets and machine learning algorithms. The primary objectives are to identify key environmental and anthropogenic drivers of wildfire occurrence, develop predictive models with high reliability, and produce high-resolution susceptibility maps. The outcomes are expected to provide actionable insights for policymakers and emergency management authorities in prioritizing monitoring efforts and implementing effective prevention strategies.

2. Related Work

Wildfire susceptibility mapping has garnered increased attention with the integration of machine learning algorithms, particularly boosting methods such as XGBoost, AdaBoost, LightGBM, and Gradient Boosting Machine (GBM). These algorithms are particularly well-suited for synthesizing diverse datasets encompassing topographic, vegetation, climatic, and anthropogenic factors, thus enhancing the predictive accuracy of wildfire risk assessments.

Yue et al. provide a comprehensive analysis of wildfire susceptibility in Guilin, China, employing XGBoost to combine multiple data sources and factor inputs. Their study reports an area under the curve (AUC) of 0.927, underscoring the model's effectiveness when compared to traditional methods. This exemplifies the capability of boosting algorithms to manage complex variables and interactions that influence wildfire occurrence(Yue et al., 2023).

He et al. extend this discussion by highlighting ensemble methods that include boosting techniques like GBM and AdaBoost. Their research assesses wildfire susceptibility in Southeast Asia and emphasizes the improved performance of boosting algorithms in modeling wildfire risks compared to traditional approaches(He, Jiang, Wang, & Liu, 2021).

Additionally, Masoudvaziri et al. demonstrate that XGBoost outperformed other machine learning models in identifying key factors related to wildfire size, integrating geophysical and anthropogenic features. This research indicates that well-

https://doi.org/10.5194/isprs-archives-XLVIII-4-W18-2025-1-2026

parameterized machine learning models can improve risk assessment by considering both topographic and socio-economic influences(Masoudvaziri, Ganguly, Mukherjee, Sun, & Assessment, 2022).

Kouassi et al. examine the impact of accessibility to urban markets and other anthropogenic factors on wildfire occurrence in Côte d'Ivoire. Their work identifies a significant correlation between these factors and wildfire frequency, showcasing the importance of integrating human variables alongside environmental data in susceptibility modeling(Kouassi, Wandan, & Mbow, 2022).

Moreover, Yang et al. employ Maxent and GIS for wildfire risk assessment in Hunan Province, China, integrating multiple environmental variables to create a risk zoning model. Their findings highlight the successful combination of machine learning algorithms with geographic information systems (GIS) to enhance understanding and forecasting of wildfire risks across varied landscapes(Yang, Jin, & Zhou, 2021).

Research by Ghorbanzadeh et al. emphasizes the role of AI and machine learning in spatial prediction of wildfire susceptibility. They advocate for methodologies that leverage field survey data to inform predictive models, reiterating the promise of incorporating decision trees and ensemble methods inherent in boosting algorithms to refine predictions(Ghorbanzadeh et al., 2019).

In summary, boosting algorithms such as XGBoost and AdaBoost play a crucial role in wildfire susceptibility mapping by effectively integrating topographic, vegetation, climatic, and anthropogenic factors. Studies showcasing their success underscore the essential role of data integration and advanced modeling techniques in developing robust wildfire risk predictions.

3. Methodology

3.1 Data Collection and Preprocessing

To construct a comprehensive wildfire susceptibility model, thirteen conditioning factors were selected based on their relevance in previous wildfire susceptibility and hazard studies. These variables were grouped into four categories: topographic, vegetation, climatic, and anthropogenic factors.

All variables were standardized to a 30 m spatial resolution and normalized using min–max scaling to ensure comparability. The Variance Inflation Factor (VIF) was applied to detect and eliminate multicollinearity among predictors before model training.

3.1.1 Topographic Factors: Elevation, higher altitudes influence microclimatic conditions and vegetation type, which directly affect fire occurrence(Chuvieco, Giglio, & Justice, 2008). Slope, steeper slopes accelerate fire spread by preheating upslope fuels(Jaiswal, Mukherjee, Raju, Saxena, & geoinformation, 2002). Aspect, South-facing slopes (in the Northern Hemisphere) receive more solar radiation, resulting in drier fuels(Vigna, Besana, Comino, & Pezzoli, 2021).

