



Article Wildfire Susceptibility Mapping in Baikal Natural Territory Using Random Forest

Olga Nikolaychuk 🗅, Julia Pestova 🕒 and Aleksandr Yurin *🕩

Matrosov Institute for System Dynamics and Control Theory, Siberian Branch of Russian Academy of Sciences (ISDCT SB RAS), Irkutsk 664033, Russia

* Correspondence: iskander@icc.ru

Abstract: Wildfires are a significant problem in Irkutsk Oblast. They are caused by climate change, thunderstorms, and human factors. In this study, we use the Random Forest machine learning method to map the wildfire susceptibility of Irkutsk Oblast based on data from remote sensing, meteorology, government forestry authorities, and emergency situations. The main contributions of the paper are the following: an improved domain model that describes information about weather conditions, vegetation type, and infrastructure of the region in the context of the possible risk of wildfires; a database of wildfires in Irkutsk Oblast from 2017 to 2020; the results of an analysis of factors that cause wildfires and risk assessment based on Random Forest in the form of fire hazard mapping. In this paper, we collected and visualized data on wildfires and factors influencing their occurrence: meteorological, topographic, characteristics of vegetation, and human activity (social factors). Data sets describing two classes, "fire" and "no fire", were generated. We introduced a classification according to which the probability of a wildfire in each specific cell of the territory can be determined and a wildfire risk map built. The use of the Random Forest method allowed us to achieve the following risk assessment accuracy indicators: accuracy—0.89, F1-score—0.88, and AUC-0.96. The comparison of the results with earlier ones obtained using case-based reasoning revealed that the application of the case-based approach can be considered the initial stage for deeper investigations with the use of Random Forest for more accurate forecasting.

Keywords: hazard of wildfires; wildfire; forest quarters; wildfire susceptibility mapping; random forest; data analysis; Baikal natural territory; Irkutsk oblast

1. Introduction

Wildfires remain a serious problem throughout the world [1–6]; they negatively affect biodiversity and air, soil, and water quality [7–9], lead to ecosystem degradation [10], and pose a threat to the safety of people and infrastructure [11,12]. Wildfires can be caused by climatic conditions, careless handling of fire by the local population, or other factors that depend on the characteristics of the regions [13–15]. In Russia, direct economic loss attributed to wildfires includes a reduction in forestry and timber on 800 thousand hectares, which is about 0.06% of GDP per year [16]. Irkutsk Oblast is a region with one of the highest forest cover rates (78%) among the constituent entities of the Russian Federation. More than 90% of the total forested area is occupied by fire-hazardous coniferous plantations [13]. In 2022, 1840 wildfires occurred in the region; the area covered by fire was 1390.9 thousand hectares, and economic loss was estimated at RUB 780.3 million.

A feature of Irkutsk Oblast is that it is a very large territory with a low population density that is mainly concentrated along rivers, roads and railways. For the timely detection of forest fires in this region, there is an aviation and ground monitoring system. Special attention is paid to preventive measures in preparation for the fire season. These activities include the construction and reconstruction of forest roads, the arrangement of fire-fighting glades, and fire-fighting mineralized strips. The entire scope of the activities performed



Citation: Nikolaychuk, O.; Pestova, J.; Yurin, A. Wildfire Susceptibility Mapping in Baikal Natural Territory Using Random Forest. *Forests* **2024**, *15*, 170. https://doi.org/10.3390/ f15010170

Received: 2 December 2023 Revised: 9 January 2024 Accepted: 11 January 2024 Published: 13 January 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/). requires significant material costs and theoretical (scientific) substantiation. Thus, the task of wildfire research is urgent to increase the efficiency of monitoring in such a large area. Wildfire research is carried out in various directions: (i) fuel characterization, fire detection, and mapping; (ii) fire weather and climate change; (iii) fire occurrence, susceptibility, and risk; (iv) fire behavior prediction; (v) fire effects; and (vi) fire management [14].

Various methods and tools are used to perform the task of risk prediction: decisionmaking methods [15], fuzzy systems [16–21], case-based reasoning [22–24], and methods based on machine learning [8,14,25], which are considered by some researchers to be the most promising.

Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data, generalize to unseen data, and thus perform tasks without explicit instructions. The following ML methods are defined [8,14,25–39]: artificial neural networks (ANNs), decision trees, supportvector machines, regression analysis, Bayesian networks, genetic algorithms, Random Forest, etc. ML methods are used to perform a wide variety of tasks [40–44], including forecasting the risk of forest fires [8,14,25,34–39].

