

## Article

# Development of a Forest Fire Diagnostic Model Based on Machine Learning Techniques

Minwoo Roh <sup>1</sup>, Sujong Lee <sup>1</sup> , Hyun-Woo Jo <sup>2</sup> and Woo-Kyun Lee <sup>2,\*</sup> 

<sup>1</sup> Department of Environmental Science and Ecological Engineering, Korea University, Seoul 02841, Republic of Korea; mw0406toto@korea.ac.kr (M.R.); leesj3610@gmail.com (S.L.)

<sup>2</sup> OJEong Resilience Institute (OJERI), Korea University, Seoul 02841, Republic of Korea; endeavor@korea.ac.kr

\* Correspondence: leewk@korea.ac.kr; Tel.: +82-2-3290-3016

**Abstract:** Forest fires have devastating effects on extensive forest areas, compromising vital ecological services such as air purification, water conservation, and recreational opportunities, thus posing a significant socioeconomic threat. Furthermore, the risk of forest fires is steadily increasing due to climate change. The most effective method for mitigating forest fire risk is proactive prevention before forest fires can occur by identifying high-risk areas based on land surface conditions. This study aimed to develop a machine learning-based forest fire diagnostic model designed for Republic of Korea, considering both satellite-derived land surface data and anthropogenic factors. For the remote sensing data, VTCI (Vegetation Temperature Condition Index) was used to reflect the land surface dryness. In addition, fire activity maps for buildings, roads and cropland were used to consider the influence of human activities. The forest fire diagnostic model yielded an accuracy of 0.89, demonstrating its effectiveness in predicting forest fire risk. To validate the effectiveness of the model, 92 short-term forest fire risk forecast maps were generated from March to May 2023 with real-time data on forest fire occurrences collected for verification. The results showed that 73% of forest fires were accurately classified within high-risk zones, confirming the operational accuracy of the model. Through the forest fire diagnostic model, we have presented the impact relationships of meteorological, topographical, and environmental data, as well as the dryness index based on satellite images and anthropogenic factors, on forest fire occurrence. Additionally, we have demonstrated the potential uses of surface condition data.

**Keywords:** forest fire; diagnostic model; remote sensing; disaster risk reduction



**Citation:** Roh, M.; Lee, S.; Jo, H.-W.; Lee, W.-K. Development of a Forest Fire Diagnostic Model Based on Machine Learning Techniques. *Forests* **2024**, *15*, 1103. <https://doi.org/10.3390/f15071103>

Academic Editors: Palaiologos Palaiologou and Viacheslav I. Kharuk

Received: 26 April 2024

Revised: 29 May 2024

Accepted: 22 June 2024

Published: 26 June 2024



**Copyright:** © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Forest fires represent a significant socioeconomic threat with the capacity to devastate expansive forest areas. These blazes not only consume the forests themselves but also eliminate the myriad ecological services they provide, such as air purification, water conservation, and recreational opportunities. Notably, the impairment of the forest carbon sequestration ability and large amounts of carbon emissions pose substantial obstacles to achieving carbon neutrality goals [1–3]. Recent data underscores an alarming trend: the incidence and severity of forest fires are escalating, a pattern that is expected to intensify with the progressing climate crisis [4,5]. According to official statistics from the Korea Forest Service, the trend of forest fire frequency and damaged area increased in the 2020s compared to the past, from the 1980s to the 2000s. Approximately 8369 ha of forest was burned, and the number of incidents reached a peak of 580 in the 2020s [6].

While climate change is undoubtedly exacerbating weather conditions conducive to forest fires, the fact that most of these fires in Republic of Korea are caused by human activity cannot be overlooked, highlighting the necessity of robust prevention measures [7]. Central to the strategy of forest fire mitigation is the accurate prediction of fire risks. Most forest fire prediction models, which largely depend on meteorological data, do not

sufficiently consider the actual conditions of the land surface, indicating a potential area for improvement in which remote sensing technologies could be of significant use.

Studies on the use of satellite imagery for fire prediction have primarily focused on monitoring fuel dryness. Myoung et al. developed an empirical model function for live fuel moisture based on the correlation of satellite vegetation indices with in-situ fuel moisture [8]. Others, such as Verbesselt et al., compared satellite indices with meteorological drought indices, revealing that satellite indices can provide superior reflections of arid conditions [9]. Although these studies have validated the relevance of satellite data, they predominantly focused on long-term risk estimation and fell short of their applicability to daily risk assessment.

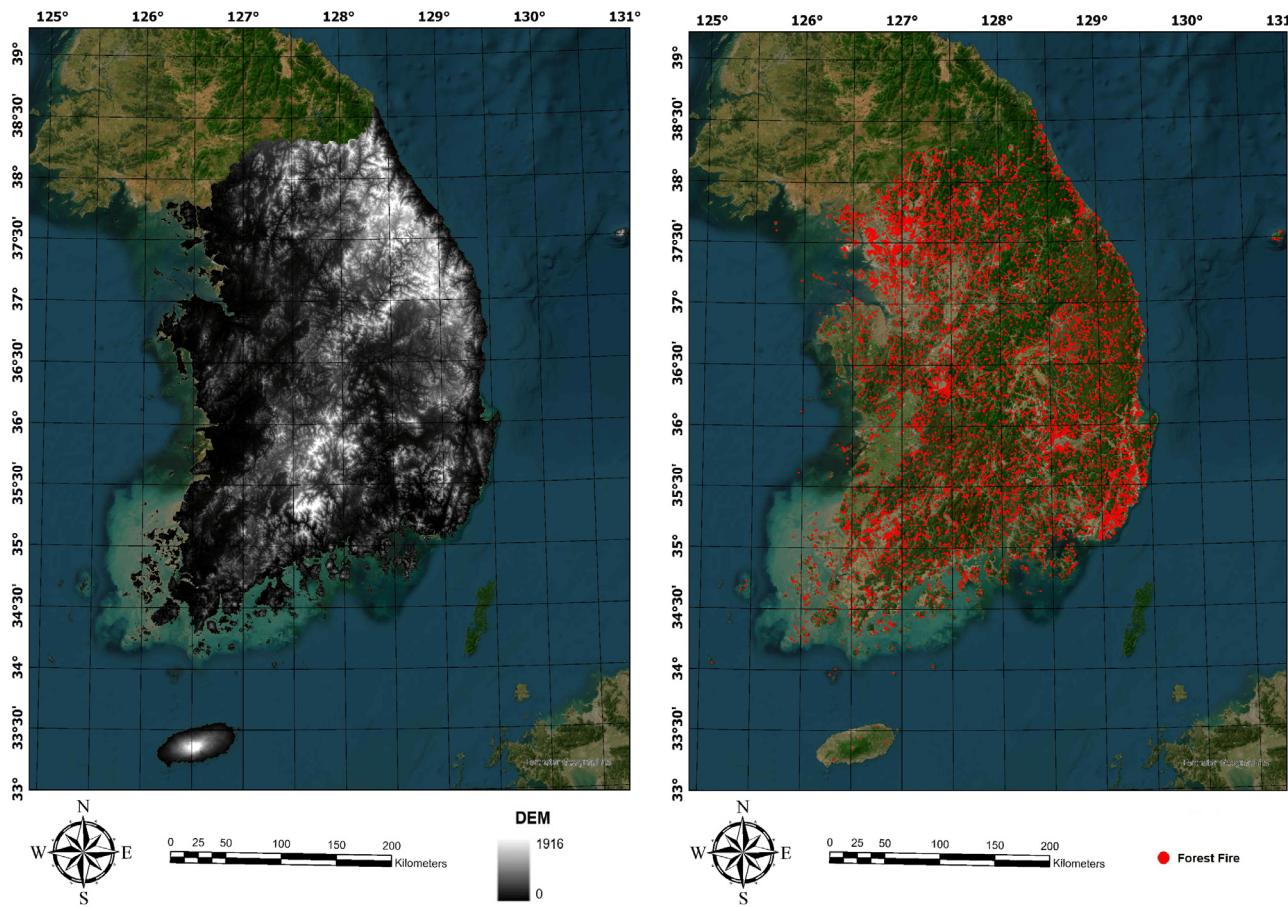
To address this limitation, Kang et al. yielded more granular spatiotemporal patterns through the integration of spatial road and population data, marking progress in fire risk prediction [7]. Similarly, Kim et al.'s machine-learning applications, incorporating socioeconomic trends, disclosed a direct relationship between population density and forest fire risk [10]. Nevertheless, these studies did not fully exploit surface condition data by using satellite imagery, and their resolutions were inadequate for identifying high-risk zones on a fine administrative scale. Forest fire risk forecasts rely on models that do not sufficiently consider the influence of anthropogenic activities, which are crucial factors affecting Korean forest fires.

Previous studies on forest fire prediction have faced limitations in fully reflecting surface conditions and have relied solely on historical data when utilizing satellite imagery. Therefore, this study developed a forest fire diagnostic model that accurately reflects the current surface conditions. To address the gaps commonly found in traditional satellite image applications, a model predicting satellite indices was developed, and its outputs were utilized as input data. Additionally, to simulate anthropogenic fire occurrences, fire activity maps that delineate the range of human activities were constructed and used.

## 2. Materials and Method

### 2.1. Study Area

The focus of this study was Republic of Korea, located between 124° and 132° East longitude and 33° and 43° North latitude, on the eastern side of the Eurasian continent and to the northwest of the North Pacific Ocean. The country's terrain is primarily defined by a mountainous spine that stretches from north to south, creating a gradient of elevation that is lower in the west and higher in the east. The average annual precipitation is 1306.3 mm, 54% of which occurs during the summer [11,12]. This uneven rainfall distribution contributes to extended periods of drought in spring and winter, which in turn increases the prevalence of forest fires [13,14]. In Republic of Korea, a significant proportion of forest fires are attributed to anthropogenic activities, particularly in regions close to residential areas bordering forests [15,16]. These areas experience more than 60% of forest fire incidents, whereas regions at higher elevations, which are more densely forested, tend to have fewer occurrences [7,17,18]. Comparing the forest fire inventory with the DEM, most points of forest fire were concentrated in low-elevation areas, while they were sparsely distributed in high-elevation areas (Figure 1).



**Figure 1.** Study area with DEM and forest fire inventory.

## 2.2. Data

To develop the forest fire diagnostic model and VTCI prediction model, forest fire label data, remote sensing data, and fire activity maps were utilized. In addition, meteorological, topographical, and environmental data were collected. The data are summarized in Table 1.