3.1.2 Vegetatin Factors: Land Use/Land Cover (LULC), different land cover classes (forest, agricultural land, grassland) exhibit varying levels of fire susceptibility(Oliveira, Pereira, San-Miguel-Ayanz, & Lourenço, 2014). Normalized Difference Vegetation Index (NDVI): NDVI indicates vegetation density and health; lower NDVI values often correspond to higher fire risk due to fuel scarcity or degradation(Chuvieco & Congalton, 1989). Fuel Type / Fuel Load, the type and amount of combustible material are critical determinants of fire ignition and spread(Chuvieco & Kasischke, 2007).

3.1.3 Climatic Factors: Precipitation, reduced rainfall lowers soil and vegetation moisture, increasing fire susceptibility(Flannigan, Stocks, Turetsky, & Wotton, 2009). Temperature, higher mean annual temperatures are strongly correlated with reduced fuel moisture and increased fire activity(Moriondo et al., 2006). Relative Humidity, lower humidity accelerates fuel drying and promotes ignition(Rovithakis et al., 2022).

3.1.4 Anthropogenic Factors: Distance to Roads, most human-induced ignitions occur near road networks(Martínez, Vega-Garcia, & Chuvieco, 2009). Distance to Settlements, proximity to populated areas increases fire frequency due to human activities(Catry, Rego, Bação, & Moreira, 2009). Distance to Rivers/Water Bodies, river proximity influences vegetation density and fuel moisture, indirectly affecting fire risk(Trucchia et al., 2023). Population Density / Human Pressure, human presence and land-use intensity increase ignition likelihood, particularly in rural–urban interface zones(Syphard et al., 2007).

3.2 Machine Learning Algorithms

In this study, five widely used ensemble learning algorithms based on the boosting approach were implemented to model wildfire susceptibility: Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Categorical Boosting (CatBoost), and Adaptive Boosting (AdaBoost).

3.2.1 Gradient Boosting Machine (GBM): GBM constructs predictive models by sequentially training decision trees, where each subsequent tree attempts to correct the errors of the previous one. This greedy stage-wise optimization minimizes prediction residuals and improves performance(Friedman, 2001).

3.2.2 Extreme Gradient Boosting (XGBoost): XGBoost extends GBM by incorporating regularization techniques (L1 and L2), which reduce overfitting, and by enabling parallel and distributed computing. Its efficiency and scalability have made it a benchmark algorithm in machine learning competitions(Chen & Guestrin, 2016).

3.2.3 Light Gradient Boosting Machine (LightGBM): LightGBM, developed by Microsoft, employs histogram-based learning and a leaf-wise growth strategy, which significantly reduces computational cost. It is particularly efficient for large-scale, high-dimensional datasets while maintaining high predictive accuracy (Ke et al., 2017).

3.2.4 Categorical Boosting (CatBoost): CatBoost, introduced by Yandex, is designed to handle categorical variables effectively without the need for extensive preprocessing such as one-hot encoding. Its symmetric tree structure provides stable and fast training, making it advantageous for heterogeneous datasets (Dorogush, Ershov, & Gulin, 2018).

https://doi.org/10.5194/isprs-archives-XLVIII-4-W18-2025-1-2026

3.2.5 Adaptive Boosting (AdaBoost): AdaBoost is one of the earliest boosting algorithms, which assigns higher weights to misclassified instances at each iteration, thereby focusing the model on harder-to-classify samples. It has proven effective in binary classification and remains a strong baseline for ensemble methods(Freund, Schapire, & sciences, 1997).

4. Results and Discussion

4.1 Performance Comparison

The predictive performance of five boosting-based machine learning algorithms was evaluated using three performance indicators: Overall Accuracy, Area Under the Curve (AUC), and Kappa Index on both training and testing datasets (Table 1).

