

Article

Not peer-reviewed version

Human Activities and Wildfires: The Impact of Forest Roads, Trails, and Forest Management on Wildfire Occurrence

[Youn Yeo-Chang](#)*, [Se-Eum Lee](#)*, [Soo-Jin Lee](#), [Hyo-Rin Kim](#)

Posted Date: 21 April 2026

doi: 10.20944/preprints202604.1451.v1

Keywords: forest fire; fire probability; forest age; road density; trail density; forest management; logistic regression; machine learning





Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Article

Human Activities and Wildfires: The Impact of Forest Roads, Trails, and Forest Management on Wildfire Occurrence

Youn Yeo-Chang ^{1,*} , Se-Eum Lee ^{2,*} , Soo-Jin Lee ³ and Hyo-Rin Kim ⁴

¹ Professor emeritus, Seoul National University, Chair Professor, aSSIST University; Director, Institute of Sustainable Social-Ecological Systems (ISSES)

² Researcher, ISSES

³ Senior Researcher, ISSES

⁴ Researcher, Indidlab

* Correspondence: youn@snu.ac.kr (Y.Y.-C.); 9_021@naver.com (S.-E.L.)

Abstract

The risk of wildfires is increasing due to high temperatures and dry weather conditions caused by climate change. Outbreaks and spread of wildfires are usually conditioned by weather, topography, and forest stand characteristics. In the Republic of Korea (hereafter ROK), most wildfires are caused by anthropogenic factors rather than by natural factors. However, the current forest fire forecasting system being operated in ROK does not account for anthropogenic factors. To analyze the impact of human factors, along with physical factors, on wildfire occurrence, a binary logistic regression model was constructed with data for the Gangwon and Gyeongbuk provinces from January 2022 to August 2025. The dependent variable was defined as the occurrence of a wildfire, while the independent variables comprised meteorological, seasonal, stand, and anthropogenic factors. To address multicollinearity, variables with high correlation coefficients were excluded from the independent variables, which were selected by three estimating approaches including logistic regression and two machine learning techniques (namely, Random Forest and XGBoost). With machine learning, the variables with high feature importance were identified. The explanatory power of the logistic regression analysis with independent variables selected by the machine learning models was about 1.3 times higher than the model using variables adjusted solely for multicollinearity. The results of logistic regression analysis revealed that weather and coniferous forests are the most important factors fostering wildfires, while the mean stand age was the most significant factor in hindering wildfires. Among the anthropogenic factors, forest road density acted as a suppressor of wildfire spread rather than a promoter of occurrence. Conversely, trail density tends to increase the risk of wildfire occurrence. Among forest management activities, artificial forests could boost forest fires, although this remains uncertain. These findings suggest that preventing wildfires requires a paradigm shift in forest resource management policies, including extending the rotation age of forests and the conversion of coniferous forests to broadleaf forests. Meanwhile, it also indicates the need to restrict the expansion of hiking trails and improve regulations regarding hiker access to prevent wildfires.

Keywords: forest fire; fire probability; forest age; road density; trail density; forest management; logistic regression; machine learning

1. Introduction

Global climate change causes extreme weather events such as droughts occurring frequently. These climatic conditions are as considered as primary causes for the increasing frequency and scale of wildfire damage [6,7,11]. Globally, the frequency of mega-fires is rising, and wildfire seasons are becoming prolonged, leading to severe ecological and economic losses. In the era of the climate

crisis, it is particularly concerning that wildfires release massive amounts of carbon dioxide into the atmosphere in a short period. This emitted carbon dioxide exacerbates climate change [8]. Boreal forest fires typically account for about 10% of global wildfire carbon dioxide emissions; however, in 2021, this figure surged to 23%, recording the highest proportion since 2000. In 2021, driven by global warming, abnormal moisture deficits led to extreme wildfires in North America and Eurasia [9]. Wildfires do not merely cause short-term damage; they trigger a severe positive feedback loop that accelerates global warming by releasing large quantities of carbon previously stored in forests [10].

In the ROK, the largest mega-fire in the nation's history occurred in March 2025, primarily affecting Gyeongbuk, Gyeongnam, and Ulsan, with the damaged area reaching approximately 104,000 hectares, equivalent to 1.64% of total forest area [12]. Furthermore, this wildfire recorded the highest number of human casualties in history, in addition to extensive forest damage [5]. Korea is constantly exposed to the risk of wildfires, which have become an everyday threat. Given that about 63% of the land is covered by forests, combined with dry spring weather, strong localized winds, and rugged topography, there is a high risk of fires spreading rapidly.

An examination of wildfire trends in Korea reveals that the interaction of climatic, structural, and social factors is making wildfires increasingly larger and more routine. According to statistics, an average of 450 to 500 wildfires occurs annually, destroying approximately 3,700 to 4,000 hectares of forests each year [22,24]. Notably, the success of past reforestation projects and continuous forest protection efforts have led to an increase in forest growing stock since the 1960s. This has resulted in fuel accumulation, acting as a structural cause that exacerbates the scale of wildfires [22,24].

Meteorological patterns, characterized by decreasing precipitation days and a sharp increase in dry weather warnings due to climate change, further aggravate the risk of mega-fires [23]. Changes in the timing and patterns of occurrence are also distinct. Wildfires, once concentrated in the spring, are expanding into May and winter season, while the increase in hikers due to the five-day workweek has led to more fires on Fridays and weekends. The decline and aging of rural populations make initial firefighting efforts increasingly difficult [22,24]. This implies that the ignition and spread of wildfires are closely linked to human activities and social structural changes, beyond simple climatic or topographical conditions.

Unlike natural causes such as lightning strikes, wildfires in Korea are predominantly caused by anthropogenic factors, such as accidental fires by hikers or the burning of agricultural waste [2]. The current National Forest Fire Danger Rating System operated by the National Institute of Forest Science issues fire risk ratings based on statistical analyses of weather, stand, and topographical factors in relation to wildfire occurrences (2000–2010) [2], but it does not incorporate anthropogenic factors. Therefore, it is necessary to account for the impact of human-induced factors on wildfire occurrence for improved forest fire forecasting systems and effective forest fire prevention plans.

Won et al. [13] analyzed variables affecting spring and autumn wildfires and found statistically significant correlations between fire probability and temperature, relative humidity, effective humidity, and wind speed. Ryu et al. [14] analyzed data of forest fire breakouts over the last 30 years and found out that wildfire risk periods have been extended due to climate change in ROK. Kwak et al. [15] showed that slope, elevation, aspect, distance to roads, and population density are significant explanatory factors for wildfire occurrence. Lee et al. [16] found out that wildfires are concentrated on south-facing slopes and socioeconomic development and fire probability are correlated. Kim et al. [17] reported that over a 30-year period (1991–2020), annual wildfire incidents have been increasing, with a worsening spatial unevenness as mega-fire damages are concentrated in the northeastern regions of Gangwon and Gyeongbuk.

However, there is a lack of empirical research that quantitatively and spatially analyzes the impact of anthropogenic factors on wildfires. Some previous studies (Hong et al. [3]) proposed a hypothesis that forest roads can exacerbate wildfire breakouts. Conversely, others suggest that forest roads hinder the spread of wildfires (Lee et al. [30]). A question of whether there is a relationship between forest

roads and wildfires or not needs to be clarified. Some argue that the ROK Government's policy of subsidizing "forest improvement tending" promotes forest fires (Park et al. [4]).

To better understand the factors associated with wildfire occurrence in ROK, this study aims to test the following three hypotheses:

1. Forest stand age and its species composition influence wildfire occurrence.
2. Artificial forest tending activities can promote wildfire occurrence.
3. Expansion of forest road networks and trail infrastructure (density and accessibility) increases wildfire risk.

2. Materials and Methods

2.1. Study Area and Research Flow

The spatial scope of this study was set to the Gangwon and Gyeongbuk provinces in the ROK (Figure 1). Both regions feature rugged terrain and have recorded frequent wildfires. The temporal scope was defined from January 2022 to August 2025. The research followed the procedure illustrated in Figure 2 below.