It is noted that the use of ML methods for predicting the risk of wildfires should take into account the physico-geographical characteristics of the territories (geographical location, relief, climatic zones, natural reservoirs, types of human activity, etc.) for which they are used. In addition, a comparison of various methods has shown that machine learning methods provide the highest prediction accuracy in the presence of large and reliable data sets.

At the moment, for the territory of Irkutsk Oblast, studies are conducted using the following techniques: analysis of statistical data on wildfires [13,45], analysis of the effectiveness of various methods for assessing meteorological factors of fire hazards in the forests of the Southern Baikal area [46,47], mapping the landscapes of the Western Baikal area [9], predicting the risk of wildfires for individual areas based on case-based reasoning [22]. However, mapping the fire hazard for the entire province, taking into account its characteristics, as well as the creation of an automated technology that implements the solution to this problem for any forecast period, has not yet been fully completed.

This task is being performed in the Matrosov Institute for System Dynamics and Control Theory, Siberian Branch of Russian Academy of Sciences (ISDCT SB RAS), and consists of several stages. The first stage was devoted to the collection and preparation of data on the territory and the selection of data analysis methods. The second stage included the tasks (steps) of applying the previously identified methods to various territorial entities, for example, forest districts, national parks or territories adjacent to settlements, and transport arteries (road and rail). At the first step of the second stage, the task of forecasting and assessing the risk of forest fires was performed on the basis of case-based reasoning [22]. In this work, one of the steps of the second stage is considered. We use the Random Forest method to map the fire hazard (predict the wildfire risk) for the territory under consideration. The results of comparing this method with other machine learning methods and its effectiveness for solving the problem are highlighted in [32,34–39]. It is believed that this method does not require large costs to justify the parameters of the model; it allows you to evaluate the contribution of each variable of the model to the overall classification result. Next, it is planned to justify the choice of neural network architecture, train it, and perform data classification based on it. The next stage of the study will be to substantiate effective methods and model parameters for each territorial entity, in particular, forestry operations and specially protected areas.

Thus, our study addresses wildfire susceptibility mapping or similar definitions of risk in Irkutsk Oblast based on data from remote sensing, meteorology, government forestry authorities, and emergency situations. Mapping the fire hazard of an area is the basis for effective monitoring and predicting the development of wildfires, planning resources, and making informed decisions to reduce risks and mitigate their consequences [15,48–50].

Our main contributions that determine the novelty of the present work are the following:

- An extended (improved) information domain model that describes information about weather conditions, vegetation type, and infrastructure of the region in the context of the possible risk of wildfires;
- An extended database that contains information about wildfires in Irkutsk Oblast from 2017 to 2020, complemented by new data on the type of vegetation (ground surface), topography, and thunderstorms. Thunderstorm data were obtained from weather station data;
- The results of an analysis of factors that cause wildfires and risk assessment of the territory of Irkutsk Oblast based on machine learning methods, in particular, Random Forest, in the form of fire hazard mapping;
- Results of assessing the effectiveness of predicting the risk of wildfires based on the Random Forest method.

For Irkutsk Oblast, the study of forest fire forecasting based on Random Forest was conducted for the first time.

The paper is organized as follows: Section 2 briefly describes the background, while Section 3 presents our results. Section 4 contains discussion and concluding remarks.

2. Materials and Methods

2.1. Background

2.1.1. Wildfire Susceptibility Mapping

The fire hazard mapping approach consists of creating a spatial model of fire hazards using remote sensing data or data from various agencies monitoring and managing fire management activities based on data on landscape, climate, and human factors.

Currently, machine learning (ML) methods are most often used to perform this task. They demonstrate greater accuracy compared to other methods, for example, logistic regression. Researchers usually either use one of the ML methods, or compare the results of several ML methods, or use ensembles of ML methods [14]. In particular, it is common to use neural networks [26–33] to build fire hazard maps for the territories of Portugal [26], Spain [27], Iran [29,30], Vietnam [31], and China [32]. The Naive Bayes method [3,51] is used for the territory of Iran, neuro-fuzzy systems are employed for the forests of Chile [17], Brazil [18], Vietnam [20] and Iran [21], the Random Forest method is applied to Mediterranean Europe [34], Ethiopia [36] and China [32,35,39], and GIS-based multicriteria decision analysis methods (GIS-based multi-criteria decision analysis (MCDA)) [52], analytical hierarchical process (AHP) [19,48,53,54] and case-based reasoning [22–24] are used for this purpose, too.