**Table 1.** Data description.

Dataset	Name	Source	Abbreviation
Label Data	Forest Fire Inventory Data	Korea Forest Service (KFS)	-
Remote Sensing Data	Vegetation Temperature Condition Index	Moderate Resolution Imaging	VTCI
	Land Surface Temperature	Spectroradiometer (MODIS)	LST
	Normalized Difference Vegetation Index		NDVI
Fire Activity Map	Road		-
	Building	National Geographic Information Institute (NGII)	-
	Cropland		-
Meteorological Data	Effective Humidity		EH
	Duff Moisture Contents		DMC
	Fine Fuel Moisture Contents	Korea Meteorological Administration (KMA)	FFMC
	Humidity		HMD
	Precipitation		PCP
	Wind Speed		WND

**Table 1.** Cont.

Dataset	Name	Source	Abbreviation
Meteorological Data	Average Temperature	Korea Meteorological Administration (KMA)	Avg TMP
	Maximum Temperature		Max TMP
	Minimum Temperature		Min TMP
Topographical and Environmental Data	Digital Elevation Model	National Geographic Information Institute (NGII)	DEM
	Slope		-
	Aspect		-
Landcover	Korea Ministry of Environment (KME)	LC	
	Coordinate (longitude and latitude)		-

### 2.2.1. Forest Fire Label Data

Forest fire inventory data was obtained from records maintained by the Korea Forest Service. These data, accessible via an API (Application Programming Interface) provided by the Forest Service, included the location addresses of past forest fires. The data were converted into a geospatial dataset containing latitude and longitude coordinates. However, given that the original data was based on address information, there were concerns regarding the accuracy of identifying the exact locations of the forest fires. To enhance the reliability of this dataset, a validation process was undertaken that included the use of historical media reports and visual inspections to refine the data. From 2003 to 2020, we collected data from 8115 locations where forest fires occurred.

To train a model for classifying forest fire occurrence areas, data on forest fire non-occurrence areas is required. Previous studies have primarily utilized ratios of occurrence to non-occurrence data ranging from 1:1 to 1:2 [19,20]. In this study, to distinguish forest fire risk areas based on as much data as possible, we constructed the data with a 1:2 ratio. In Republic of Korea, over 50% of wildfires occur in the spring and fall, while the risk significantly decreases in the summer. To clearly distinguish the actual differences in wildfire risk and train the model effectively, we applied different extraction ratios for non-occurrence points monthly. This approach ensures that fewer non-occurrence samples are extracted during high-risk months and more samples are extracted during low-risk periods.

The total number of non-occurrence data points was 16,000, accounting for approximately 66% of the label data. In Table 2, each column is marked with the signs (a) to (e). Column (a) represents the monthly counts of forest fires in the inventory data, and column (b) shows the monthly ratio of forest fires. Column (c) contains the values obtained by inverting the ratios in column (b), and column (d) represents the ratios of the values in column (c). The final counts of non-occurrence data are calculated by multiplying the values in column (d) by 16,000, the total number of non-occurrence cases. The number of non-occurrence labels for the monthly sample numbers was calculated as shown in Table 2.

**Table 2.** Method for calculating the monthly non-occurrence label sample sizes.

Month	(a) Forest Fires	(b) Ratio of Forest Fire	(c) Value	(d) Ratio of Value	(e) Non-Occurrence Data
January	634	8%	1279.96	3%	439
February	999	12%	812.31	2%	279
March	2053	25%	395.27	1%	136

**Table 2.** Cont.

Month	(a) Forest Fires	(b) Ratio of Forest Fire	(c) Value	(d) Ratio of Value	(e) Non-Occurrence Data
April	1990	25%	407.78	1%	140
May	697	9%	1164.27	2%	400
June	467	6%	11,737.68	4%	597
July	52	1%	15,605.76	33%	5358
August	87	1%	9327.58	20%	3203
September	94	1%	8632.97	19%	2964
October	266	3%	3050.75	7%	1048
November	390	5%	2080.76	4%	714
December	386	5%	2102.33	5%	722
Total	8115	100%	46,597.49	100%	16,000

### 2.2.2. Remote Sensing Data

We used the Vegetation Temperature Condition Index (VTCI) to monitor the dryness of forest areas as a component of land surface data. The VTCI measures the variance in Land Surface Temperature (LST) across pixels designated with specific Normalized Difference Vegetation Index (NDVI) values within a broad study region [21–23]. The VTCI was formulated as follows:

$$VTCI = \frac{LST_{NDVIi\max} - LST_{NDVIi}}{LST_{NDVIi\max} - LST_{NDVIi\min}} \quad (1)$$

$$LST_{NDVIi\max} = a + bNDVI_i \quad (2)$$

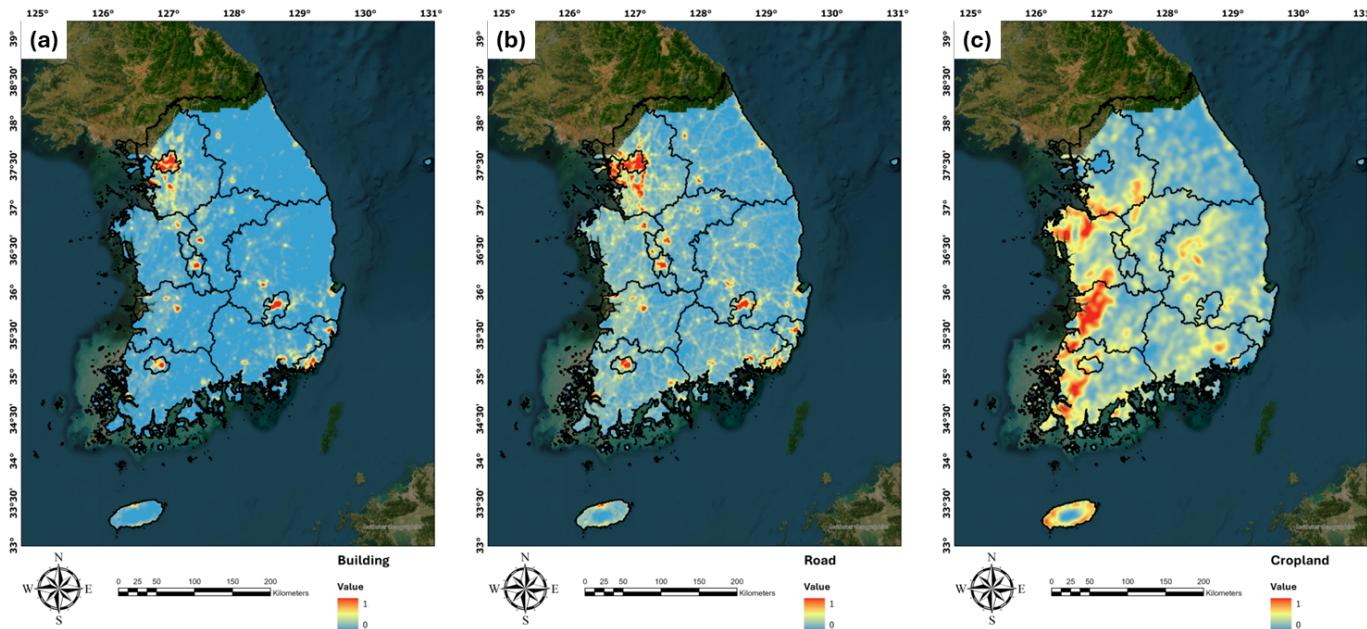
$$LST_{NDVIi\min} = a' + b'NDVI_i \quad (3)$$

$LST_{NDVIi\max}$  and  $LST_{NDVIi\min}$  represent the maximum and minimum LSTs of the pixels that have the same  $NDVI_i$ , respectively.  $LST_{NDVIi}$  is the LST of a single pixel with a specific  $NDVI_i$  value. The coefficients  $a$ ,  $b$ ,  $a'$ , and  $b'$  are determined from the study area, where soil moisture ranges from the wilting point to the field capacity [24].

The NDVI was calculated using bands 1 (red) and 2 (near-Infrared) from the MOD11A1 dataset provided by NASA's Terra Moderate Resolution Imaging Spectroradiometer (MODIS), whereas LST was obtained from MOD09GA data from the same source. All satellite image data were provided daily, and the VTCI values were calculated by excluding locations obscured by cloud cover. For missing values, a deep learning-based VTCI prediction model was employed for interpolation.

### 2.2.3. Forest Fires Activity Map

According to Korea Forest Service statistics, 33% of forest fires in Republic of Korea are caused by hikers' negligence, with other significant causes including agricultural residue burning (13%), garbage incineration (13%), and cigarette-related incidents (6%). These data highlight the prevalence of human influences on forest fire ignition [17]. The forest fire activity map is a dataset constructed to estimate the range of human activities based on anthropogenic data. In this study, data on buildings, roads, and agricultural land from digital topographic maps was used to create the forest fire activity data for each element. Three types of forest fire activity maps were developed using the kernel density method and applied to spatial data on agricultural land, buildings, and road infrastructure sourced from the National Geographic Information Institute of Republic of Korea (Figure 2).



**Figure 2.** Forest Fire Activity Maps. (a) Building; (b) Road; (c) Cropland.

#### 2.2.4. Meteorological Data

We collected and utilized the daily average humidity, daily total precipitation, daily average wind speed, daily average temperature, daily maximum temperature, and daily minimum temperature provided by the Korea Meteorological Administration (KMA). All meteorological data were sourced from KMA observation stations and converted into nationwide raster spatial data with a 1 km resolution using the Inverse Distance Weighting (IDW) interpolation method.

Meteorological factors such as Effective Humidity (EH), Fine Fuel Moisture Code (FFMC), and Duff Moisture Code (DMC) were incorporated. EH is a specialized index used to gauge the potential risk of forest fires based on the atmospheric moisture content. It is calculated by aggregating the relative humidity levels over a set period, typically incorporating data from both the current day and several preceding days, to provide a comprehensive picture of the atmospheric moisture availability that affects fire conditions [25,26].

The FFMC and DMC are two critical components of the Canadian Forest Fire Weather Index System (CFFDRS), designed to assess the moisture content of forest fuels, which is a key factor in determining the risk and behavior of forest fires [27,28]. The FFMC specifically targets the moisture content of fine fuels such as leaves, grasses, and small twigs that are less than 1/4 inch in diameter. These materials are highly responsive to weather conditions and can dry rapidly under appropriate conditions, thereby becoming highly flammable. The FFMC value ranges from 0 to 100, where lower values indicate a higher moisture content and a lower risk of fire ignition, whereas higher values suggest drier conditions and a higher susceptibility to ignition and support fire spread. This index is particularly sensitive to relative humidity and temperature changes, and can significantly fluctuate over the course of a single day [29–31].