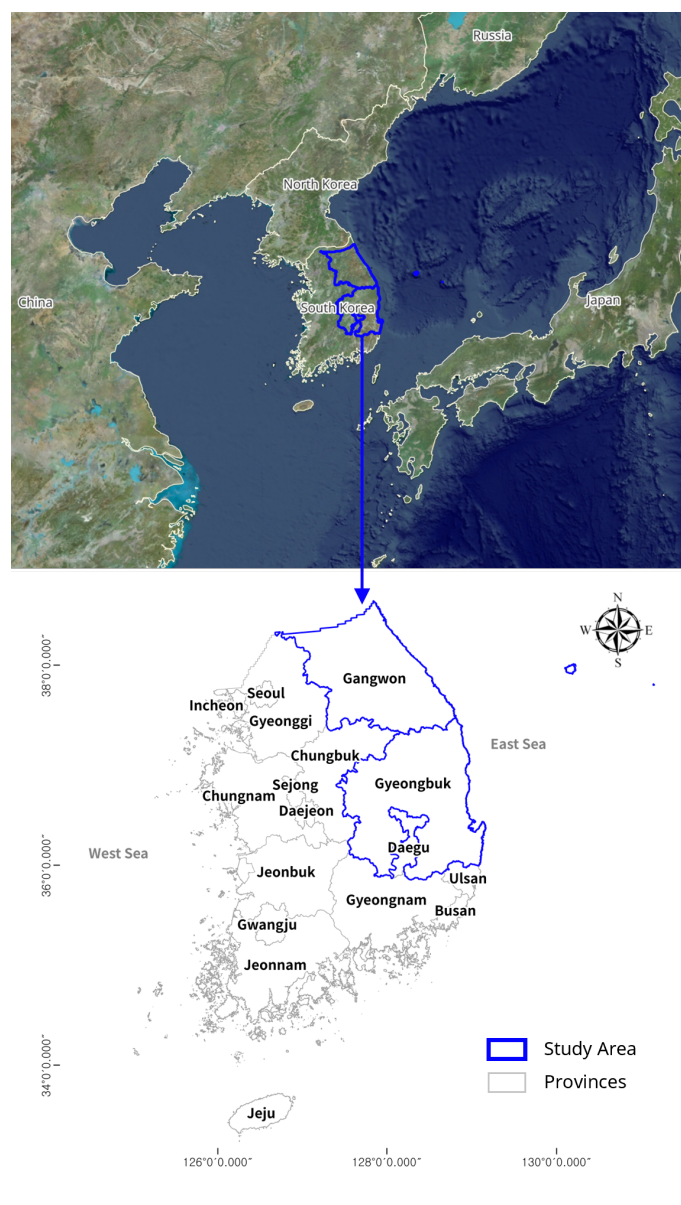


Figure 1. Study Area.

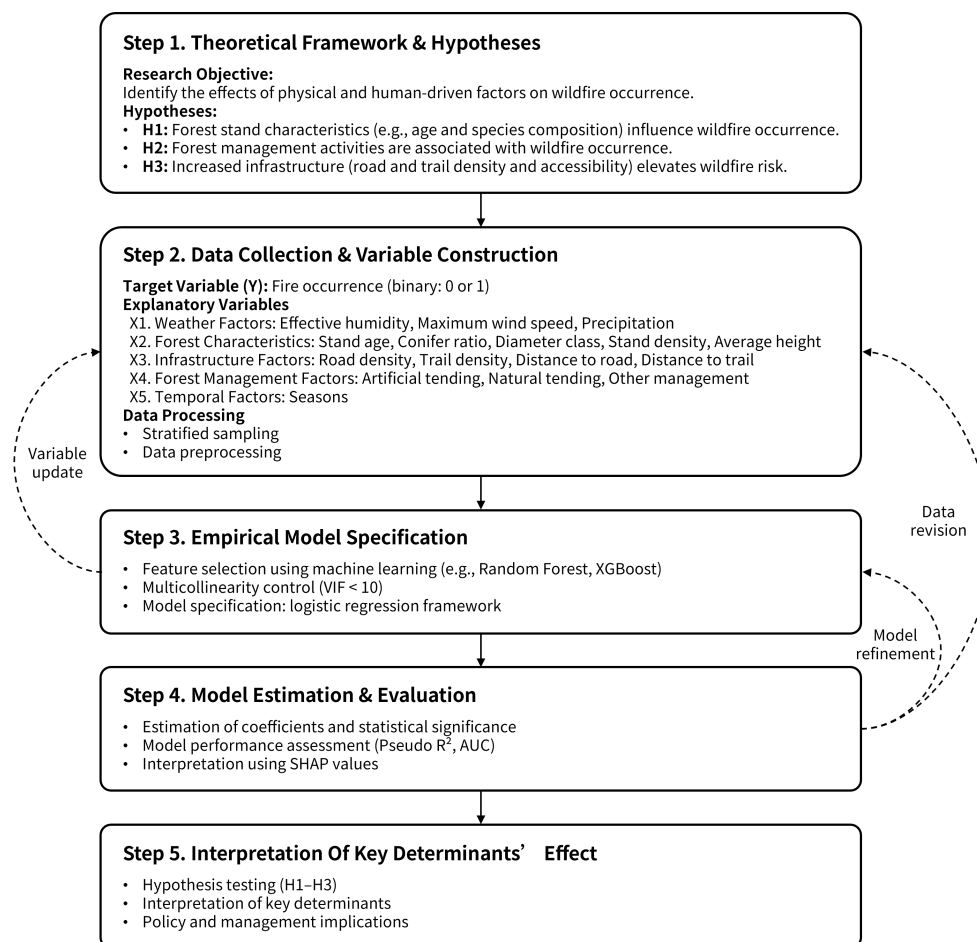


Figure 2. Research framework for analyzing the impact of physical and human-driven factors on forest fire occurrence.

Variables were selected and collected from relevant organizations, followed by a data refinement process to determine their impact on wildfire occurrence. In the data preprocessing stage, the study areas of Gangwon and Gyeongbuk were divided into 1 km × 1 km square grids to construct grid data. All spatial data preprocessing and variable mapping were performed using the open-source software QGIS 3.40.15.

The dependent variable (Target) was set as binary classification, defining grids with a wildfire occurrence during the period as 1, and those without as 0. For grids where wildfires occurred (Target=1), addresses provided by the Korea Forest Service's wildfire statistics were converted into latitude and longitude coordinates using Geocoder¹, a web-based geocoding site. The coordinates represent the centroid of the address parcel area. The geocoded location data were mapped as Point data to the grids to identify Target grids. Independent variables X_1 to X_5 were also mapped to these grids along with the dependent variable Y . Data collection and processing methods for these variables are detailed in Section 2.2. The constructed spatial information was then transformed into a dataset structured suitably for machine learning model training and binary logistic regression analysis.

First, data on wildfire occurrence locations were obtained through an information disclosure request to the Korea Forest Service. Initially, a total of 492 wildfire cases were recorded in the Gangwon and Gyeongbuk provinces from January 2022 to August 2025. However, after excluding cases where the exact address was difficult to identify and duplicate cases occurring within the same 1 km × 1 km grid, the final number of wildfire occurrence grids (Target=1) was established as 471.

¹ <https://geocoder.gimi9.com/>

Second, non-occurrence grids (Target=0) were extracted for machine learning and binary logistic regression analysis based on the refined data. Simple random sampling of non-occurrence grids could lead to spatial bias regarding regional meteorological and topographical characteristics. Therefore, to resolve the class imbalance caused by the difference in quantities between occurrence and non-occurrence grids, Region-based Stratified Random Sampling was employed, extracting 1,000 non-occurrence grids. The specific sampling process is as follows.

The study area was divided into three zones based on topographical and meteorological similarities: Yeongdong (Gangwon East Coast), Yeongseo (Gangwon Inland), and Gyeongbuk (Figure 3). The wildfire occurrences in each zone were examined. Out of 471 wildfires between January 2022 and August 2025, Gyeongbuk accounted for 266 (56.48%), Yeongseo for 139 (29.51%), and Yeongdong for 66 (14.01%). To ensure the model balanced by learning each region's unique environmental characteristics according to occurrence weights, the actual occurrence ratio was set as the extraction weight. The target of 1,000 non-occurrence grids was allocated accordingly: 566 from Gyeongbuk, 296 from Yeongseo, and 138 from Yeongdong were randomly selected. Through this process, a final analysis dataset of 1,471 grids was constructed, comprising 471 occurrence grids and 1,000 stratified non-occurrence grids (Table 1). Notably, for the Yeongdong region—characterized by unique weather conditions like the Yangganjipung (local strong winds)—138 non-occurrence grids (about double the 66 occurrence grids) were allocated to effectively train the model on the complex ignition risk factors of the area.

Modeling, feature selection, logistic regression analysis, and SHAP visualization were implemented using the pandas, scikit-learn, xgboost, statsmodels, and shap libraries in the Python programming language.