It should be noted that works [38,39,55,56] compare ML methods and state that the Random Forest method is superior in accuracy to neural networks and the support vector machine (SVM). In [36,57], the authors highlight the effectiveness of ensemble methods. The result of these methods and models is that fire hazard mapping and wildfire risk assessment are based on various factors.

The following groups of factors influence the occurrence of wildfires [14,25,58]: meteorological factors, topographic factors, characteristics of vegetation, and human activity (social factors). The group of meteorological factors includes the topographic wetness index, average annual temperature, average annual precipitation, air temperature (average daily and maximum), dates of transition of average daily temperatures through threshold values, dates of onset and disappearance of stable snow cover, relative humidity (average daily and minimum), lack of air humidity, number of days with relative humidity \leq 30% in one of the observation periods for a certain period, annual precipitation period, number of days with rain, dryness index, weather regime, number of days with thunderstorms, etc. The set of topographic factors includes altitude, terrain slope, etc. Social factors are the following: distance from urban areas, land use, distance from roads, population density, GDP per capita, etc. It is common to use various weather indices, e.g., the Canadian Forest Fire Weather Indices System (CFFWIS) [59]. The application of the factors described should take into account the individual properties of the initial data and the study areas [60].

In [22], we forecast the risk of wildfires in certain areas of Irkutsk Oblast (Bodaibinsky and Kazachinsko-Lensky Districts) based on case-based reasoning. This method was chosen as the most effective for preliminary analysis of the obtained data set. The wildfire risk forecasting accuracy score was 0.874.

The purpose of this study is to complement the model with new variables and increase the description accuracy of the type of vegetation, the topography of the territory, and thunderstorm activity. We use the Random Forest method for data characterizing the entire territory of Irkutsk Oblast and compare the results obtained with the results of previous studies.

2.1.2. Related Works: Using Random Forest for Wildfire Susceptibility Mapping

Random Forest (RF) is a machine learning algorithm based on the use of an ensemble of decision trees, each of which has a very low classification quality. The algorithm is applied to classification, regression, and clustering tasks [61].

This method is actively employed to perform the task of fire hazard mapping [32,34–39,62] (Table 1).

Territory	Accuracy
Yunnan Province (China) [32]	84.36
European Mediterranean region [34]	96.3
Global ecosystem [35]	78.33
China's boreal forest, located in the Daxing'an Mountains of Northeastern China [37]	70.1
Yunnan Province (China) [38]	88.3
Hunan Province (China) [39]	91.68
Ethiopia [62]	67.2

Table 1. Examples of the accuracy score using the Random Forest method.

These works demonstrated that RF has high predictive accuracy and high robustness to outliers and "noise" [63]. These factors conditioned its use for wildfire research [34].

The mentioned works differ in the period of consideration of fire data from 8 to 25 years, the volume, and the content of variables that are taken into account in the created models. The authors of these papers note that the assessment of the importance of the model factors varies depending on the part of the territory under consideration, since climatic, environmental, and social factors differ from region to region. The same factors can affect estimates in different ways, depending on the location and scale of the analysis. For example, in Pakistan, precipitation, soil moisture, unemployment rate, livestock density, and density of local roads are important [8]; in the European Mediterranean region, precipitation, soil moisture, road density, vegetation type [34]; in the Daxing'an Mountains, forest type, and distances to railways. However, gross national product, unemployment, and population density did not play a decisive role in the occurrence of fires [37]. The work of [38] differs in the volume of factors under consideration; the model contains 42 variables, where climatic factors describing average, maximum, and minimum values for a 10-day period with data gradation for 6 and 24 h are taken into account. The identified main factors that influenced Yunnan wildfire occurrence were forest coverage ratio, month, season, surface roughness, 10-day minimum of the 6 h maximum humidity, and 10-day maxima of the 6 h average and maximum temperatures. Interesting conclusions were obtained in [62] on the weak influence of climatic factors on the level of fire danger compared with human influence on the environment.

Thus, the analysis of the work allows us to conclude that the value of each solution consists not only in obtaining the results of the fire risk assessment but also in identifying the causes of forest fires characteristic of the territories under consideration.