Metric/ Model	XGBoost	CatBoost	Light GBM	GBM	Ada Boost
ACC (Train)	99.5	99.4	99.8	89.2	82.1
ACC (Test)	94.5	94.5	94.2	89.1	81.0
AUC (Train)	0.999	0.999	0.999	0.915	0.830
AUC (Test)	0.939	0.939	0.937	0.889	0.775
Kappa (Train)	0.990	0.988	0.997	0.784	0.642
Kappa (Test)	0.890	0.890	0.885	0.782	0.620

Table 1: Performance comparison of applied machine learning models

The results clearly demonstrate that XGBoost, CatBoost, and LightGBM outperformed the other models across all evaluation metrics. In particular, XGBoost achieved the highest generalization ability with a testing accuracy of 94.5%, AUC of 0.939, and Kappa index of 0.890. LightGBM achieved the highest training accuracy (99.8%), but showed slightly lower performance on the testing set, indicating a mild tendency toward overfitting. CatBoost exhibited a performance profile very similar to XGBoost, confirming its robustness in handling categorical and heterogeneous data. The Receiver Operating Characteristic (ROC) curves provided an additional performance comparison among the models (Figure 1).

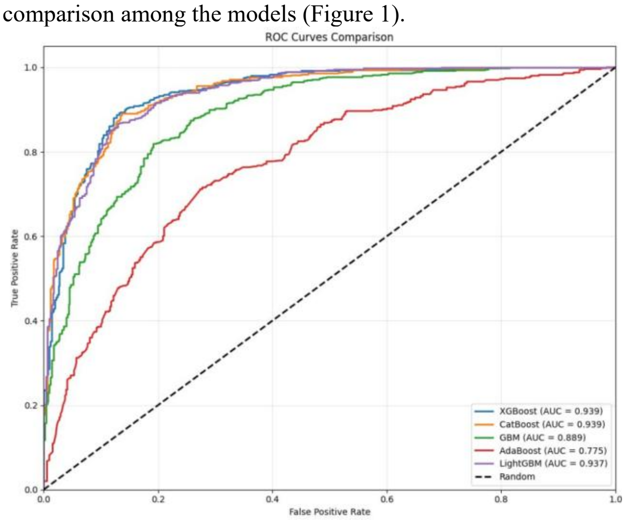

Figure 1: ROC curve comparison of XGBoost, CatBoost, LightGBM, GBM, and AdaBoost models

On the other hand, GBM and AdaBoost exhibited lower predictive capabilities, with test accuracies of 89.1% and 81.0% respectively, and significantly lower Kappa indices. These results confirm that newer boosting algorithms (XGBoost, LightGBM, CatBoost) offer substantial improvements in stability, generalization, and handling complex non-linear interactions compared to earlier ensemble methods.

4.2 Feature Importance Analysis

Feature importance derived from the XGBoost model indicated that land cover, NDVI, and temperature were the most influential variables in wildfire susceptibility prediction (Figure 2). Specifically, land cover emerged as the dominant factor, accounting for more than 25% of the model's predictive power. Climatic factors such as temperature and wind speed, along with vegetation-related indices (NDVI, LST), also played a critical role. Topographic features such as slope, aspect, and elevation, together with anthropogenic variables (distance to residential areas), showed moderate but significant contributions. These findings are consistent with previous studies highlighting the combined impact of environmental and human-induced factors on wildfire occurrence.

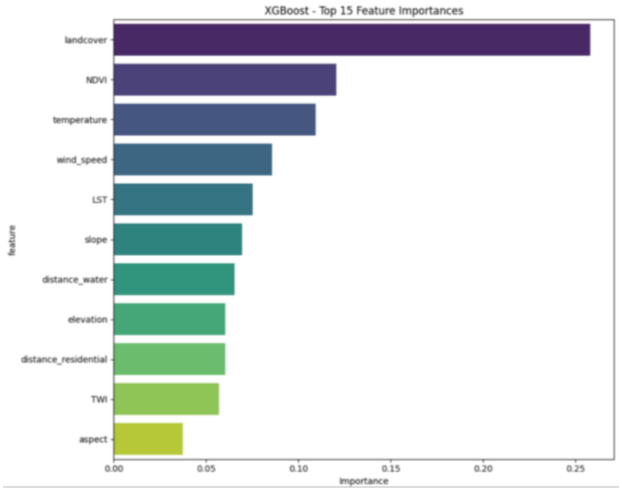

Figure 2: Feature importance values of the XGBoost model

4.3 Susceptibility Mapping

The spatial distribution of wildfire susceptibility, as predicted by the five algorithms, is shown in Figure 3. The maps illustrate that XGBoost, CatBoost, and LightGBM provide clearer and more spatially consistent susceptibility patterns compared to GBM and AdaBoost. In particular, XGBoost produced well-differentiated zones with high-risk areas predominantly concentrated in regions with dense vegetation cover and proximity to residential and water sources.