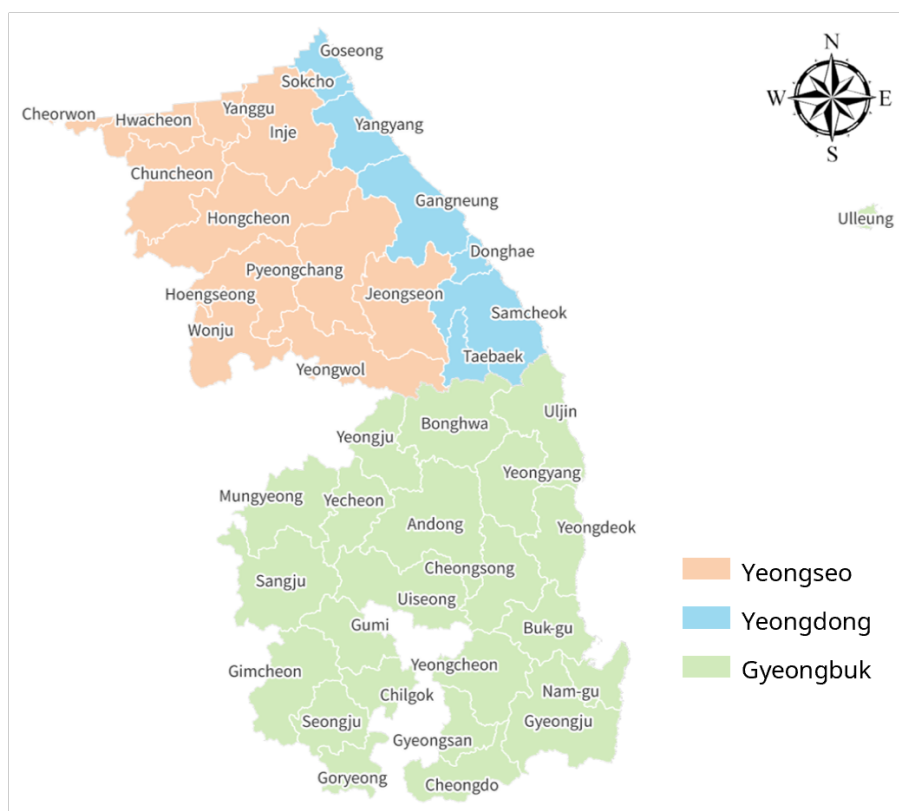


Figure 3. Regional division map of Gangwon-Yeongseo, Gangwon-Yeongdong, and Gyeongbuk.

Table 1. Composition of the final analysis dataset by region (Target=1 vs. Target=0).

Region	Target=1 (Occurrence)	Target=0 (Non-occurrence)	Total Grids
Gyeongbuk	266 (56.48%)	566	832
Yeongseo	139 (29.51%)	296	435
Yeongdong	66 (14.01%)	138	204
Total	471 (100%)	1,000	1,471

2.2. Data Collection and Processing Methods

To analyze the multidimensional factors affecting wildfire occurrence, independent variables were categorized into four groups based on raw data: X_1 (Weather), X_2 (Forest Characteristics), X_3 (Infrastructure), X_4 (Forest Management) and X_5 (Temporal Factors) (Table 2). Each independent variable was spatially joined to the grids. Missing values were replaced with 0 to complete preprocessing. Detailed collection and processing methods for each variable are described below.

Table 2. Definitions of dependent and independent variables for logistic regression analysis.

Types	Variables	Specifications
Weather Factors (X_1)	Effective Humidity	Daily effective humidity (%)
	Maximum Wind Speed	Daily maximum wind speed (m/s)
	Precipitation	Daily total precipitation (mm)
Forest Characteristics (X_2)	Stand Age Mean	Mean age of forest stands (years)
	Conifer Ratio	Proportion of coniferous forest area (%)
	Diameter Class	Mean diameter class of trees (cm/dmcls)
	Stand Density	Degree of forest stocking/density (%)
Infrastructure Factors (X_3)	Average Height	Mean height of forest stands (m)
	Road Density	Total length of forest roads per grid (km/km ²)
	Trail Density	Total length of hiking trails per grid (km/km ²)
	Distance to Road	Euclidean distance to the nearest road (km)
Forest Management Factors (X_4)	Distance to Trail	Euclidean distance to the nearest trail (km)
	Artificial Forest Tending	Area ratio of artificial forest management activities (%)
	Natural Forest Tending	Area ratio of natural forest management activities (%)
Temporal Factors (X_5)	Other Management	Area ratio of other management activities (%)
	Season	Categorized as Spring, Summer, Fall, and Winter based on occurrence date (Reference: Summer)
Target (Y)	Fire Occurrence	Daily wildfire occurrence (Binary: 0 or 1)

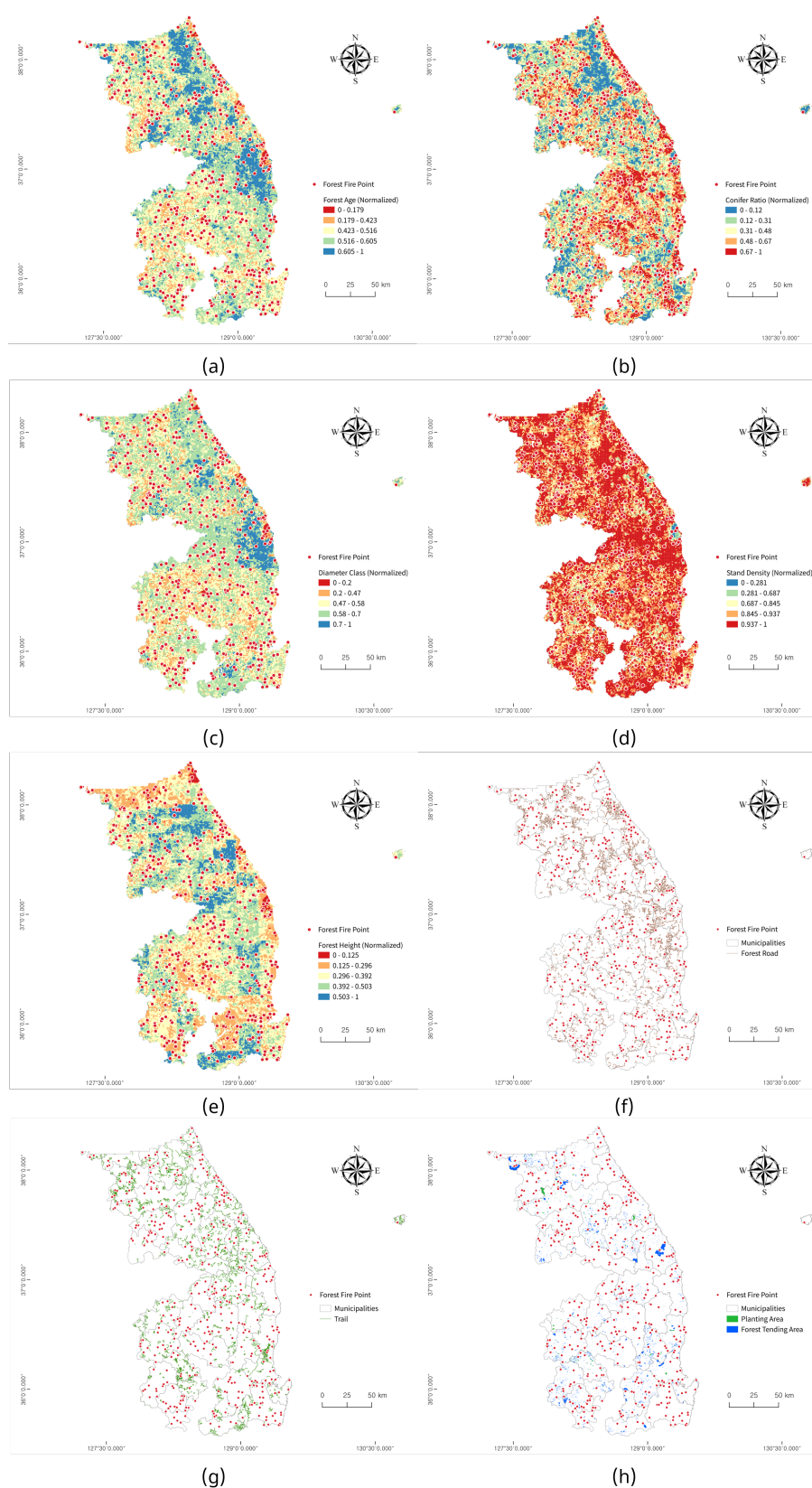


Figure 4. Maps showing the distribution of forest attributes, infrastructure, and management areas with forest fire breakout points: (a) normalized stand age mean, (b) normalized conifer ratio, (c) normalized diameter class, (d) normalized stand density, (e) normalized forest height, (f) municipal boundaries and forest road, (g) municipal boundaries and hiking trails, and (h) municipal boundaries with forest management areas (planting and forest tending zones).

2.2.1. Weather Factors (X_1)

Daily weather data provided by the Korea Meteorological Administration (KMA) were utilized. For each Grid point, the shortest distance to weather stations was calculated to extract data from the nearest station. The extracted data included effective humidity (eff_hum; %), daily precipitation (daily_precip; mm), and maximum wind speed (max_wind; m/s). For occurrence grids, weather data from the day before the fire (D-1) were used. For non-occurrence grids, weather data from a random day between January 2022 and August 2025 were extracted.