Having summarized the results of using RF, we can formulate the following stages of implementing the method for the task at hand:

- (1) Selection (justification) of factors influencing the occurrence of a wildfire.
- (2) Collection of data on fires and factors for a territory (class "fire").
- (3) Data generation to form the "absence of fire" class, taking into account spatial and temporal criteria.
- (4) Selection of parameters for the RF method (using recommendations from the literature).
- (5) Application of the RF method. It is possible to change the method parameters.
- (6) Analysis of the accuracy of the problem solution.
- (7) Risk assessment.
- (8) Visualization of the results of applying the method.
- (9) Analysis of the results of applying the method depending on the seasonality and territoriality of the fire.

2.2. Wildfire Susceptibility Mapping Using Random Forest

2.2.1. The Study Area

The study area is Irkutsk Oblast, located in Eastern Siberia between 51 and 65 parallels of north latitude in the southeast of the Central Siberian Plateau along Lake Baikal. The area of the region is 774,846 km² (4.52% of the territory of Russia), the population is 2,357,134 people (2022), and the population density is 3.04 people per km² (2022). The territory is divided into taiga, forest-steppe, and South Siberian mountain zones. The lands of the Forest Fund include 37 forest districts, 2 nature reserves, a national park, and about a dozen reserves. The forest cover of the region is 82%, which is one of the highest indicators in Russia. Coniferous species predominate in 76% of the forest area, while soft-leaved species account for 19% [64].

The territory of Irkutsk Oblast is a region with a high level of fire hazards and high growth rates of wildfires, against the backdrop of increasing atmospheric factors of climate aridity in recent decades [64]. According to research data from 2001–2020 in Eastern Siberia, a positive trend in air temperature (0.11 °C year⁻¹) and a negative trend in precipitation (–1.64 mm year⁻¹) were observed in June [65]. The wildfire season in the region lasts, on average, 174 days.

In recent years, there has been an increase in the area of wildfires in the region (Figure 1) [15]. In Russia, 70%–90% of all forest areas burned by fire are recorded in Siberia, where the bulk of Russia's boreal forests is concentrated, which plays an important role in the absorption and sequestration of carbon [66].

Wildfires significantly changed the taiga forests of Central Siberia [64], leading to a transformation of the landscapes of the Lake Baikal drainage basin and a decrease in forest reserves [67]. Over the course of a number of years, during wildfires in the region, distinctive signs of deterioration in the health of people in the smoky zone were detected. This demonstrates an increase in the values of the following indicators: the number of requests for medical help in connection with respiratory diseases by 6.5%; the number of exacerbations of chronic bronchitis by 4.2%; and bronchial asthma by 5.2% [66].

The largest areas of wildfires on average over 10 years have been recorded in the Katangsky, Kirensky, Mamsko-Chuysky, Bodaibinsky, Bokhansky, Ust-Kutsky, and Chunsky districts [67]. This fact is confirmed by the statistics from recent years (Figure 2).



Figure 1. Statistics of wildfires in Irkutsk Oblast [45]: (a) number of forest fires; (b) fire damage; (c) area of fires.

Researchers highlight that significant forest cover, complex orography, and differences in climatic, plant, and socio-economic indicators of mountain, foothill, and lowland areas have a complex impact on the formation of conditions on which the frequency, intensity, and prevalence of forest fires in Irkutsk Oblast depend. It was revealed that meteorological conditions favorable for the occurrence and spread of wildfires are most often observed in May–June and August–September. The areas most prone to wildfires due to weather and climatic factors are the northern regions and the coast of Lake Baikal, where the environmental protection zones are located [68].

Representatives of the Ministry of Emergency Situations point out that 70%–80% of wildfires occur due to the "human factor" [69], because the accessibility of forest areas is quite high both for loggers and for the population as a whole. The second important cause of wildfires is lightning discharges (on average, 29%) that result in large-scale forest burns in hard-to-reach areas with no ground protection [70].

This work will examine and take into account the main factors describing the causes of wildfires in the entire territory of Irkutsk Oblast.



Figure 2. Distribution of some wildfires across municipalities of Irkutsk Oblast for the period March–November 2017–2020.

2.2.2. Data

A digital terrain map at a scale of 1:200,000 is used as an electronic topographic basis. The main thematic layers are vector layers obtained from various institutions.