DMC, on the contrary, focuses on the moisture content of moderately decomposed organic matter located below the litter of freshly fallen leaves or needles, which is commonly referred to as the duff layer. This layer is deeper and denser than the materials targeted by the FFMC and therefore responds more slowly to weather changes. This serves as a good indicator of the drying trends in these medium fuels, which are crucial for the sustained burning of fires. The DMC also ranges from 0 to a high value, typically approximately 300, with higher numbers indicating drier conditions and a greater potential for a fire to consume deeper fuel layers and persist once ignited [32,33].

### 2.2.5. Topographical and Environmental Data

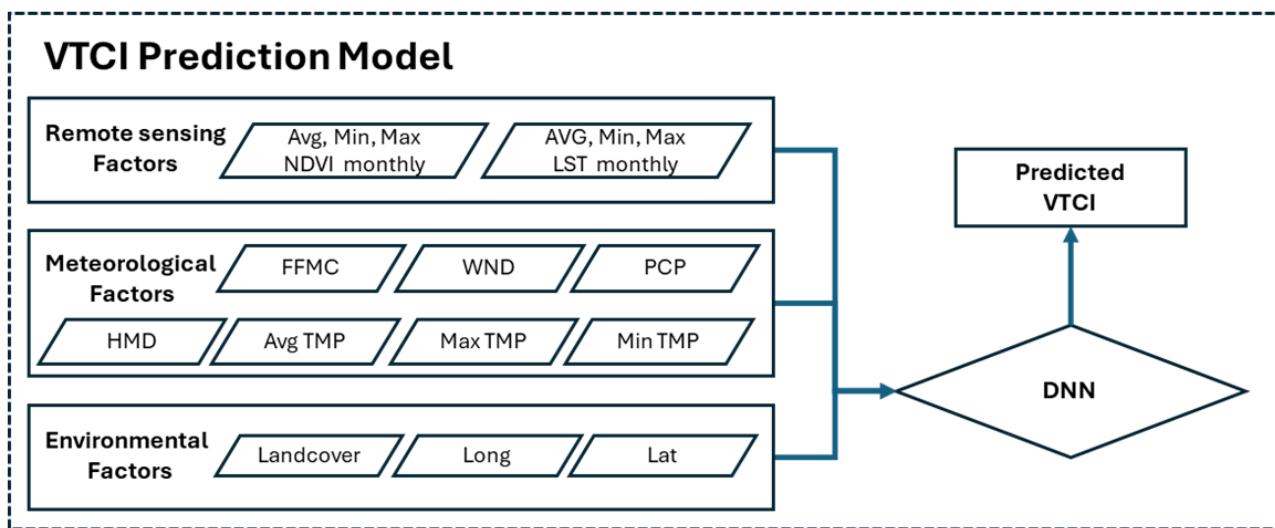
To evaluate the influence of topographic characteristics on forest fire occurrence, we employed data from a Digital Elevation Model (DEM) along with slope and aspect information, all of which were provided by the National Geographic Information Institute. Topographical factors influence human accessibility, solar radiation, and wind speed, which in turn correlate with soil moisture and humidity levels in forests, closely linking them to the occurrence of forest fires [10,15,34,35].

Additionally, we considered the ecological impact of various forest tree species and patterns of initial fire outbreaks across different types of land cover. This analysis was facilitated by a land-cover map from the Korean Ministry of Environment, which categorized the landscape into nine distinct types: urban areas, agricultural lands, various forest types (coniferous, deciduous, and mixed), grasslands, wetlands, barren lands, and water bodies. This detailed categorization helps understand how land cover variability contributes to fire dynamics.

For further analysis, longitudinal and latitudinal data were used to identify and track the patterns of recurrent fire occurrences. This geographical information aided in identifying high-risk areas and understanding the spatial distribution of fire events over time.

### 2.3. Development of the VTCI Prediction Model

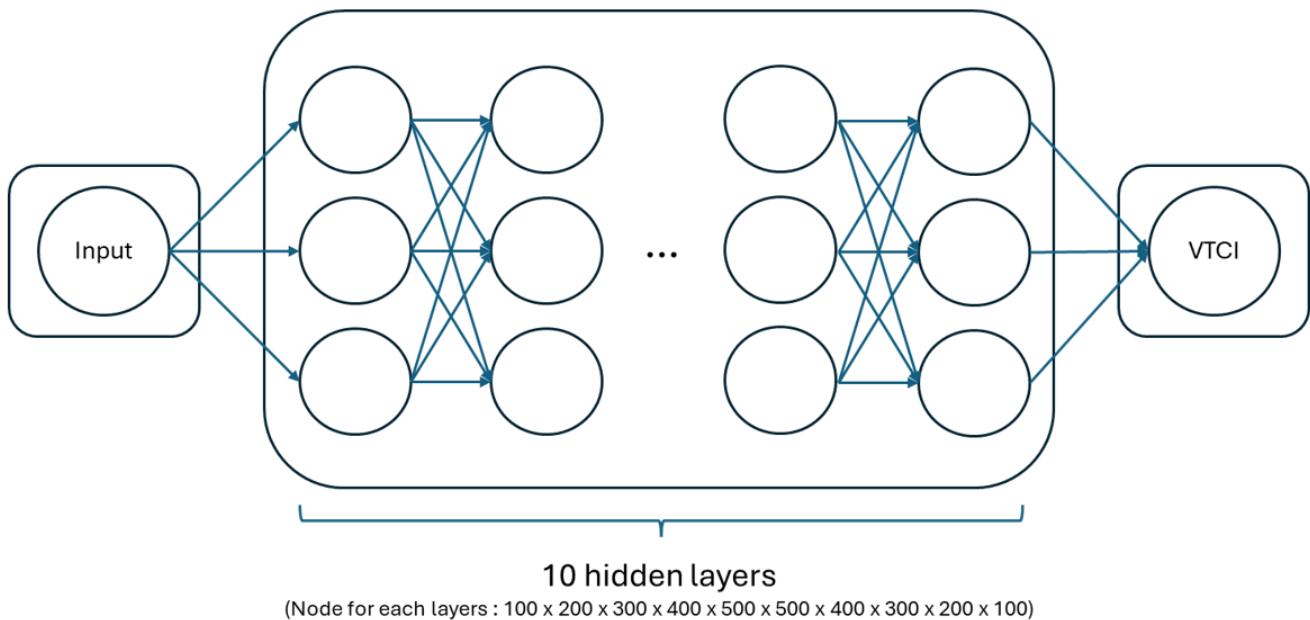
In response to the need for updated and accurate VTCI data, a deep-learning-based model was developed to predict VTCI values using satellite imagery. Due to the limited number of studies that have directly predicted satellite imagery in existing research, methodologies and input variables were selected by referencing studies predicting soil moisture [36–38]. Figure 3 shows the workflow of the VTCI prediction model, which incorporates remote sensing, meteorological, and environmental factors. For the remote sensing factors, NDVI and LST data were utilized, with each satellite index being constructed from the average, maximum, and minimum values of data for the same month from 2003 to 2020. In the case of meteorological factors, average temperature, maximum temperature, minimum temperature, humidity, precipitation, and wind speed were used, along with the FFMC index, which measures the dryness of the surface layer. Additionally, land cover maps and a latitude-longitude coordinate system were utilized.



**Figure 3.** The Workflow of the VTCI Prediction Model.

The model incorporated a DNN (Deep Neural Network) algorithm. DNN, a highly prevalent regression model in machine learning, has proven effective in soil moisture retrieval [39,40]. An illustration in Figure 4, DNN model constructed with 10 hidden layers, where each layer contained nodes between 100 and 500. The model utilized the LeakyReLU

activation function and the Adam optimization algorithm to minimize the Mean Squared Error (MSE) loss function. The performance of the final model was evaluated using two metrics: Mean Absolute Error (MAE) and  $R^2$ . Twelve models were developed to effectively account for monthly weather variations. Each model was trained with a batch size of 1000 and ran for a default of 10 epochs.



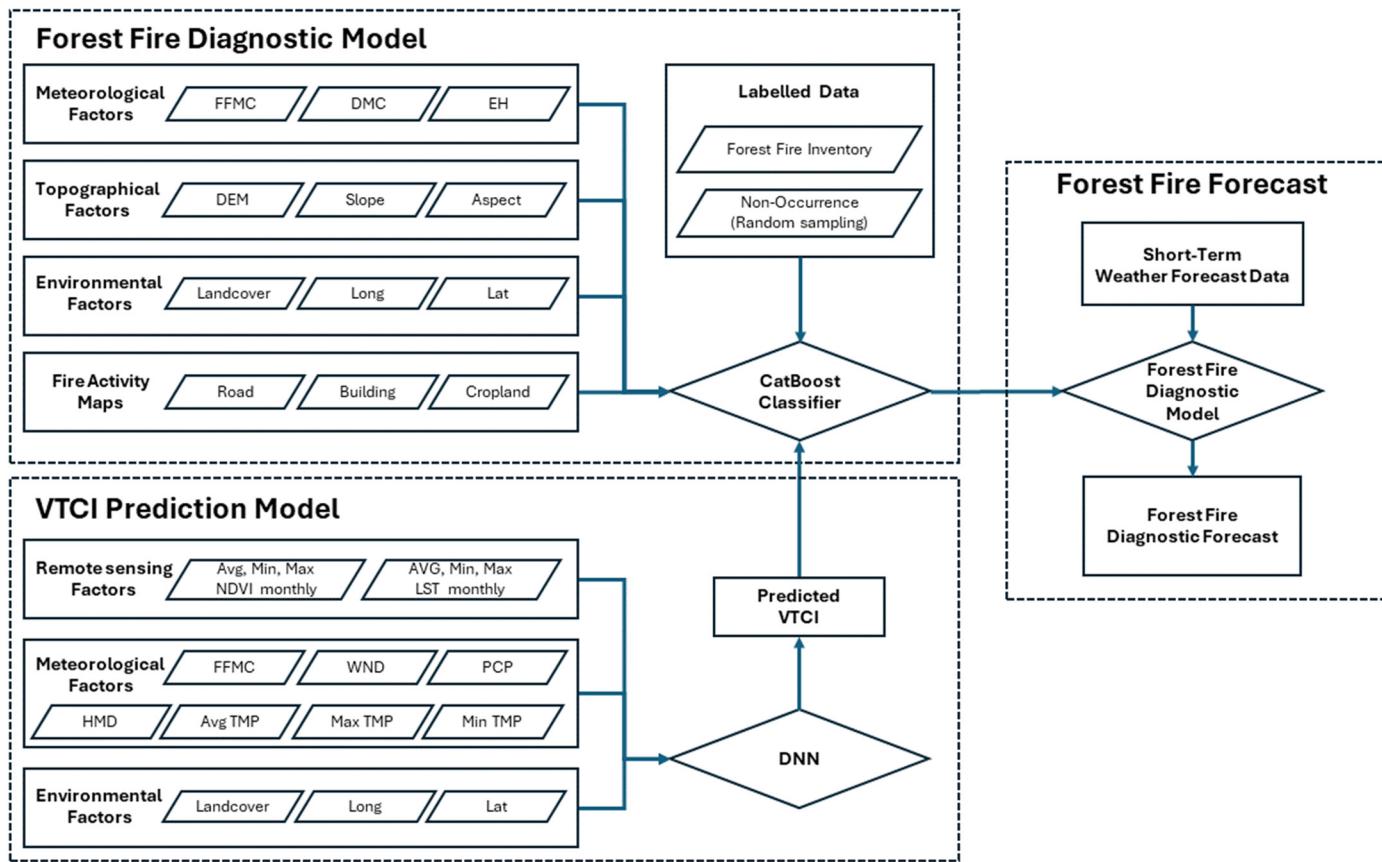
**Figure 4.** Structure of the DNN Algorithm for the VTCI Prediction Model.