Particularly, effective humidity (eff_hum) was employed as a crucial indicator of the moisture content in forest fuels, which reflects the cumulative dryness of the environment. Unlike simple daily average humidity (avg_hum), effective humidity is calculated as a weighted moving average of the daily relative humidity over a specific preceding period. Following the KMA standard, this study calculated effective humidity using a 5-day window ending on the analysis date (D-1 for occurrence grids). A decay coefficient ($r = 0.7$) was applied to assign higher weights to more recent days according to the following formula:

$$H_e = (1 - r) \sum_{i=0}^4 r^i H_i \quad (1)$$

where avg_hum_{*i*} represents the average humidity *i* days prior to the analysis date (with $i = 0$ being the analysis date itself, up to $i = 4$). This 5-day cumulative approach effectively captures the persistent drying conditions that critically influence wildfire ignition probabilities.

2.2.2. Forest Characteristics (X_2)

To reflect the ecological structure and physical state of the forest, the 2024 large-scale forest type map (1:5,000) produced by the Korea Forest Service was used. Based on the spatial data, five variables were calculated per grid: mean stand age (stand age mean; years) indicating maturity, coniferous tree ratio (conifer_ratio; ratio) representing species composition, mean diameter class (dmcls; cm), stand density (dnst; %), and mean tree height (height; m).

2.2.3. Infrastructure Factors (X_3)

Data for national forest roads and trails were sourced from the Korea Forest Service's Forest Spatial Information Service. Data for public and private forest roads were obtained via information disclosure requests from officials in Gangwon and Gyeongbuk provinces. Infrastructure factors act as indicators of human accessibility and firefighting resources. The forest road density (road_density; km/km²), trail density (trail_density; km/km²), and the distance to the nearest road and trail (dist_road, dist_trail; km) per grid were calculated through GIS spatial analysis².

Table 3. Characteristics of forest roads and forest area in the study area (2023).

Region (Si-Do)	Forest Area (ha)	Forest Road Length (km)	Forest Road Density (m/ha)	Hiking Trail Length (km)	Hiking Trail Density (m/ha)
Gangwon-do	1,365,746	5,496.47	4.02	4,973.95	3.64
Gyeongsangbuk-do	1,286,222	4,464.91	3.47	4,562.94	3.55
Total / Average	2,651,968	9,961.38	3.76	9,536.89	3.60

Source: Korea Forest Service (2024) Statistical Yearbook of Forestry.

2.2.4. Forest Management Factors (X_4)

Spatial data on tending and afforestation projects in public and private forests from 2015 to 2017 provided by the Korea Forest Service were utilized. The implementation records for detailed forest

² Infrastructure density was calculated as Total Length per grid area using the QGIS 'Sum line lengths' function, while infrastructure distance was calculated as the Euclidean distance (km) from the grid Centroid to the infrastructure object using the 'Distance to nearest hub' function.

management activities in the study areas are shown in Table 4. The 12 detailed activities from the collected data were calculated as Area Ratios, dividing the total activity area performed within each grid (1 km × 1 km) by the grid's total area. These variables were classified into three groups based on their purpose:

- **Artificial Forest Tending:** Pruning, thinning, tending of young trees, forest debris clearance, weeding, planting.
- **Natural Forest Tending:** Tending for public benefits, natural forest improvement, natural forest conservation.
- **Other Forest Management:** Others, vine removal, byproduct collection.

Table 4. Summary of forest management activities by region (2015–2017).

Management Groups	Specific Activities	Gangwon (ha)	Gyeongsangbuk (ha)
Artificial Tending	Pruning	13.7	1.5
	Thinning	3,852.6	2,258.4
	Young tree tending	2,344.5	1,737.5
	Stand cleaning	681.6	11.2
	Weeding	4,993.6	5,347.6
	Planting	2,017.7	1,311.8
Natural Tending	Public forest tending	6,189.5	2,910.3
	Natural forest improvement	693.5	10,437.3
	Natural forest tending	2,456.0	5,005.2
Other Management	Others	0.0	279.9
	Vine removal	867.8	931.8
	Logging residue collection	55.7	840.1
Total Area		24,166.2	30,132.6

A preliminary validation was conducted using the AUC (Area Under the ROC Curve) to evaluate which configuration—treating activities individually or grouping them—best predicts wildfire risk. The validation showed that the model's predictive performance was superior when utilizing the three grouped categories (Artificial Tending, Natural Tending, and Other Management; AUC: 0.7590) compared to inputting 12 individual variables or integrating them into a single category (AUC: 0.7491). Thus, the three-group classification system was adopted.

2.2.5. Temporal Factors (X_5)

To account for temporal variations in wildfire occurrence, a 'Season' variable was derived from the mapped date of each event. The months were grouped into four categories: Summer (June–August), Spring (March–May), Fall (September–November), and Winter (December–February). For statistical analysis, these were transformed into dummy variables, with Summer serving as the reference category to prevent multicollinearity.

2.3. Feature Selection Based on Machine Learning

To extract the core variables substantially impacting wildfire occurrence among various independent variables, statistical multicollinearity diagnostics and machine learning algorithms were used concurrently. First, to prevent variables that behave similarly from degrading the model's statistical accuracy, Variance Inflation Factor (VIF) values were calculated to diagnose multicollinearity. Stable variables with a VIF of less than 10 were initially selected (Table 5).

Table 5. VIF-based Primary Selection.

Category	Selected Independent Variables	VIF
Infrastructure	Road density, Trail density, Distance to road, Distance to trail	< 10
Management treatment	Other Management, Artificial forest tending, Natural forest tending	< 10
Environmental factors	Conifer ratio, Max wind speed, Daily precipitation	< 10
Temporal factors	Season (Spring, Fall, Winter)	< 10
Model Fit	Pseudo R ² = 0.1505, AUC = 0.7494	

Second, ensemble machine learning models (Random Forest, XGBoost) were introduced to capture complex interactions among variables. Wildfires are non-linear phenomena where weather, topography, and infrastructure intertwine. To overcome the limitations of traditional statistical methods that only consider linear relationships, tree-based models, which excel at finding hidden patterns, were utilized. The entire dataset was split into training (80%) and validation (20%) sets using stratified sampling.

Feature importance was extracted from each trained model to identify the top 10 variables (Table 6). The results demonstrated a high degree of consensus between the two algorithms, with 9 out of the top 10 variables overlapping. Notably, while the VIF-based model retained the 'Season' variables as significant linear predictors, both machine learning models excluded all seasonal variables from the top 10. Instead, effective humidity and stand age mean—which were excluded in the VIF diagnostics due to collinearity issues—emerged as the absolute top-tier variables in both models. This indicates that the machine learning algorithms accurately identified the fundamental physical trigger of wildfires (i.e., extreme dryness represented by low effective humidity) rather than relying on the superficial temporal proxy of 'Season'.

To construct the final probability function, the variable set from the Random Forest (RF) model was adopted, as it exhibited a slightly higher predictive performance (AUC: 0.8001, Pseudo R²: 0.1950) compared to XGBoost (AUC: 0.7960, Pseudo R²: 0.1818).

Table 6. Final Selection based on ML Feature Importance.

Variables Finally Selected	RF Rank	XGB Rank	Remarks
Effective humidity	1	1	Excluded based on VIF criteria
Stand age mean	2	3	Excluded based on VIF criteria
Conifer ratio	3	2	
Distance to road	4	4	
Distance to trail	5	5	
Max wind speed	6	8	
Road density	7	7	
Daily precipitation	8	-	Selected only by RF
Artificial forest tending	9	6	
Trail density	10	9	
Natural forest tending	-	10	Selected only by XGB
Model Fit (using Top 10)	AUC = 0.8001 Pseudo R² = 0.1950	AUC = 0.7960 Pseudo R² = 0.1818	The RF variable set was selected for the final logistic regression.

2.4. Estimation of Forest Fire Probability Function and Hypothesis Verification

The core objective of this study lies in causal inference rather than predictive accuracy per se. Specifically, this research aims to identify the direction and statistical significance of the impacts of anthropogenic factors—such as forest road density, trail density, and forest management activities—on wildfire occurrence. To achieve this inferential goal, a model that generates interpretable coefficients is essential. While 'black-box' models, such as Random Forest or deep neural networks, offer high raw predictive power, they lack the coefficient-level transparency required to refine the algorithms of the Korea Forest Service's National Forest Fire Danger Rating System.

Therefore, in this study, a Logistic Regression Model was constructed to determine the directionality and statistical significance of the top variables selected via machine learning on actual wildfire occurrence probability. Logistic regression is suitable for binary dependent variables, with wildfire occurrence set as 1 and non-occurrence as 0. The regression equation comprising the selected 10 explanatory variables is expressed as Equation (2):

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{10} X_{10} + \epsilon \quad (2)$$

Here, P is the probability of wildfire occurrence, β_0 is the constant term, β_1 to β_{10} are the regression coefficients representing the influence of each independent variable, and ϵ represents unobserved factors and the random error term not explained by the model's independent variables.

McFadden's Pseudo R^2 was used to evaluate the goodness-of-fit. The final machine learning-based model, which exhibited the highest explanatory power, was adopted. The regression coefficients (Coefficient), significance probability (P-value), and Odds Ratios derived from this model were calculated to verify the practical impact of each factor.