The study uses data on 45,000 thermal hotspots identified by analyzing satellite imagery from 2017 to 2020. This information is used as input. A thermal point is a significant (when compared to neighboring points) increase in temperature on the Earth's surface, recorded at the time of the satellite's passage. The thermal point is recorded by the satellite in the form of polygonal type objects (a set of quadrangles).

Information about thermal points is generated using software and hardware, which includes a satellite telemetry receiving unit. The complex consists of a data receiving station from NOAA satellites of the Alisa-SC series produced by Scanex LLC (Moscow, Russia)

and specialized software tools for receiving satellite telemetry. At the moment, the satellite complex receives and decrypts data from the AVHRR device of the NOAA18 and NOAA19 satellites. The visibility zone of the Alisa–SK satellite complex extends from the Urals in the west to the Far East in the east of the Russian Federation, which ensures monitoring of fire hazardous conditions throughout Irkutsk Oblast.

The first stage of the monitoring system is the direct reception of data from the NOAA satellite through the Alisa–SK satellite complex. Afterwards, several stages of data processing are carried out (demodulation, synchronization, decoding, etc.). At the final stage, the FirePro program (Institute of Solar-Terrestrial Physics of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia) [71] is used to perform the analysis and identify thermal points (Figure 3). The data obtained automatically are further corrected by the operator to reduce possible errors in the algorithm and are stored in the database that provided the information for this study.



Figure 3. A map of Irkutsk Oblast displaying thermal points for 2017–2020.

The data on thermal points contain information about the date, area, and time of their discovery.

A thermal point can indicate not only a wildfire but also a garbage fire, a man-made process, or a man-made fire. For this reason, the data set is cleaned based on the data on the boundaries of settlements and man-made objects.

Thermal points can be recorded 3–4 times a day, so it is necessary to process and merge information about thermal points characterizing the same fire, taking into account time and coordinate parameters. Further processing uses data on the first (in terms of time) thermal point obtained by the merging algorithm. This fusion resulted in a dataset of 9001 wildfire records for the period March–November 2017–2020. Data on thermal points were obtained from the Institute of Solar and Earth Physics SB RAS [72] (Table 2).

Index	Initial Date	Day Number	Initial Day	Initial Decade	Month	Year	Lat	Lon
0	9 August 2017	221	221	22	8	2017	58.29611	104.35694
1	12 August 2020	222	1320	22	8	2020	57.87000	108.04694
2	5 May 2020	123	1221	12	5	2020	56.17500	97.71500
3	25 September 2018	268	633	26	9	2018	57.87389	101.20306
4	30 June 2018	181	546	18	6	2018	58.89611	116.01306

Table 2. The data on thermal points.

In Irkutsk Oblast, each forestry operation is divided into forest areas and blocks. Data on forestry and forest blocks were obtained from the state forest inventory materials.

In studies on wildfires, the following groups of factors influencing the occurrence of fires are defined: meteorological, topographic, characteristics of vegetation, and human activity (social factors) [14,25,58].

The significance of meteorological factors in the occurrence and development of wildfires is well known; therefore, in all works, they are assessed as fire hazard factors, and this assessment can be expressed in the form of special indicators. In this study, meteorological data were obtained from the Federal State Budgetary Institution "Irkutsk Department of Hydrometeorology and Environmental Protection" [73]. The selected average daily indicators describe air temperature, atmospheric pressure, relative humidity, wind direction (points), wind speed (determined on the 12-point F. Beaufort scale: calm (0–0.2), quiet (0.3–1.5), light (1.6–3.3), light (3.4–5.4), moderate (5.5–7.9), fresh (8.0–10.7), heavy (10.8–13.8)), amount of precipitation (light rain (0.0–2), rain (3–14), heavy rain (15–49), very heavy rain (more than 50)), and weather phenomena (thunderstorm, fog, rain, haze, snow, cloudy, drizzle, dust, hail) (Table 3). Maps in Figure 4 show average monthly weather indicators.



Figure 4. Cont.



Figure 4. Maps of average monthly weather indicators for 2017–2020: (**a**) maximum temperature in June; (**b**) maximum temperature in July; (**c**) maximum temperature in August; (**d**) precipitation in June; (**e**) precipitation in July; (**f**) precipitation in August.