#### 2.4. Development of a Forest Fire Diagnostic Model

To develop a forest fire diagnostic model, we utilized the Python PyCaret module, which is an open-source, low-code machine-learning framework that facilitates the rapid deployment of models after data preparation. PyCaret was designed to simplify the development, assessment, comparison, and deployment of machine learning models, making the process as efficient and effective as possible [41,42]. In PyCaret, the performance of the model was evaluated against seven criteria: Accuracy, Area Under the Receiver Operating Characteristic Curve (AUC), Recall, Precision, F1 score, Kappa Value, and Matthews Correlation Coefficient (MCC). For this project, use PyCaret version 3.0.0.

Using PyCaret, we compared various machine learning algorithms to select the best-performing algorithm for the forest fire diagnostic model. CatBoost was selected, which utilizes ordered boosting, a novel approach for managing categorical variables, and automatically handles them without requiring preprocessing such as one-hot or label encoding [43,44]. Although this study does not utilize multiple categorical variables, CatBoost still incorporates advanced techniques, such as the robust handling of missing values, feature importance estimation, and efficient GPU support for faster training [45].

Figure 5 illustrates the comprehensive workflow of the study, which comprises three primary components. The first is the previously detailed VTCI prediction model, which serves as a foundational element. The second is the Forest Fire Diagnostic Model, which utilizes the predicted VTCI as input data. Finally, the wildfire risk forecasting section relies on 3-day short-term weather forecast data sourced from the KMA.



**Figure 5.** Workflow of the Forest Fire Diagnostic Model and Forecast Process.

### 2.5. SHapley Additive exPlanations (SHAP)

The SHAP values were utilized to analyze the impact of each variable in the model. SHAP is a versatile approach used to explain the output of machine learning models. It provides insights on the importance and contribution of each feature in making predictions, offering a broader understanding of the model behavior. SHAP values are based on the Shapley values from cooperative game theory and provide a fair way to allocate credit among features. By examining the SHAP values, users can discern how changes in individual features affect model predictions, aiding model interpretation and debugging. Moreover, SHAP values can be aggregated to explain global model behavior or used to analyze specific predictions on a per-instance basis, thereby enhancing model transparency and trustworthiness [46,47]. The formula for the SHAP value of the variable ( $i$ ) is as follows:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(F - |S| - 1)!}{F!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \quad (4)$$

$\phi_i$  is the SHAP value of the  $i$ -th data point and  $F$  represents the entire dataset.  $S$  denotes the set of all subsets obtained by removing the  $i$ -th data point from the dataset.  $f_{S \cup \{i\}}(x_{S \cup \{i\}})$  is the contribution of the entire set (including the  $i$ -th data point), and  $f_S(x_S)$  is the contribution of the subset obtained by removing the  $i$ -th data point [48].

## 3. Results and Discussion

### 3.1. VTCI Prediction Model Performance

The final VTCI prediction model demonstrated an average Mean Absolute Error (MAE) of 0.0038, Mean Squared Error (MSE) of 0.1103, and an  $R^2$  of 0.86. The best model performance occurred in February, whereas the lowest was in August (Table 3). Performance

gradually declined from February to August, and increased from September onwards. In Republic of Korea, rainfall increases from June to August leading to a greater amount of cloud cover and making satellite image acquisition more challenging. Months with heavy rainfall typically have lower data availability than relatively drier periods, leading to decreased utilization of data and, consequently, lower model performance.

**Table 3.** Performance of the VTCI prediction model.

Metric	January		February		March		April		May		June	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
MSE	0.002	0.002	0.001	0.001	0.002	0.002	0.002	0.002	0.003	0.003	0.004	0.004
MAE	0.033	0.036	0.029	0.029	0.035	0.035	0.037	0.037	0.04	0.04	0.05	0.05
R <sup>2</sup>	0.921	0.899	0.937	0.937	0.913	0.912	0.908	0.907	0.88	0.879	0.818	0.809
Metric	July		August		September		October		November		December	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
MSE	0.005	0.005	0.005	0.008	0.004	0.004	0.004	0.004	0.003	0.003	0.002	0.002
MAE	0.052	0.053	0.053	0.065	0.046	0.049	0.049	0.049	0.426	0.426	0.035	0.035
R <sup>2</sup>	0.834	0.83	0.816	0.715	0.847	0.826	0.865	0.865	0.899	0.898	0.911	0.813

Despite some errors, the performance of the VTCI prediction model was high. Considering that forest fires in Republic of Korea are concentrated between January and early June, it is valid to use the VTCI derived from this model. Furthermore, it allows for the assessment of forest dryness in areas not observed in satellite imagery owing to cloud cover and can provide future results using weather forecast data.

### 3.2. Forest Fire Diagnostic Model Performance

The results of comparing various machine learning algorithms showed that boosting algorithms demonstrated relatively high performance, with the CatBoost model exhibiting the highest performance in terms of Accuracy, AUC, F1 Score, Kappa Value, and MCC. In addition, it demonstrated commendable performance in terms of Recall and Precision.

After selecting the CatBoost algorithm, the model was tuned for 50 iterations based on the kappa criterion. Based on validation data, the final forest fire diagnostic model yielded the following performance metrics: Accuracy of 0.8898, AUC of 0.9541, Recall of 0.8515, Precision of 0.8279, F1 Score of 0.8395, Kappa of 0.7557, and MCC of 0.7558 (Table 4).

**Table 4.** Comparison of Machine Learning Algorithma in PyCaret.

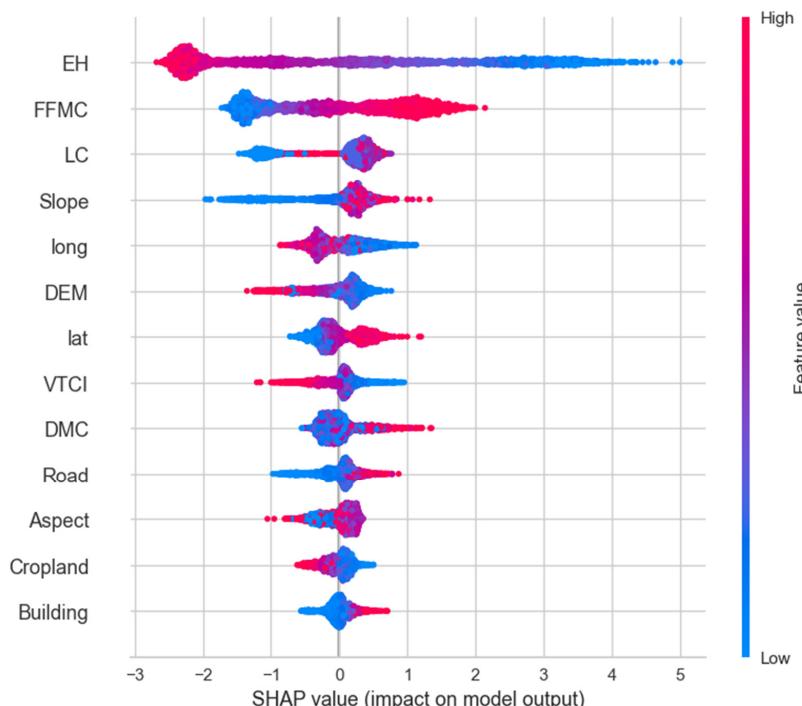
Model	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC
<b>CatBoost Classifier</b>	0.8899	0.9524	0.8423	0.8324	0.8372	0.7541	0.7543
<b>Light Gradient Boosting Machine</b>	0.8875	0.9502	0.8414	0.8272	0.8340	0.7490	0.7493
<b>Random Forest Classifier</b>	0.8844	0.9473	0.8281	0.8280	0.8279	0.7408	0.7410
<b>Extreme Gradient Boosting</b>	0.8840	0.9480	0.8319	0.8248	0.8282	0.7407	0.7408
<b>Gradient Boosting Classifier</b>	0.8839	0.9479	0.8294	0.8258	0.8275	0.7399	0.7401
<b>Extra Tree Classifier</b>	0.8817	0.9460	0.8132	0.8313	0.8220	0.7334	0.7337
<b>Ada Boost Classifier</b>	0.8777	0.9410	0.8155	0.8199	0.8175	0.7256	0.7257
<b>Linear Discriminant Analysis</b>	0.8761	0.9371	0.7993	0.8262	0.8124	0.7199	0.7202
<b>Ridge Classifier</b>	0.8757	0	0.7932	0.8292	0.8107	0.7182	0.7187
<b>Logistic Regression</b>	0.8364	0.9099	0.7494	0.7630	0.7353	0.6325	0.6334

**Table 4.** Cont.

Model	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC
Decision Tree Classifier	0.8267	0.8067	0.7458	0.6354	0.7430	0.6123	0.6124
Naïve Bayes	0.7932	0.9043	0.9023	0.5857	0.7456	0.5801	0.6056
Quadratic Discriminant Analysis	0.7147	0.7738	0.6828	0.8023	0.6001	0.38886	0.4174
K Neighbors Classifier	0.6657	0.6513	0.3976	0.5026	0.4438	0.2095	0.2125
Dummy Classifier	0.6643	0.5	0	0	0	0	0
SVM–Linear Kernel	0.5520	0	0.3586	0.2278	0.2202	0.0094	0.0110