2.5. Contribution Analysis to Wildfire Risk Using SHAP

Despite high predictive performance, machine learning models possess a 'Black-box' characteristic, making internal decision-making processes hard to grasp. To overcome this, the SHAP (Shapley Additive exPlanations) technique, based on Game Theory, was introduced. SHAP quantitatively decomposes and explains each variable's contribution at the individual prediction level using Shapley Values [18]. SHAP allows for the observation of how wildfire risk changes as variable values shift at the individual grid level.

3. Results

3.1. Seasonal Distribution of Wildfire Occurrences

A total of 471 wildfire events were recorded across the study area from January 2022 to August 2025. Analysis of the monthly distribution revealed a pronounced seasonal concentration during the spring period (Figure 5). April recorded the highest share at 20.4% (96 events), followed by March (18.5%, 87 events), February (17.8%, 84 events), and January (12.1%, 57 events). Collectively, the January–April peak season accounted for 68.8% of all recorded wildfires, consistent with Korea's climatological pattern of low humidity and strong winds in spring.

In contrast, the summer and early autumn months (June–October) showed markedly suppressed activity, with monthly shares ranging from 1.1% (July, 5 events) to 3.6% (June, 17 events). A secondary, modest increase was observed in November (4.7%, 22 events) and December (6.2%, 29 events), reflecting the dry conditions of the winter season.

Regional disaggregation revealed that Gyeongbuk consistently dominated fire counts throughout the year, peaking at 51 events in February. Yeongseo exhibited its highest activity in April (39 events), while Yeongdong maintained relatively low but persistent counts across all months, with a maximum of 13 events in February. All three regions followed the same unimodal seasonal pattern, confirming that the spring peak is a province-wide phenomenon driven by shared meteorological conditions rather than region-specific factors.

Seasonal Wildfire Pattern — Korean Provinces (2022–2025)

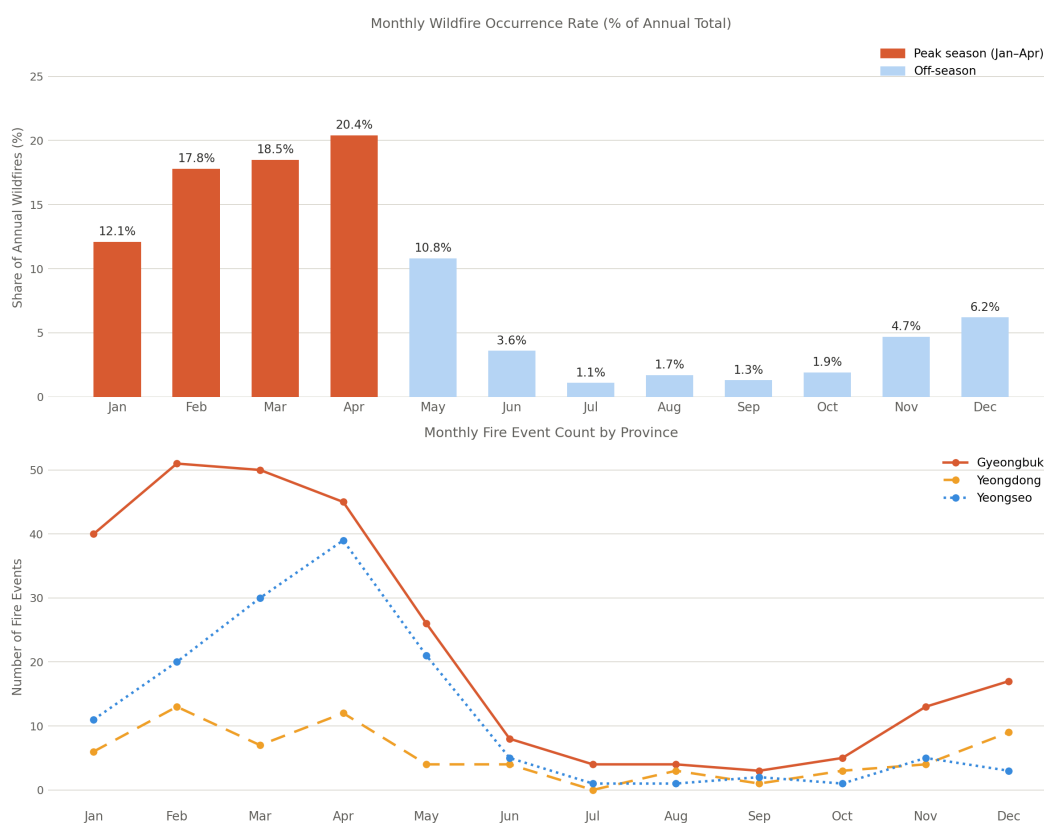


Figure 5. Seasonal wildfire pattern in Korean provinces (2022–2025). (Top) Monthly distribution of wildfire occurrences as a percentage of the annual total ($n = 471$). Peak season (January–April) is highlighted in red. (Bottom) Monthly fire event counts disaggregated by province (Gyeongbuk, Gangwon-Yeongdong, Gangwon-Yeongseo).

To further investigate the spatial distribution of these seasonal patterns, a heatmap was generated analyzing wildfire occurrences across municipal districts (Si/Gun/Gu) during the peak (January–April) and non-peak seasons (Figure 6). The heatmap reveals a dense concentration of peak-season wildfires in specific municipalities, particularly within the Gyeongbuk and Yeongseo regions. A prominent commonality among these high-frequency areas is their extensive agricultural land area. In the context of South Korea, the spring peak season precedes the annual agricultural cycle. The traditional, albeit illegal, practice of burning agricultural residues and field margins in these rural communities serves as a primary anthropogenic ignition source.

Furthermore, the spatial vulnerability of these agricultural areas is exacerbated by surrounding environmental factors. Based on the constructed dataset, municipalities with high fire frequencies consistently exhibit high proportions of coniferous forests and low effective humidity during the spring months. Consequently, the interface between expansive agricultural lands and dry, highly flammable coniferous forests creates an extremely susceptible environment where human-induced sparks from farming preparations rapidly escalate into significant wildfires during the peak season.

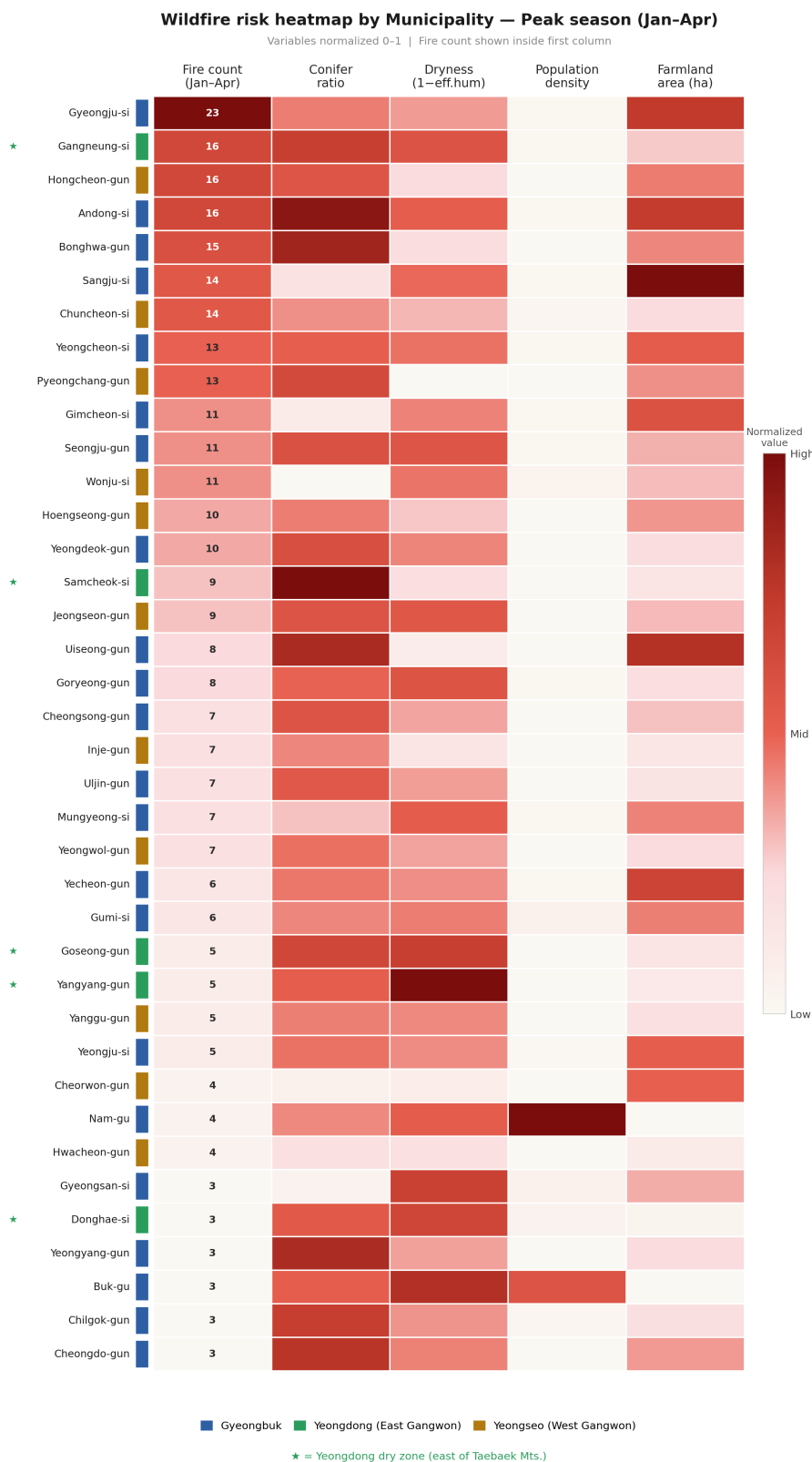


Figure 6. Heatmap of wildfire occurrences by municipalities (Si/Gun/Gu) highlighting the concentration during the peak season. *Note:* Population density data from the official website of Gyeongsangbuk-do (2025) and the Basic Statistics 2023 of Gangwon State (via KOSIS); agricultural area data from the Agricultural Area Survey (National Data Agency, 2025). All data were accessed on April 9, 2026. Data on the coniferous forest ratio and effective humidity were derived from the dataset constructed in this study.