CatBoost showed nearly identical spatial patterns, while LightGBM tended to slightly overestimate susceptibility in fragmented land cover areas. In contrast, GBM and AdaBoost produced noisier maps with lower discrimination between high and low susceptibility zones, reflecting their lower predictive performance.

https://doi.org/10.5194/isprs-archives-XLVIII-4-W18-2025-1-2026

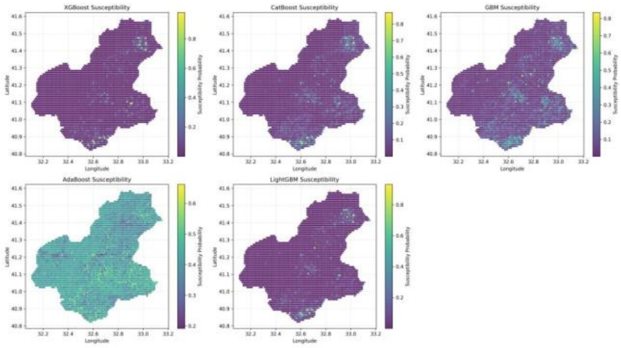

Figure 3: Wildfire susceptibility maps generated by XGBoost, CatBoost, GBM, AdaBoost, and LightGBM models

The combination of high predictive performance and spatial interpretability confirms that boosting-based machine learning approaches are highly effective for wildfire susceptibility assessment. Among them, XGBoost stands out as the most reliable tool, offering both robust generalization and meaningful feature contributions. These results can directly support forest management strategies, land-use planning, and early warning systems, contributing to the mitigation of wildfire risks in vulnerable landscapes.

5. Conclusion

This study demonstrated the effectiveness of boosting-based machine learning algorithms in wildfire susceptibility mapping. By incorporating 13 topographic, vegetation, climatic, and anthropogenic factors, five ensemble models (GBM, AdaBoost, LightGBM, CatBoost, and XGBoost) were trained and evaluated. The results clearly revealed that XGBoost, CatBoost, and LightGBM outperformed the traditional GBM and AdaBoost approaches, with XGBoost emerging as the most reliable algorithm, achieving the highest predictive accuracy and robust generalization performance.

The feature importance analysis highlighted that land cover, NDVI, and temperature were the most influential predictors of wildfire susceptibility, supported by secondary contributions from wind speed, slope, and anthropogenic proximity factors. These findings emphasize the necessity of considering both environmental and human-related drivers in wildfire risk modeling.

The susceptibility maps generated from the best-performing models provided spatially consistent and interpretable patterns, identifying high-risk areas that require urgent management attention. Such outputs can be directly applied to forest management, land-use planning, and the development of early warning systems, supporting proactive strategies to mitigate wildfire hazards.

In conclusion, the integration of advanced ensemble learning methods, particularly XGBoost, into wildfire risk assessment frameworks offers a powerful tool for enhancing decision-making processes and improving resilience against increasing wildfire threats under changing climatic conditions.

References

Catry, F. X., Rego, F. C., Bação, F. L., & Moreira, F. J. I. J. o. W. F. (2009). Modeling and mapping wildfire ignition risk in Portugal. 18(8), 921-931.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. Paper presented at the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.

Chuvieco, E., & Congalton, R. G. J. R. s. o. E. (1989). Application of remote sensing and geographic information systems to forest fire hazard mapping. 29(2), 147-159.

Chuvieco, E., Giglio, L., & Justice, C. J. G. c. b. (2008). Global characterization of fire activity: toward defining fire regimes from Earth observation data. 14(7), 1488-1502.

Chuvieco, E., & Kasischke, E. S. J. J. o. G. R. B. (2007). Remote sensing information for fire management and fire effects assessment. 112(G1).