Index	Weather Station Index	Daily Average Wind Speed	Daily Average Air Pressure	Daily Precipitation	Daily Average Temperature	Daily Average Relative Wetness	Horizontal Sight Range	Weather Events
0	30,127	0.625	748.7375	0.00	19.2750	73.625	19.75	haze
1	30,230	0.750	754.8875	0.00	18.6375	65.375	25.00	storm
2	29,594	5.500	755.5500	0.15	10.6375	48.875	45.00	rain
3	30,210	3.500	759.1000	0.25	11.1875	79.000	20.00	rain
4	30,069	0.875	755.3000	4.00	16.2000	80.375	41.75	cloudy

Table 3. The data on weather indicators.

Topographic data describe the relief of the territory. In our study, altitude, aspect, and slope are taken into account. To collect information on altitude, we used data from the

WorldClim resource [74] and the Institute of Geography of the SB RAS [75]. The digital elevation model of the QGIS analysis tools [76] helped generate slope and aspect maps (Figure 5). The aspect values are divided into eight categories in accordance with the criteria given in Table 4. Selected data on the relief are in Table 5.



Figure 5. A relief map of Irkutsk Oblast: (a) elevation; (b) aspect; (c) slope.

Table 4.	Categories	of the	"aspect"	variable.
----------	------------	--------	----------	-----------

Aspect	Azimuth (°)
North	337.5~22.5
Northeast	22.5~67.5
East	67.5~112.5
Southeast	112.5~157.5
South	157.5~202.5
Southwest	202.5~247.5
West	247.5~292.5
Northwest	292.5~337.5

Index	Aspect	Slope	Elevation
0	230.516327	3.406513	510.0
1	157.435410	2.461879	350.0
2	168.702484	1.773175	314.0
3	309.590942	1.614365	492.0
4	88.886177	3.588046	819.0

Table 5. Selected of data on the relief.

The vegetation of the territory is described according to forest regulations, in particular, data on forest seed zones: the territory is divided into relatively homogeneous parts according to natural factors that determine the formation of populations of a certain genotypic composition in the process of evolution. It is done, in particular, for pine, spruce, larch, and cedar (Irkutsk Oblast has 19 such zones) (Figure 6) (Table 6).



Figure 6. Division of the territory into zones of evolution: (a) cedar; (b) spruce; (c) pine; (d) larch.

Table 6. Selected data on the forest zones.

Index	Zones of Evolution	
0	pine-18, spruce-10, lurch-10, cedar-6	
1	pine-16, spruce-12, lurch-11, cedar-7	
2	pine-16, spruce-11, lurch-8, cedar-5	
3	pine-16, spruce-9, lurch-10, cedar-6	
4	pine-18, spruce-10, lurch-10, cedar-9	

Additionally, we take into account the Earth surface type as an indicator obtained by machine learning methods for the territory in question. The classification of Earth surface types is based on a convolutional neural network of the ResNet-50 architecture. The neural network was trained on labeled satellite images of the Irkutsk Oblast and the Republic of Buryatia during the summer period of 2018–2020, information on spectral indices (NDVI and LBP), and image texture in the form of local binary templates. As a result, we obtained the data on 11 surface classes: water, clouds, residential zones, mixed forest, coniferous forest, deciduous forest, open forest, bare rock, clearings, pasture, and agricultural land. The classification results are presented at https://geos.icc.ru/remotesensing (accessed on 27 August 2023) (Figure 7) [77].



Figure 7. Vegetation (surface) classes of Irkutsk Oblast.

The data on the water system of Irkutsk Oblast (Figure 8), roads, and railways were provided by the Institute of Geography of SB RAS [75] (Table 7).

Table 7. Selected data on the vegetation classes and the distance to rivers.

Index	Vegetation	Distance to Rivers
0	13.0	2.694398
1	11.0	10.831042
2	12.0	20.520019
3	13.0	22.165932
4	6.0	33.092805



Figure 8. A map of distances to water bodies in Irkutsk Oblast.

Social factors include distances to populated areas, roads, and railways (Figure 9). For factors that account for distances, an interval measurement scale was selected: 0–2; 2–5; 5–10; 10–20; 20–50; more than 50 km (Table 8).



Figure 9. A map of distances to roads and settlements in Irkutsk Oblast: (**a**) A map of distances to roads. (**b**) A map of distances to settlements.

Index	Distance to Roads	Distance to Settlements
0	87.405866	87.737006
1	6.679721	7.532349
2	12.783439	12.380512
3	69.071455	70.777155
4	4.500935	50.845274

Table 8. Selected data on the distance to roads and settlements.