3.2. Machine Learning-Based Forest Fire Probability Function Analysis Results

To reflect the non-linear and complex mechanisms of wildfire occurrence, the top 10 core variables derived from machine learning algorithms were incorporated into the final logistic regression model. The model's Pseudo R^2 was 0.1950, demonstrating valid analytical reliability with an explanatory power about 1.3 times higher than the logistic regression model excluding variables with high VIF values (0.1505). The final model's AUC score reached 0.8001, indicating excellent predictive performance. The results of the logistic regression and the Odds Ratios of each variable are shown in Table 7.

Table 7. Logistic regression analysis results for forest fire occurrence.

Variables	Coefficient	Std. Error	z-value	$P > z $	Odds Ratio
Intercept	3.7124	0.549	6.764	< 0.001***	-
Effective humidity	-0.0713	0.006	-12.012	< 0.001***	0.931
Conifer ratio	1.4446	0.293	4.926	< 0.001***	4.240
Stand age mean	-2.8588	0.737	-3.881	< 0.001***	0.057
Distance to road	1.1206	2.680	0.418	0.676	3.067
Distance to trail	-2.6321	1.632	-1.613	0.107	0.072
Road density	-2.7202	0.988	-2.752	0.006**	0.066
Max wind speed	0.0117	0.035	0.334	0.739	1.012
Artificial tending	0.0501	0.862	0.058	0.954	1.051
Trail density	1.4625	0.873	1.676	0.094*	4.317
Daily precipitation	-0.1881	0.067	-2.826	0.005**	0.829

Note: Pseudo $R^2 = 0.1950$. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Odds Ratio is calculated as $\exp(\text{Coefficient})$.

To precisely interpret non-linear patterns between variables and individual data contributions that logistic regression struggles to capture, a SHAP Summary Plot was analyzed. As a result, effective humidity (eff_hum), mean stand age (stand age mean), and coniferous tree ratio (conifer_ratio) were identified as the top contributing variables to fire prediction. Variables shifting towards the positive (+) direction when the point color is red (high variable value), such as coniferous ratio and trail density, aggravate wildfire risk. Conversely, variables moving towards the negative (-) direction, such as stand age and road density, suppress the risk (Figure 7).

Figure 1. SHAP summary plot for forest fire occurrences

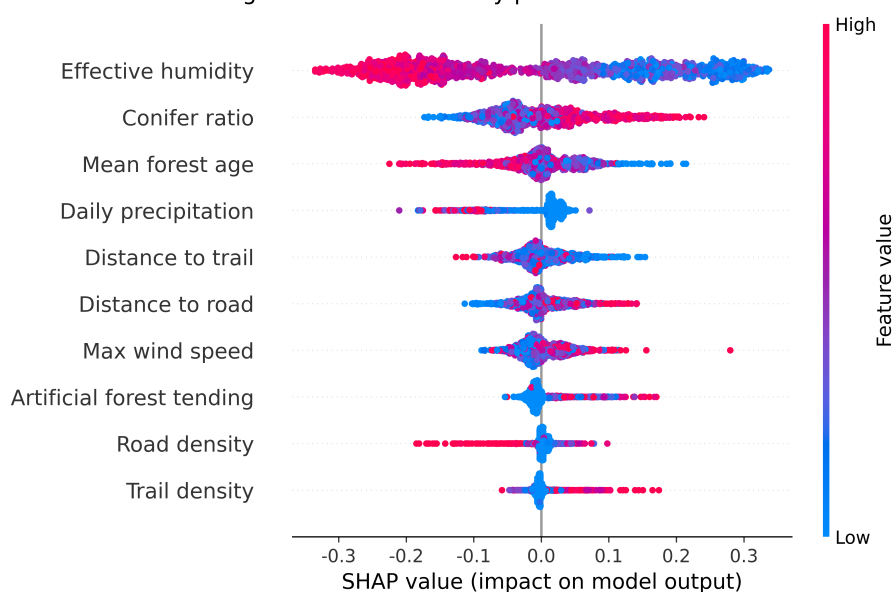


Figure 7. SHAP summary plot (beeswarm plot).

3.3. Hypothesis 1 Testing: Impact of Forest Age and Species Composition

Based on the regression results in Table 7, Hypothesis 1 was statistically supported. The regression coefficient for mean stand age was -2.8588 ($p < 0.001$), identifying it as the strongest suppressor of wildfire occurrence among the model's variables. This implies that as forests mature, canopy layers develop, and the understory microclimates stabilize, significantly enhancing the forest's ecological resistance to fires. The coniferous tree ratio had a coefficient of 1.4446 ($p < 0.001$), acting as a positive (+) factor that significantly increases fire risk. Odds ratio analysis indicated that areas with a high proportion of conifers are roughly 4.24 times more likely to experience wildfires compared to other areas.

In the SHAP plot regarding age and species, the mean stand age showed a trend of SHAP values falling below 0 during the maturation stage (Figure 8). This indicates high vulnerability in young forests but a decreasing risk as age increases. In contrast, as the coniferous ratio increased, SHAP values rose linearly, showing a clear pattern of heightening fire risk.

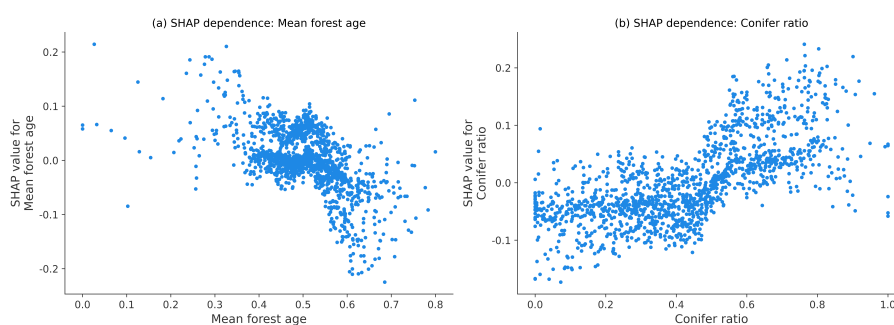


Figure 8. SHAP dependence plots for Stand age mean and conifer ratio.

3.4. Hypothesis 2 Testing: Impact of Artificial Forest Tending Activities

Hypothesis 2, which posited an association between artificial forest tending activities and wildfire occurrence, was rejected in this analysis. The regression coefficient for artificial forest tending was 0.0501 ($p = 0.954$), indicating no significant relationship. This does not support the argument that anthropogenic forest management activities like afforestation or thinning directly cause wildfires. This can be interpreted to mean that other factors, such as meteorological conditions and species composition at the time, have an overwhelmingly dominant influence on ignition and spread.

On the SHAP plot for artificial forest tending, the majority of data points were densely clustered around a SHAP value of 0. This suggests that the marginal contribution of fluctuations in forest management ratios to the prediction of wildfire occurrence in individual grids is negligible.

Figure 3. SHAP dependence: Artificial forest tending

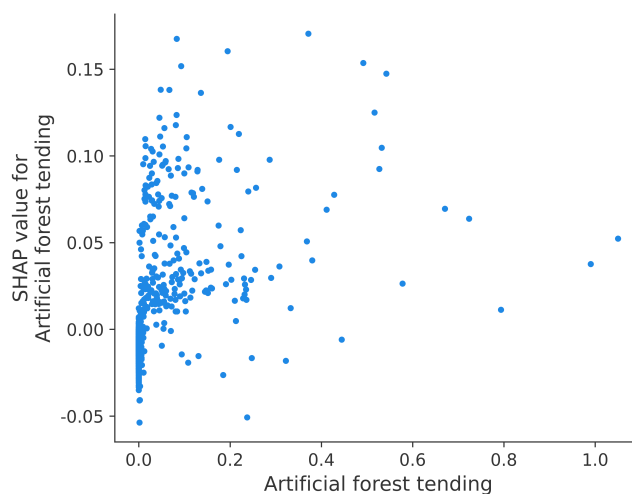


Figure 9. SHAP dependence plot for artificial forest tending.