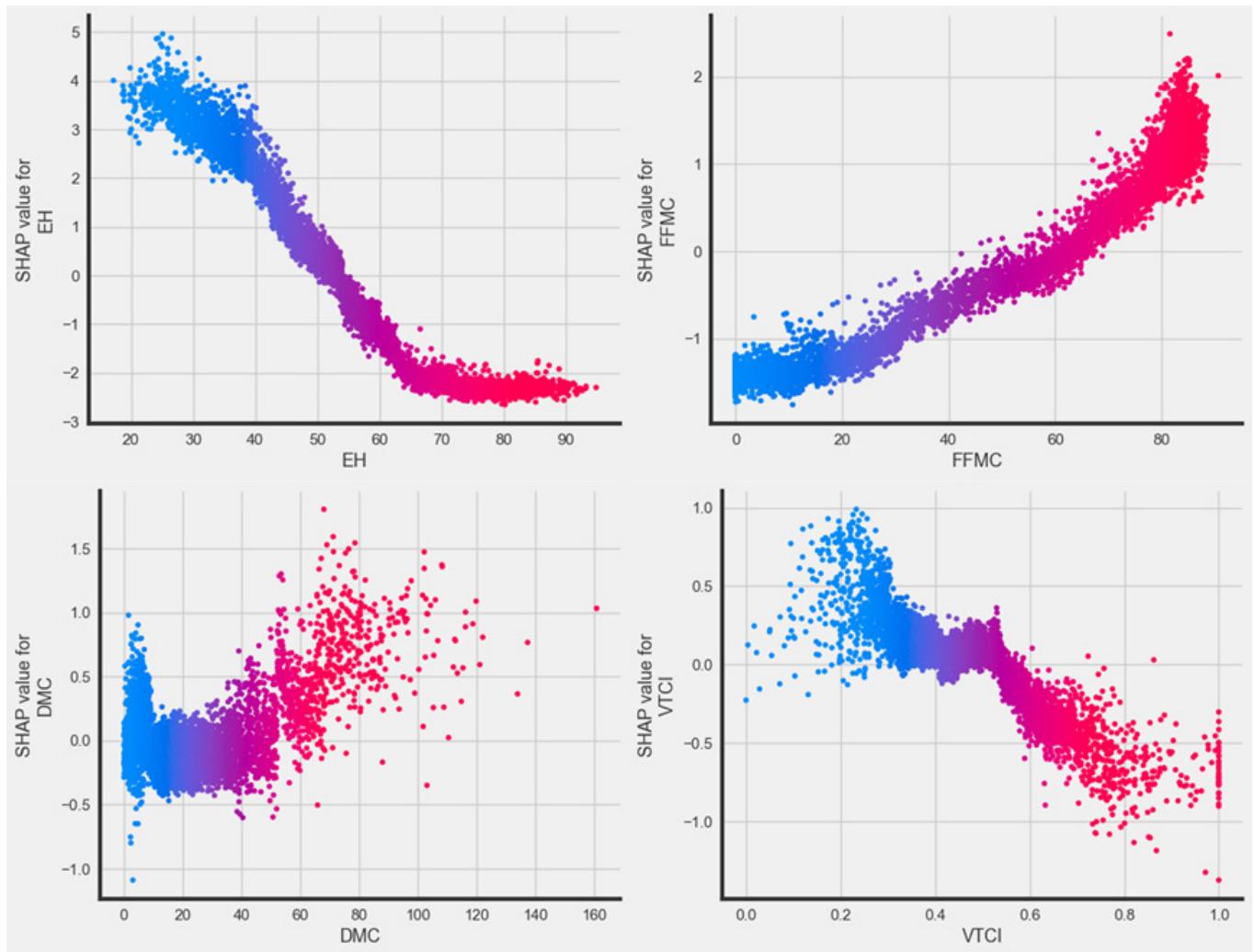
### 3.3. Variable Impact Analysis Based on SHAP Value

Figure 6 presents a SHAP summary plot of the input feature factors derived from the CatBoost classifier. The feature factors were ranked based on their contributions. The X-axis represents the SHAP value, and the y-axis represents the feature factors. Each dot in the plot corresponds to a sample of forest fires from the test dataset, with the color indicating the value of a specific factor. Blue signifies a lower value, whereas red indicates a higher value. The horizontal position of the dot indicates whether the feature factor has a positive or negative influence on prediction [48]. A positive value indicates an impact on the occurrence of wildfires (1), increasing the risk of fire. Conversely, a negative value affects cases where no wildfires occur (0), reducing the fire risk. In the example of EH, the model assesses a higher forest fire risk as the EH value decreases (indicated by blue), whereas a decrease in the EH value (indicated by red) leads to a lower assessment of forest fire risk. Furthermore, the influence of EH can be evaluated by examining the numerical values along the x-axis. Upon analyzing the impact of variables in the forest fire diagnostic model based on the SHAP Summary Plots, EH emerged as the variable with the most significant influence among all variables, followed by FFMC, which also showed a substantial impact. The satellite imagery index VTCI, although relatively less influential on the results, revealed that differences in dryness affect forest fire occurrence.

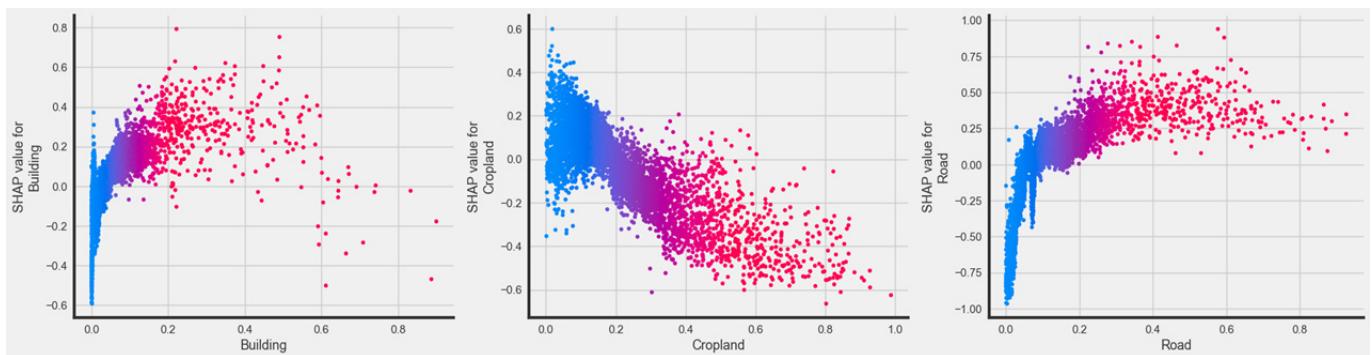


**Figure 6.** SHAP summary plots of the forest fire diagnostic model based on the CatBoost classifier.

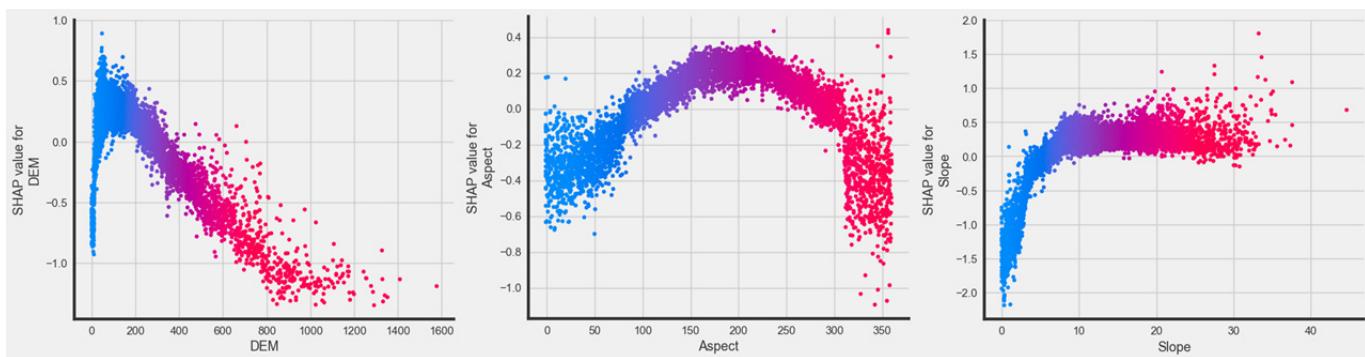
Figures 7–10 show the SHAP-dependence plot for these factors. The SHAP dependence plot identifies the relationship between a single factor (X-axis) and the corresponding SHAP values generated (Y-axis) to evaluate the effect of each feature factor.



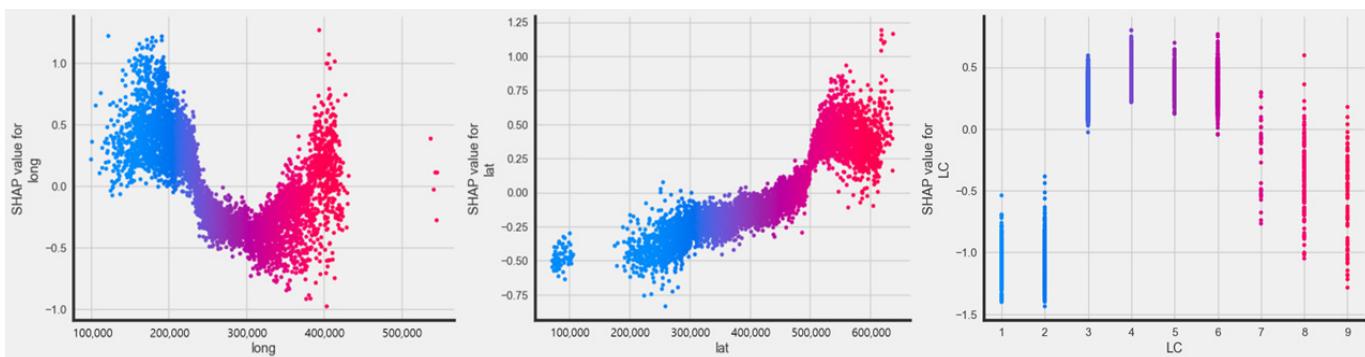
**Figure 7.** SHAP dependence plots for meteorological and satellite image data.



**Figure 8.** SHAP dependence plots for forest fire activity map.



**Figure 9.** SHAP dependence plots for topographic data.



**Figure 10.** SHAP dependence plots for environmental data.

In the graph, the positions of the dots represent the SHAP values corresponding to each variable, indicating their association with the forest fire occurrence. Notably, for EH values below 50, FFMC showed the risk increased significantly when the value exceeded 60. On the contrary, the DMC indicated risks above 60, and even values between 0 and 10 were considered risky, suggesting that forest fires in Korea are heavily influenced by surface dryness rather than prolonged drought, which affects underground dryness in forests.