Dorogush, A. V., Ershov, V., & Gulin, A. J. a. p. a. (2018). CatBoost: gradient boosting with categorical features support.

Flannigan, M., Stocks, B., Turetsky, M., & Wotton, M. J. G. c. b. (2009). Impacts of climate change on fire activity and fire management in the circumboreal forest. 15(3), 549-560.

Freund, Y., Schapire, R. E. J. J. o. c., & sciences, s. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. 55(1), 119-139.

Friedman, J. H. J. A. o. s. (2001). Greedy function approximation: a gradient boosting machine. 1189-1232.

Ghorbanzadeh, O., Valizadeh Kamran, K., Blaschke, T., Aryal, J., Naboureh, A., Einali, J., & Bian, J. J. F. (2019). Spatial prediction of wildfire susceptibility using field survey GPS data and machine learning approaches. 2(3), 43.

He, Q., Jiang, Z., Wang, M., & Liu, K. J. R. S. (2021). Landslide and wildfire susceptibility assessment in Southeast Asia using ensemble machine learning methods. 13(8), 1572.

Jaiswal, R. K., Mukherjee, S., Raju, K. D., Saxena, R. J. I. j. o. a. e. o., & geoinformation. (2002). Forest fire risk zone mapping from satellite imagery and GIS. 4(1), 1-10.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., . . . Liu, T.-Y. J. A. i. n. i. p. s. (2017). Lightgbm: A highly efficient gradient boosting decision tree. 30.

Kouassi, J. L., Wandan, N., & Mbow, C. J. A. J. o. E. (2022). Exploring spatio-temporal trends and environmental drivers of wildfire occurrence and impacts in Côte d'Ivoire, West Africa. 60(4), 1218-1236.

Martínez, J., Vega-Garcia, C., & Chuvieco, E. J. J. o. e. m. (2009). Human-caused wildfire risk rating for prevention planning in Spain. 90(2), 1241-1252.

Masoudvaziri, N., Ganguly, P., Mukherjee, S., Sun, K. J. S. E. R., & Assessment, R. (2022). Impact of geophysical and anthropogenic factors on wildfire size: a spatiotemporal data-driven risk assessment approach using statistical learning. 36(4), 1103-1129.

Moriondo, M., Good, P., Durao, R., Bindi, M., Giannakopoulos, C., & Corte-Real, J. J. C. r. (2006). Potential impact of climate change on fire risk in the Mediterranean area. 31(1), 85-95.

Oliveira, S., Pereira, J. M., San-Miguel-Ayanz, J., & Lourenço, L. J. A. G. (2014). Exploring the spatial patterns of fire density in Southern Europe using Geographically Weighted Regression. 51, 143-157.

Rovithakis, A., Grillakis, M. G., Seiradakis, K. D., Giannakopoulos, C., Karali, A., Field, R., . . . Voulgarakis, A. J. E. R. L. (2022). Future climate change impact on wildfire danger over the Mediterranean: the case of Greece. 17(4), 045022.

Syphard, A. D., Radeloff, V. C., Keeley, J. E., Hawbaker, T. J., Clayton, M. K., Stewart, S. I., & Hammer, R. B. J. E. a. (2007). Human influence on California fire regimes. 17(5), 1388-1402.

Trucchia, A., Meschi, G., Fiorucci, P., Provenzale, A., Tonini, M., & Pernice, U. J. I. J. o. W. F. (2023). Wildfire hazard mapping in the eastern Mediterranean landscape. 32(3), 417-434.

Vigna, I., Besana, A., Comino, E., & Pezzoli, A. J. S. (2021). Application of the socio-ecological system framework to forest fire risk management: A systematic literature review. 13(4), 2121.

Yang, X., Jin, X., & Zhou, Y. J. F. (2021). Wildfire risk assessment and zoning by integrating Maxent and GIS in Hunan province, China. 12(10), 1299.

Yue, W., Ren, C., Liang, Y., Liang, J., Lin, X., Yin, A., & Wei, Z. J. R. S. (2023). Assessment of wildfire susceptibility and wildfire threats to ecological environment and urban development based on GIS and multi-source data: A case study of Guilin, China. 15(10), 2659.