3.5. Hypothesis 3 Testing: Impact of Forest Road and Trail Infrastructure

Hypothesis 3, which suggested that infrastructure factors increase fire risk, was partially supported as conflicting results emerged depending on the infrastructure type. Trail density recorded a coefficient of 1.4625 ($p = 0.094$), showing a tendency to increase fire probability within a 10% significance level. The odds ratio was high at 4.317, confirming that areas with frequent hiker access face aggravated fire risks due to accidental ignitions. In stark contrast, forest road density (road_density) significantly decreased the occurrence probability with a coefficient of -2.7202 ($p = 0.006$). This implies that forest roads facilitate the rapid deployment of firefighting equipment and personnel, thereby preventing the spread of flames and reducing damage probabilities.

The SHAP graphs for infrastructure variables clearly illustrate these opposing roles. As trail density increases, SHAP values rise in the positive (+) direction, indicating heightened risk; whereas for road density, higher densities push SHAP values in the negative (-) direction, exhibiting a suppressive trend on occurrences.

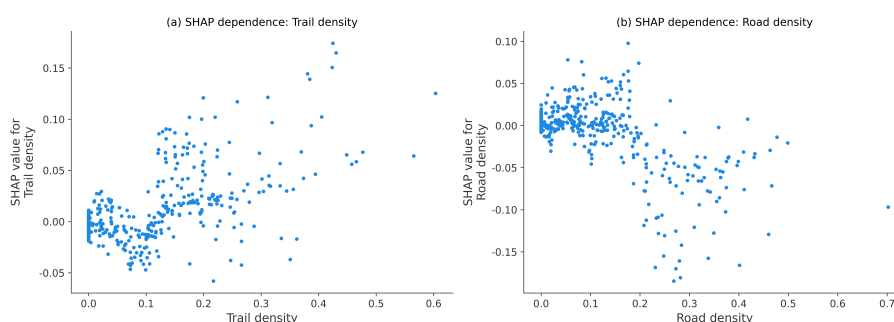


Figure 10. SHAP dependence plots for trail density and road density.

3.6. Impact of Other Environmental Variables: Effective Humidity and Precipitation

Meteorological factors acted as critical control variables determining wildfire occurrence in both machine learning importance evaluations and logistic regression. Effective humidity (eff_hum) had a coefficient of -0.0713 ($p < 0.001$), and daily precipitation (daily_precip) was -0.1881 ($p = 0.005$), confirming that drier atmospheres drastically and significantly increase the probability of fires. These results align with previous studies (Kang et al. [31]). On the effective humidity SHAP plot, dropping below a specific dryness threshold caused SHAP values to spike, aggravating the risk (Figure 11). In zones with sufficient humidity, risk was consistently suppressed. This suggests that even under identical topographical, structural, and infrastructure conditions, reaching meteorological tipping points has a profound impact on triggering wildfires.

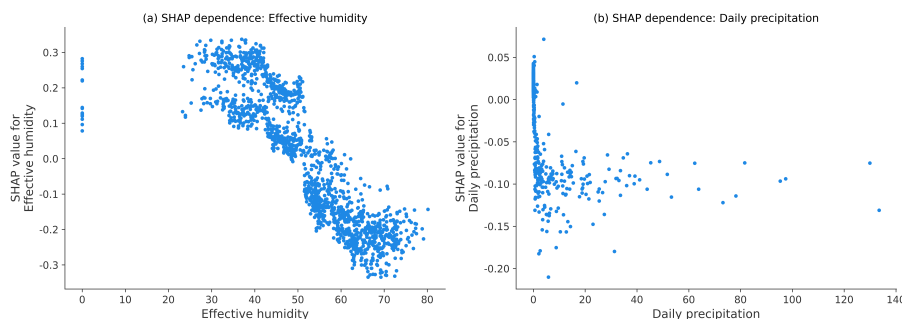


Figure 11. SHAP dependence plots for effective humidity and daily precipitation.

4. Discussion

4.1. Wildfire Suppression Effect of Old-Growth Forests and Ecological Mechanisms

The analysis revealed that among ecological factors, the mean stand age lowered fire probability second only to meteorological variables. This supports our hypothesis that mature forests ecologically

suppress wildfires, aligning with Zald and Dunn's [19] findings that young forests heavily impact fire severity. Immature forests or homogeneous artificial stands often have canopies close to the ground, serving as ladder fuels that carry flames upward, and they tend to have abundant dry fine debris, making them highly vulnerable to ignition.

According to the JRC/EU [21] report and Carroll et al. [20], old-growth forests represent not merely a senescent stage but rather sustain ecosystem integrity across spatial and temporal scales. The vertical and horizontal heterogeneity and dense canopies of mature forests provide an insulating effect by shading the forest floor, lowering temperatures, and retaining moisture. The coarse woody debris accumulated in old-growth forests retains large amounts of water, physically impeding ignition and spread even during climatic disasters.

In Korea, forest growing stock was officially recorded at 1,040,447,000 m³ (165.2 m³/ha) as of the end of 2020. This is an 18.4-fold increase compared to 1946 and a 14-fold increase from 1973. The trees planted during the national reforestation efforts are now mature forests aged 31 to 50 years. Recently, there has been active debate between forest policies focused on short-rotation clearcutting for economic timber and carbon absorption of young forests versus ecological preservation. At this juncture, forest policies must be established through scientific evaluations of stand age, climate change mitigation, and ecosystem services. Our findings present crucial evidence that forest management extending stand age can serve as a Nature-based Solution (NbS) for suppressing mega-fires in the climate crisis era.

However, since topographical characteristics were not jointly considered, this study could not clarify whether the suppressive effect of long-rotation management purely stems from ecological structures like insulation or from a combination of topographical isolation that restricts human activity. Further research is needed.

Currently, research verifying whether aging forests suppress wildfires in Korea is scarce. Therefore, in-depth follow-up studies utilizing remote sensing technologies to explore the relationship between forest structure, stand age, and wildfire resistance are required.

4.2. *The Paradox of Human Activity Infrastructure: Conflicting Roles of Forest Roads and Trails*

Anthropogenic infrastructures demonstrated opposing impacts depending on their characteristics. An increase in hiking trail density was identified as an ignition factor raising fire risk. In Korea, dry spring and autumn seasons coincide with peak hiking periods. Fine fuels like dry fallen leaves are easily exposed around trails. Thus, accidental fires caused by human negligence, such as discarded cigarette butts or illegal cooking, readily escalate into actual wildfires.

Conversely, increased forest road density served as a suppressor. While forest roads could act as potential ignition sources by increasing human access, they simultaneously perform a vital fire prevention role regarding emergency response. During a fire, roads enable rapid access for fire trucks and personnel, facilitating mopping-up operations and restricting spread. They are functionally indispensable, especially at night or when helicopters cannot be deployed.

Nevertheless, arguments exist advocating for minimizing road construction due to increased landslide risks and ecosystem fragmentation. Studies investigating the link between roads and fires also present varied outcomes. Hong et al. [3] suggested roads could be primary ignition sources by increasing accessibility. Yet, others (Lee et al. [30]) argue they inhibit the spread. Thus, sophisticated empirical research proving the exact effects of forest roads is required, alongside the development of construction methods that minimize ecological damage.

5. Conclusions

To combat the increasing scale and routine nature of wildfire disasters driven by climate change, this study investigated the complex impacts of meteorological, ecological, and anthropogenic factors on wildfire occurrences in Gangwon and Gyeongbuk provinces of KOR via logistic regression analysis. The results confirm that, excluding weather, the most critical fire-suppressing factor is forest stand age. We found that a high proportion of coniferous forests and increased trail density could serve as primary ignition and spread factors of forest fires. Furthermore, increased forest road density

significantly reduces occurrence probability, identifying it as a core firefighting asset. The hypothesis that artificial forest tending increases fuel loads and fire risk requires further follow-up research.

These findings suggest the necessity of re-evaluating Korea's current forest policies centered on economic timber and short-rotation logging. To effectively mitigate fire damage and enhance climate resilience, forests should be managed until old-age classes by extending rotation period. Additionally, ecological forest tending that transitions highly flammable, uniform artificial coniferous forests into fire-resistant broadleaf ecosystems is necessary.

Since this study conducted macroscopic spatial analyses at a 1 km grid resolution, the effects of microscopic changes in understory microclimates conditioned by topographical factors like slope and aspect were not addressed. Future studies are necessary for understanding the roles of canopy strata structure and forest floor fuel loads on microscopic wildfire spread mechanisms.