The VTCI exhibited a distinction in forest fire risk around a threshold of 0.5, indicating a correlation with forest fire occurrence. Despite VTCI's lower influence compared to meteorological data, the predicted VTCI data demonstrated their utility.

In the case of forest fire activity maps, the risk of forest fires increased as the density of buildings and roads increased. However, for buildings, the trend indicated a decrease in forest fire risk at densities above approximately 0.6, whereas the impact of roads on risk levels remained consistent. Unlike buildings and roads, farmland showed a tendency to reduce risk levels as density increased, although a higher risk of forest fire occurrences was observed in the range of approximately 0.0 to 0.3.

Among the topographical factors, a DEM between 0 and 350 m showed a higher forest fire risk, particularly in low-lying areas accessible to humans, reflecting the tendency for forest fires to occur in lowland areas rather than in higher-elevation forests. Aspect angles between 100° and 300°, closer to south-facing, exhibited higher forest fire risks, likely due to higher solar radiation and the resulting evaporation rates in those areas. Slope angles between approximately 5° and 30° were associated with forest fire occurrence, indicating their influence on distinguishing between forested and non-forested land.

Analysis using SHAP values revealed that longitude significantly influenced the risk of forest fire occurrence in Republic of Korea. Areas along the western coast and adjacent inland regions in the south, as well as coastal areas in the east, exhibited higher forest fire risks, whereas the central regions showed relatively lower risks. This pattern reflects the topographical characteristics of Republic of Korea, where forest fires predominantly occur in urban areas adjacent to forests rather than in centrally forested areas. This phenomenon

is also evident from the higher SHAP values observed in regions with higher latitudes, indicating the influence of the Gyeonggi and Gangwon provinces, where forest fires are more frequent [49]. Regarding land cover, coniferous forests (LC 4), grassland (LC 6), mixed forests (LC 5), and deciduous forests (LC 3) had a higher impact on forest fire risk, in descending order, which is consistent with historical forest fire occurrence statistics based on tree species [15,50].

### 3.4. Forest Fire Forecast for Republic of Korea

To conduct short-term forest fire forecasting, an automated process was developed by integrating the short-term weather forecasting process with the forest fire diagnostic model established in this study. This automated forecasting process involves collecting and processing meteorological data to obtain process information, such as the average relative humidity, daily precipitation, average/maximum/minimum temperature, and average wind speed announced at 5 PM the day before the forecast. This information was then used as the input for the model, and the process culminated in the production and mapping of the final forecast results.

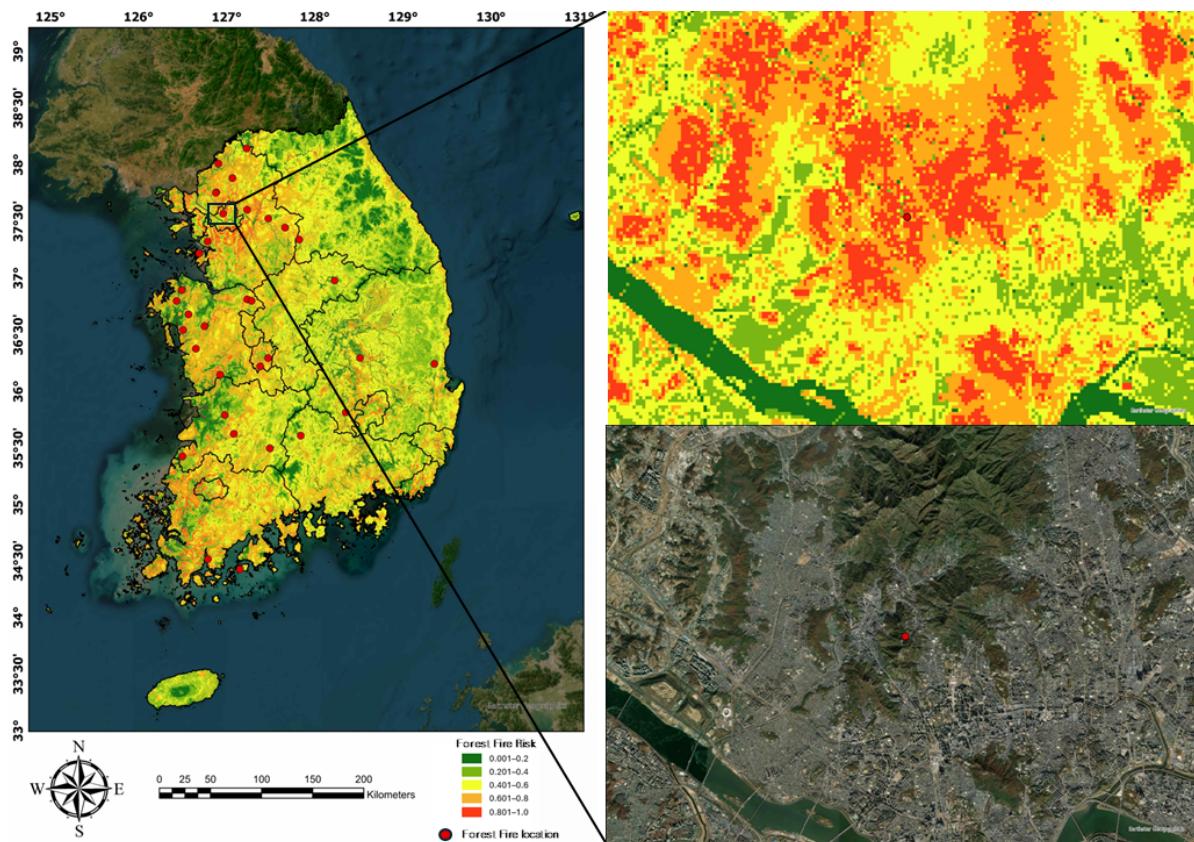
Utilizing short-term weather forecast data provided by the Meteorological Administration at 5 PM on the previous day, the process calculates the forecast results one to three days ahead. A total of 92 short-term forest fire risk forecast maps were generated for the period of 1 March–31 May 2023. Forest fire inventory data was collected from the Korea Forest Service Real-Time Forest Fire Information Platform. The data were compiled by referencing the map locations provided by the platform for verification purposes. During the forecasted period, 362 forest fires occurred.

Based on short-term forecast data for forest fire diagnoses, additional validation was conducted on the forest fire diagnostic model, with 362 forest fire occurrences recorded from March to May 2023. The classification of these forest fire occurrences into the categories in Table 5. showed 165 cases as “Very High”, 99 cases as “High”, 21 cases as “Moderate”, 40 cases as “Low”, and 37 cases as “Very Low”. With “High” and “Very High” cases making up 264 of the total, accounting for 73% of all incidents, it was confirmed that the forest fire diagnostic model significantly and accurately diagnoses actual forest fire risks.

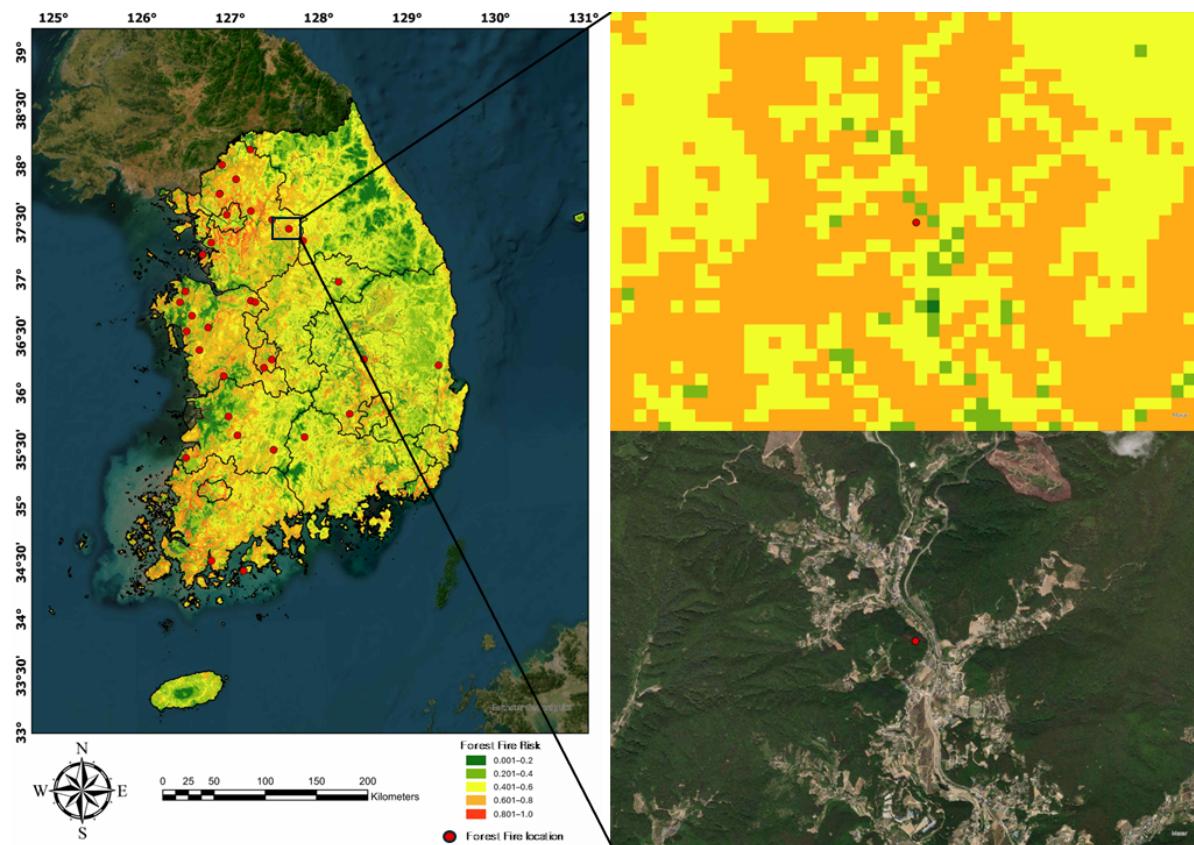
**Table 5.** Forest Fire Forecast Result Verification.

Risk	Risk Level	Number of Forest Fire Occurrences	Percentage
0.8–1.0	Very High	165	46%
0.6–0.8	High	99	27%
0.4–0.6	Moderate	21	6%
0.2–0.4	Low	40	11%
0–0.2	Very Low	37	10%

On 2 April 2023, 34 forest fires occurred, the highest number of forest fires on a single day in 2023. Comparing the forest fire risk map with the actual forest fire occurrence locations, forest fire risks were effectively differentiated between urban (Seoul City) and rural areas (Yangpyeong County). In both cases, forests located near residential areas had a higher fire risk than those located in the central parts of the forests. However, despite the close proximity of the two areas, there is a significant difference in the distribution of forest fire risk levels, which is likely influenced by the distribution of the forest fire activity map. In the forest fire activity map of buildings and roads, areas marked in red, representing high density, coincide with major cities in Korea and show higher forest fire risk levels centered around these urban areas. In particular, Seoul shows a significant difference in forest fire risk levels in forests near residential areas compared to Yangpyeong (Figures 11 and 12).

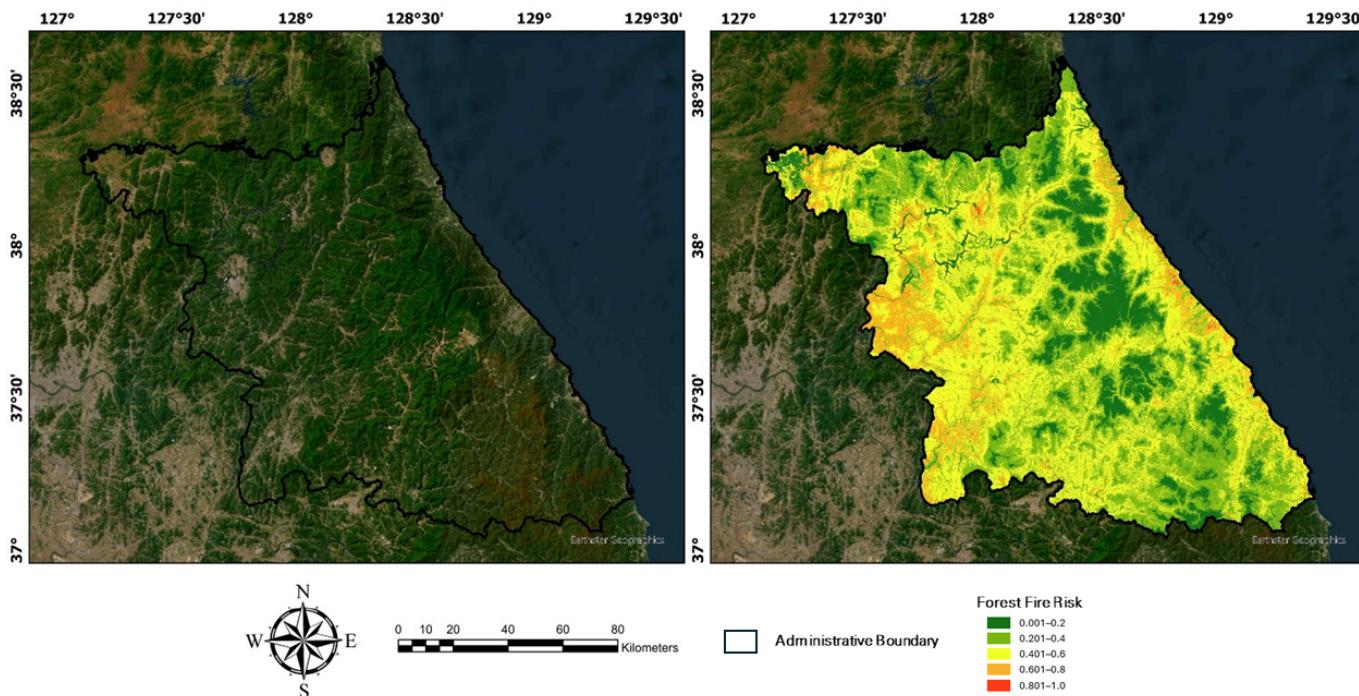


**Figure 11.** Forest fire forecast result for 2 April 2023 in Seoul City.



**Figure 12.** Forest fire forecast result for 2 April 2023 in Yangpyeong County.

On the other hand, Gangwon Province has the highest proportion (82%) of forests in Republic of Korea, and most areas shown in green in the forest fire diagnostic forecast results correspond to forested areas (Figure 13). Historical forest fire records also indicate that, apart from the central forested areas of Gangwon Province, forest fires have primarily occurred near the eastern coastline and western residential areas (Figure 1). A similar trend has been observed in the current distinctions of areas at risk of forest fires.



**Figure 13.** Ortho imagery (**left**) and Forest Fire Diagnostic forecast result (**right**) for 2 April 2023 in Gangwon Province.

#### 4. Conclusions

In this study, we developed a forest fire diagnostic model using land surface condition data (which were VTCI satellite image index) and forest fire activity maps that can diagnose forest fire risks at a resolution of 100 m. One of the key components of the forest fire diagnostic model, the VTCI prediction model, demonstrated its usability with an average performance of R-square 0.83, MSE 0.1103, and MAE 0.0038, thereby validating the model's applicability. The forest fire diagnostic model yielded an accuracy of 0.89. The performance of the forest fire prediction was validated using short-term weather forecasts nationwide from March 2023 to May 2023, showing that 73% of forest fires were included in high-risk areas.

The forecast results of the forest fire diagnostic model showed that areas adjacent to human activity showed a higher risk of forest fire occurrence, demonstrating the potential for simulating anthropogenic fire events. However, since the wildfire activity map is constructed based on density, it appears that urban areas have a higher risk of wildfires compared to small towns or rural areas, which is an aspect that needs to be improved in further research. The results derived from SHAP values indicate that the impact of VTCI is relatively lower compared to meteorological factors. Nevertheless, the dryness indicated by VTCI has shown a correlation with actual forest fire occurrences, highlighting its utility in applying satellite image predictions to diagnose forest fire occurrences. Additionally, by utilizing the VTCI prediction model, it has become possible to address issues related to missing data points caused by clouds or snow, which were limitations previously identified when using satellite images for forest fire risk prediction.

Despite the limitations mentioned earlier, the model developed in this study can diagnose forest fire risks that reflect not only natural environmental factors but also an-

thropogenic influences by incorporating the condition of the land surface. Furthermore, by utilizing the forecast results, it is believed that the model can contribute to more effective forest fire prevention and response activities.

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

**Funding:** This study was carried out with the support of ‘R&D Program for Forest Science Technology (project No. RS-2024-00402509)’ provided by the Korea Forest Service (Korea Forestry Promotion Institute), OJEong Resilience Institute (OJERI) at Korea University under the Basic Science Research Program through the National Research Foundation of Korea (NRF) (NRF-2021R1A6A1A10045235), and the R&D Program for Forest Science Technology (project No. 2021345B10-2323-CD01) provided by the Korea Forest Service (Korea Forestry Promotion Institute).

**Data Availability Statement:** The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

**Acknowledgments:** We would like to express our gratitude to the National Institute of Forest Science (NIFoS) for providing the data and for their valuable advisory support related to this research.

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

## References

1. Liu, Y.; Ding, A. Contrasting Trends of Carbon Emission from Savanna and Boreal Forest Fires during 1999–2022. *Meteorol. Appl.* **2024**, *31*, e2177. [[CrossRef](#)]
2. Bennett, L.T.; Bruce, M.J.; MacHunter, J.; Kohout, M.; Krishnaraj, S.J.; Aponte, C. Assessing Fire Impacts on the Carbon Stability of Fire-Tolerant Forests. *Ecol. Appl.* **2017**, *27*, 2497–2513. [[CrossRef](#)] [[PubMed](#)]
3. Tilman, D.; Reich, P.; Phillips, H.; Menton, M.; Patel, A.; Vos, E.; Peterson, D.; Knops, J. Fire Suppression and Ecosystem Carbon Storage. *Ecology* **2000**, *81*, 2680–2685. [[CrossRef](#)]
4. Halofsky, J.E.; Peterson, D.L.; Harvey, B.J. Changing Wildfire, Changing Forests: The Effects of Climate Change on Fire Regimes and Vegetation in the Pacific Northwest, USA. *Fire Ecol.* **2020**, *16*, 4. [[CrossRef](#)]
5. Abatzoglou, J.T.; Williams, A.P.; Barbero, R. Global Emergence of Anthropogenic Climate Change in Fire Weather Indices. *Geophys. Res. Lett.* **2019**, *46*, 326–336. [[CrossRef](#)]
6. Korea Forest Service. Forest Fire Occurrence Status. Available online: <https://smartcity.go.kr/en/2020/01/22/%EC%82%BC%EB%A6%BC%EC%B2%AD-2020%EB%85%84-%EC%A0%84%EA%B5%AD-%EC%82%BC%88%EB%80%A9%EC%A7%80-%EC%A2%85%ED%95%A9%EB%8C%80%EC%B1%85-%EB%80%9C%ED%91%9C/> (accessed on 13 April 2024).
7. Kang, Y.; Jang, E.; Im, J.; Kwon, C.; Kim, S. Developing a New Hourly Forest Fire Risk Index Based on Catboost in Republic of Korea. *Appl. Sci.* **2020**, *10*, 8213. [[CrossRef](#)]
8. Myoung, B.; Kim, S.H.; Nghiêm, S.V.; Jia, S.; Whitney, K.; Kafatos, M.C. Estimating Live Fuel Moisture from MODIS Satellite Data for Wildfire Danger Assessment in Southern California USA. *Remote Sens.* **2018**, *10*, 87. [[CrossRef](#)]
9. Verbesselt, J.; Jönsson, P.; Lhermitte, S.; Van Aardt, J.; Coppin, P. Evaluating Satellite and Climate Data-Derived Indices as Fire Risk Indicators in Savanna Ecosystems. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1622–1632. [[CrossRef](#)]
10. Kim, S.J.; Lim, C.H.; Kim, G.S.; Lee, J.; Geiger, T.; Rahmati, O.; Son, Y.; Lee, W.K. Multi-Temporal Analysis of Forest Fire Probability Using Socio-Economic and Environmental Variables. *Remote Sens.* **2019**, *11*, 86. [[CrossRef](#)]
11. Azam, M.; Maeng, S.J.; Kim, H.S.; Lee, S.W.; Lee, J.E. Spatial and Temporal Trend Analysis of Precipitation and Drought in Republic of Korea. *Water* **2018**, *10*, 765. [[CrossRef](#)]
12. Chang, H.; Kwon, W.T. Spatial Variations of Summer Precipitation Trends in Republic of Korea, 1973–2005. *Environ. Res. Lett.* **2007**, *2*, 045012. [[CrossRef](#)]
13. Jeon, B.R.; Chae, H.M. The Analysis on Forest Fire Occurrence Characteristics by Regional Area in Korea from 1990 to 2014 Year. *J. For. Environ. Sci.* **2016**, *32*, 149–157. [[CrossRef](#)]
14. Won, M.; Yoon, S.; Jang, K. Developing Korean Forest Fire Occurrence Probability Model Reflecting Climate Change in the Spring of 2000s. *Korean J. Agric. For. Meteorol.* **2016**, *18*, 199–207. [[CrossRef](#)]
15. Kwak, H.; Lee, W.K.; Saborowski, J.; Lee, S.Y.; Won, M.S.; Koo, K.S.; Lee, M.B.; Kim, S.N. Estimating the Spatial Pattern of Human-Caused Forest Fires Using a Generalized Linear Mixed Model with Spatial Autocorrelation in Republic of Korea. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1589–1602. [[CrossRef](#)]
16. Kwak, H.-B.; Lee, W.-K.; Won, M.-S.; Koo, K.-S.; Lee, M.-B.; Lee, S.-C. Spatial and Temporal Pattern of the Human-Caused Forest Fire Occurrences in Korea. In Proceedings of the ESRI User Conference, San Diego, CA, USA, 13–17 July 2009.