Author Contributions: Conceptualization, Y.-C.Y. and S.-E.L.; methodology, S.-J.L.; software, H.-R.K.; validation, Y.-C.Y. and S.-J.L.; formal analysis, S.-E.L.; investigation, S.-E.L.; resources, Y.-C.Y.; data curation, S.-E.L. and H.-R.K.; writing—original draft preparation, S.-E.L.; writing—review and editing, Y.-C.Y.; visualization, S.-E.L.; supervision, Y.-C.Y.; project administration, Y.-C.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors gratefully acknowledge the Korea Forest Service and the Korea Meteorological Administration for providing the official national wildfire statistics and climatological data used in this study.

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

References

1. Byrne, B.; Liu, J.; Bowman, K.W.; et al. Carbon emissions from the 2023 Canadian wildfires. *Nature* **2024**, *633*, 835–839. <https://doi.org/10.1038/s41586-024-07878-z>.
2. Korea Forest Service. *Explanation of the Risk Index Calculation Algorithm for the National Forest Fire Danger Rating System*; Forest Fire Research Division: Daejeon, Republic of Korea, 2024. (In Korean)
3. Hong, S.; Ahn, M.; Hwang, J. The Effect of Road Density and Vegetation Type on Large Forest Fire Damage—Centered on the 2023 Hongseong Forest Fire. *Korean J. Environ. Ecol.* **2024**, *38*, 634–645. <https://doi.org/10.13047/KJEE.2024.38.6.634>. (In Korean)
4. Park, J.; Kwon, S.; Cho, S.; Lee, G.; Ryu, S. Deriving Improvement Directions through Case Analysis of Forest Fire Response in Korea. *Crisisonomy* **2025**, *15*, 45–58. (In Korean)
5. National Assembly Research Service. *National Response Tasks for Large-Scale Wildfires: In the Wake of the 2025 Yeongnam Region Large Wildfire (Special Report of the Forest Fire Response Research TF)*; National Assembly Research Service: Seoul, Republic of Korea, 2025. (In Korean)
6. Abatzoglou, J.T.; Williams, A.P. Impact of anthropogenic climate change on wildfire across western US forests. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 11770–11775.
7. Bowman, D.M.J.S.; Kolden, C.A.; Abatzoglou, J.T.; Finn, M.; Johnston, F.H.; van der Werf, G.R.; Flannigan, M. Vegetation fires in the Anthropocene. *Nat. Rev. Earth Environ.* **2020**, *1*, 500–515.
8. Lee, C.B.; et al. *Scientific Understanding of Forest Fire Management*; Jieul: Seoul, Republic of Korea, 2023. (In Korean)
9. Zheng, B.; et al. Record-high CO₂ emissions from boreal fires in 2021. *Science* **2023**, *379*, 912–917. <https://doi.org/10.1126/science.ade0805>.
10. IPCC. *Climate Change and Land: An IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*; Intergovernmental Panel on Climate Change: Geneva, Switzerland, 2019.

11. IPCC. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2021.
12. Provisional scale of forest fire damage in Gyeongbuk, Gyeongnam, and Ulsan is 104 thousand ha, Korea Forest Service is doing its best for restoration. Available online: <https://buly.kr/1xzblGH> (accessed on 21 March 2026).
13. Won, M.; Jang, K.; Yoon, S. Development of a National Integrated Forest Fire Occurrence Probability Model Based on Spring and Autumn Meteorology. *Korean J. Agric. For. Meteorol.* **2018**, *20*, 348–356. <https://doi.org/10.5532/KJAFM.2018.20.4.348>. (In Korean)
14. Ryu, J.; Kim, S.; Lim, C.; Kwon, C. A Study on Resetting the Forest Fire Caution Period According to Climate Change. *Crisisonomy* **2024**, *20*, 83–91. <https://doi.org/10.14251/crisisonomy.2024.20.1.83>. (In Korean)
15. Kwak, H.; Lee, W.; Saborowski, J.; Lee, S.; Won, M.; Koo, K.; Lee, B.; Kim, S. Estimating the spatial pattern of human-caused forest fires using a generalized linear mixed model with spatial autocorrelation in South Korea. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1589–1606. <https://doi.org/10.1080/13658816.2011.642799>.
16. Lee, J.; Lim, C.; Kim, G.; Kafatos, M.; Lee, W. Multi-temporal analysis of forest fire probability using socio-economic and environmental variables. *Remote Sens.* **2019**, *11*, 86. <https://doi.org/10.3390/rs11010086>.
17. Kim, S.; Lim, C.; Kim, G.; Lee, W. Spatial and temporal variability of forest fires in the Republic of Korea over 1991–2020. *Nat. Hazards* **2025**. <https://doi.org/10.1007/s11069-025-07169-4>.
18. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
19. Zald, H.S.J.; Dunn, C.J. Severe fire weather and intensive forest management increase fire severity in a multi-ownership landscape. *Ecol. Appl.* **2018**, *28*, 1068–1080. <https://doi.org/10.1002/eap.1710>.
20. Carroll, C.; Noon, B.R.; Masino, S.A.; Noss, R.F. Coordinating old-growth conservation across scales of space, time, and biodiversity: Lessons from the US policy debate. *Front. For. Glob. Change* **2025**. <https://doi.org/10.3389/ffgc.2025.14938794>.
21. European Union. *Primary and old-growth forests are more resilient to natural disturbances: Perspective on wildfires (JRC133970)*; Joint Research Centre: Ispra, Italy, 2023.
22. Yang, C. A Study on the Improvement of Forest Fire Response System in Korea. Master's Thesis, University of Seoul, Seoul, Republic of Korea, 2017. (In Korean)
23. Lee, Y.; Kwak, C.; Kim, Y.; Kim, K. Analysis of the Current Status and Trends of Forest Fires in Korea. In Proceedings of the 2022 KICS Winter Conference, Pyeongchang, Republic of Korea, 9–11 February 2022. (In Korean)
24. Lee, M.-W.; Lee, S.-Y.; Lee, J.H. Study of the Characteristics of Forest Fire Based on Statistics of Forest Fire in Korea. *J. Korean Soc. Hazard Mitig.* **2012**, *12*. <https://doi.org/10.9798/kosham.2012.12.5.185>. (In Korean)
25. Kang, R.-Y.; Hong, S.-H. The Effect on the Forest Temperature by Reduced Biomass Caused by Natural Forest Thinning. *Korean J. Environ. Ecol.* **2018**, *32*, 303–312. <https://doi.org/10.13047/KJEE.2018.32.3.303>. (In Korean)
26. Lee, S.-Y.; Lee, M.-W.; Lee, H.-P. Comparative Analysis of Forest Fire Danger Rating on Accumulation Types of the Leaving of Thinning Slash. *J. Korean Inst. Fire Sci. Eng.* **2008**, *22*. (In Korean)
27. Lee, S.-Y.; Lee, M.-W.; Yeom, C.-H.; Kwon, C.-G.; Lee, H.-P. Comparative Analysis of Forest Fire Danger Rating on Forest Characteristics of Thinning Area and Non-thinning Area on Forest Fire Burnt Area. *J. Korean Inst. Fire Sci. Eng.* **2009**, *23*. (In Korean)
28. Lee, S.J.; Kwon, C.G.; Seo, K.W.; Lee, Y.J.; Kim, S.Y. Thinning Effect on Fuel Load and Crown Fire Hazard - A Case Study of Pinus densiflora in Goseong, Gangwon Province. *Crisisonomy* **2023**, *19*, 27–37. <https://doi.org/10.14251/crisisonomy.2023.19.5.27>. (In Korean)
29. Lee, Y.E.; Lee, S.J.; Kwon, C.G.; Seo, K.W.; Bang, C.A.; Kim, S.Y. The Effects of Thinning Slash on Wildfire Fuel Type. *Crisisonomy* **2020**, *16*, 61–69. <https://doi.org/10.14251/crisisonomy.2020.16.10.61>. (In Korean)
30. Lee, H.-E.; Kwon, S.; Lim, C.-H. Investigating the Environmental Influencing Factors on Large Wildfire Spread Rate Considering the Spatial Configuration of Forest Roads. *J. Korean Soc. For. Sci.* **2025**, *114*, 558–569. (In Korean)
31. Kang, S.-C.; Won, M.; Yoon, S. Large Fire Forecasting Depending on the Changing Wind Speed and Effective Humidity in Korean Red Pine Forests Through a Case Study. *J. Korean Assoc. Geogr. Inf. Stud.* **2016**, *19*, 146–156. (In Korean)

32. Basic statistics of Gangwon State: Household and registered population by municipality. Available online: https://stat.kosis.kr/statHtml_host/statHtml.do?orgId=211&tblId=DT_211002_B002 (accessed on 9 April 2026).
33. Gyeongsangbuk-do statistics: Population. Available online: https://www.gb.go.kr/Main/page.do?mnu_uid=6816&LARGE_CODE (accessed on 9 April 2026).
34. Agricultural area survey [Data set]. Available online: http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1EB002&tmpScrId=20260409191452219_2b270226f3464dcc (accessed on 9 April 2026).

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.