17. Korea Forest Service. Causes and Impacts of Forest Fires. Available online: <https://www.globalforestwatch.org/dashboards/country/KOR/?category=fires> (accessed on 16 April 2024).
18. Choung, Y.; Lee, B.-C.; Cho, J.-H.; Lee, K.-S.; Jang, I.-S.; Kim, S.-H.; Hong, S.-K.; Jung, H.-C.; Choung, H.-L. Forest Responses to the Large-Scale East Coast Fires in Korea. *Ecol. Res.* **2004**, *19*, 43–54. [CrossRef]
19. Si, L.; Shu, L.; Wang, M.; Zhao, F.; Chen, F.; Li, W.; Li, W. Study on Forest Fire Danger Prediction in Plateau Mountainous Forest Area. *Nat. Hazards Res.* **2022**, *2*, 25–32. [CrossRef]
20. Piao, Y.; Lee, D.; Park, S.; Kim, H.G.; Jin, Y. Forest Fire Susceptibility Assessment Using Google Earth Engine in Gangwon-Do, Republic of Korea. *Geomat. Nat. Hazards Risk* **2022**, *13*, 432–450. [CrossRef]
21. Patel, N.R.; Mukund, A.; Parida, B.R. Satellite-Derived Vegetation Temperature Condition Index to Infer Root Zone Soil Moisture in Semi-Arid Province of Rajasthan, India. *Geocarto. Int.* **2022**, *37*, 179–195. [CrossRef]
22. Peng, J.; Loew, A.; Zhang, S.; Wang, J.; Niesel, J. Spatial Downscaling of Satellite Soil Moisture Data Using a Vegetation Temperature Condition Index. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 558–566. [CrossRef]
23. Wang, P.-X.; Li, X.-W.; Gong, J.-Y.; Song, C. Vegetation Temperature Condition Index and Its Application for Drought Monitoring. In Proceedings of the IEEE 2001 International Geoscience and Remote Sensing Symposium, Sydney, Australia, 9–13 July 2001.
24. Zhou, X.; Wang, P.; Tansey, K.; Zhang, S.; Li, H.; Wang, L. Developing a Fused Vegetation Temperature Condition Index for Drought Monitoring at Field Scales Using Sentinel-2 and MODIS Imagery. *Comput. Electron. Agric.* **2020**, *168*, 105144. [CrossRef]
25. Won, M.S.; Miah, D.; Koo, K.S.; Lee, M.B.; Shin, M.Y. Meteorological Determinants of Forest Fire Occurrence in the Fall, Republic of Korea. *J. Korean Soc. For. Sci.* **2010**, *99*, 163–171.
26. Kang, S.-C. 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. [CrossRef]
27. Van Nest, T.A.; Alexander, M.E. Systems for Rating Fire Danger and Predicting Fire Behavior Used in Canada. In Proceedings of the National Interagency Fire Behavior Workshop, Phoenix, AZ, USA, 1–5 March 1999.
28. Van Wagner, C.E. *Structure of the Canadian Forest Fire Weather Index*; Environment Canada, Forestry Service: Ottawa, ON, Canada, 1974.
29. Jo, H.W.; Krasovskiy, A.; Hong, M.; Corning, S.; Kim, W.; Kraxner, F.; Lee, W.K. Modeling Historical and Future Forest Fires in Republic of Korea: The FLAM Optimization Approach. *Remote Sens.* **2023**, *15*, 1446. [CrossRef]
30. Cawson, J.G.; Nyman, P.; Schunk, C.; Sheridan, G.J.; Duff, T.J.; Gibos, K.; Bovill, W.D.; Conedera, M.; Pezzatti, G.B.; Menzel, A. Estimation of Surface Dead Fine Fuel Moisture Using Automated Fuel Moisture Sticks across a Range of Forests Worldwide. *Int. J. Wildland Fire* **2020**, *29*, 548–559. [CrossRef]
31. Fan, J.; Hu, T.; Ren, J.; Liu, Q.; Sun, L. A Comparison of Five Models in Predicting Surface Dead Fine Fuel Moisture Content of Typical Forests in Northeast China. *Front. For. Glob. Chang.* **2023**, *6*, 1122087. [CrossRef]
32. Kang, Z.; Quan, X.; Lai, G. Assessing the Effects of Fuel Moisture Content on the 2018 Megafires in California. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 868–877. [CrossRef]
33. Ruffault, J.; Martin-StPaul, N.; Pimont, F.; Dupuy, J.L. How Well Do Meteorological Drought Indices Predict Live Fuel Moisture Content (LFMC)? An Assessment for Wildfire Research and Operations in Mediterranean Ecosystems. *Agric. For. Meteorol.* **2018**, *262*, 391–401. [CrossRef]
34. Jung, J.; Kim, C.; Jayakumar, S.; Kim, S.; Han, S.; Kim, D.H.; Heo, J. Forest Fire Risk Mapping of Kolli Hills, India, Considering Subjectivity and Inconsistency Issues. *Nat. Hazards* **2013**, *65*, 2129–2146. [CrossRef]
35. Pourtaghi, Z.S.; Pourghasemi, H.R.; Rossi, M. Forest Fire Susceptibility Mapping in the Minudasht Forests, Golestan Province, Iran. *Environ. Earth Sci.* **2015**, *73*, 1515–1533. [CrossRef]
36. Xu, M.; Yao, N.; Yang, H.; Xu, J.; Hu, A.; Gustavo Goncalves de Goncalves, L.; Liu, G. Downscaling SMAP Soil Moisture Using a Wide & Deep Learning Method over the Continental United States. *J. Hydrol.* **2022**, *609*, 127784. [CrossRef]
37. Gao, L.; Gao, Q.; Zhang, H.; Li, X.; Chaubell, M.J.; Ebtehaj, A.; Shen, L.; Wigneron, J.P. A Deep Neural Network Based SMAP Soil Moisture Product. *Remote Sens. Environ.* **2022**, *277*, 113059. [CrossRef]
38. Achieng, K.O. Modelling of Soil Moisture Retention Curve Using Machine Learning Techniques: Artificial and Deep Neural Networks vs Support Vector Regression Models. *Comput. Geosci.* **2019**, *133*, 104320. [CrossRef]
39. Rodríguez-Fernández, N.J.; Aires, F.; Richaume, P.; Kerr, Y.H.; Prigent, C.; Kolassa, J.; Cabot, F.; Jiménez, C.; Mahmoodi, A.; Drusch, M. Soil Moisture Retrieval Using Neural Networks: Application to SMOS. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5991–6007. [CrossRef]
40. Albergel, C.; de Rosnay, P.; Gruhier, C.; Muñoz-Sabater, J.; Hasenauer, S.; Isaksen, L.; Kerr, Y.; Wagner, W. Evaluation of Remotely Sensed and Modelled Soil Moisture Products Using Global Ground-Based In Situ Observations. *Remote Sens. Environ.* **2012**, *118*, 215–226. [CrossRef]
41. Whig, P.; Gupta, K.; Jiwani, N.; Jupalle, H.; Kousser, S.; Alam, N. A Novel Method for Diabetes Classification and Prediction with PyCaret. *Microsyst. Technol.* **2023**, *29*, 1479–1487. [CrossRef]
42. Sarangpure, N.; Dhamde, V.; Roge, A.; Doye, J.; Patle, S.; Tamboli, S. Automating the Machine Learning Process Using PyCaret and Streamlit. In Proceedings of the 2023 2nd International Conference for Innovation in Technology, INOCON 2023, Bangalore, India, 3–5 March 2023; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2023.
43. Hancock, J.T.; Khoshgoftaar, T.M. CatBoost for Big Data: An Interdisciplinary Review. *J. Big Data* **2020**, *7*, 94. [CrossRef]
44. Ding, Y.; Wang, M.; Fu, Y.; Zhang, L.; Wang, X. A Wildfire Detection Algorithm Based on the Dynamic Brightness Temperature Threshold. *Forests* **2023**, *14*, 477. [CrossRef]

45. Adler, A.I.; Painsky, A. Feature Importance in Gradient Boosting Trees with Cross-Validation Feature Selection. *Entropy* **2022**, *24*, 687. [[CrossRef](#)]
46. Wang, D.; Thunell, S.; Lindberg, U.; Jiang, L.; Trygg, J.; Tysklind, M. Towards Better Process Management in Wastewater Treatment Plants: Process Analytics Based on SHAP Values for Tree-Based Machine Learning Methods. *J. Environ. Manag.* **2022**, *301*, 113941. [[CrossRef](#)]
47. Abdollahi, A.; Pradhan, B. Explainable Artificial Intelligence (XAI) for Interpreting the Contributing Factors Feed into the Wildfire Susceptibility Prediction Model. *Sci. Total Environ.* **2023**, *879*, 163004. [[CrossRef](#)]
48. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 4765–4774.
49. Korea Forest Service. The Forest Fire Occurrences by Region over the Past Decade. Available online: <https://www.forest.go.kr/kfsweb/kfi/kfs/frfr/selectFrfrStatsArea.do> (accessed on 14 April 2024).
50. Lee, S.-Y.; Won, M.-S.; Han, S.-Y. Developing of Forest Fire Occurrence Danger Index Using Fuel and Topographical Characteristics on the Condition of Ignition Point in Korea. *Fire Sci. Eng.* **2005**, *19*, 75–79.

**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.