Thus, our study describes wildfires in Irkutsk Oblast by topographic and climatic indicators, as well as by the type of vegetation comprising the land cover and distances to natural and anthropogenic objects (Table 9).

Variable Type	Variable Name	Code	Unit
Topographic	Elevation	elevation	m
	Slope	slope	degree
	Aspect	aspect	degree
Land cover	Vegetation cover	vegetation	classes of land cover
	The distance to river	distance_to_river	km
Climatic	Daily precipitation	RRR	mm
	Daily average temperature	Т	°C
	Daily average wind speed	FF	m/s
	Daily average air pressure	Р	mm Hg
	Daily average relative wetness	U	%
	Horizontal sight range	VV	km
	Daily weather events	WW_code	category
Social	The distance to road and railway	distance_to_road	km
	The distance to settlement	distance_to_set	km

3. Results

The fire hazard mapping approach implies building a spatial model of fire hazards using remote sensing data or data from various agencies that provide monitoring and fire management depending on landscape, climatic, and anthropogenic factors [14].

Currently, machine learning methods that are most often used to perform this task demonstrate greater accuracy. The Random Forest (RF) method is recognized as an effective method in many works.

RF is a popular machine learning algorithm proposed by Breiman in 2001 [61]. It is based on the traditional decision tree method and is capable of analyzing and assessing the relative importance of input factors with high classification accuracy, computation speed, and robustness to outliers [39].

The RF method is implemented using the Python programming language of the Sckitlearn module. The input data are presented in the form of CSV (Comma-Separated Values) files.

3.1. Data Preparation and Processing

Wildfire prediction is considered as a classification problem, where the dependent variable has two values: "fire" and "no fire". The available wildfire data belong to the first class. To build the second class, "absence of fire", random points were generated in the territory under consideration; the randomness was determined in time and space. To determine the randomness of the location, two approaches were explored: dividing the territory into sectors 50 by 50 km and introducing a buffer zone for locating the fire. The

generated event coordinates were located outside of the buffer zone around fires of 5 km in size, outside populated areas and man-made objects. The number of generated points for each area corresponded to the number of fires in that area. As a result, the following numbers of representatives from each class were used for the study: "fire"—9001, "no fire"—9001 (Figure 10).



Figure 10. Historical fires, "fire" points, and generated "no fire" points with a distance of 5 km from the fires.

While selecting the time parameter, the following variables were considered: year, season (spring–summer, summer–autumn), month (from March to November), decade, and day. No-fire time points were generated for dates in the wildfire season when the number of fires was less than 10.

At the next stage, we added data on factors influencing the fire occurrence to the obtained points. Due to the fact that the factor (WW_code) describing the weather is categorical, the data set covered 27,522 records, since each record contained all possible unique categories of current weather that occurred for the day in question. Thus, the ratio of records for the classes "fire" and "no fire" was 44% to 56%.

3.2. Multicollinearity and Correlation

The evaluation of the correlation [78] between the model variables showed a strong linear dependence between distances to the road (distance_to_road) and settlements (distance_to_set) (0.82), slope and elevation (0.66), temperature (T), and barometric pressure (P) (-0.68) (Figure 11).



Figure 11. Correlation matrix for model variables.

As a result of this evaluation, we eliminated the variables of barometric pressure (P), distance to settlements (distance_to_set), and slope from the study. Since the correlation values of these variables are modulo more than 0.6, this means that they are not independent and reflect the same information; therefore, when using them, the model will be retrained.

The next stage was to assess the collinearity of the available fire factors for the territory under consideration. Multicollinearity in regression analysis occurs when two or more independent variables are so highly correlated with each other that they do not provide independent information in the regression model. One way to detect multicollinearity is to use a metric known as the variance inflation factor (VIF), which measures the strength of the correlation between independent variables in a regression model. A VIF value of 1 is considered to indicate that there is no correlation between the given independent variable and any other independent variables in the model. A value greater than 5 indicates a potentially strong correlation, in which case the coefficient estimates and p-values in the regression output are likely to be unreliable. In this study, the parameters for building a wildfire risk classification model were selected as a result of a collinearity assessment. They are presented in Tables 10 and 11 [73].

Table 10. Results of assessing	g collinearity of a	all model variables.
--------------------------------	---------------------	----------------------

Variable	Description	VIF
Т	daily average temperature	2.259900
Ff	daily average wind speed	1.248514
RRR	daily precipitation slope	1.186837
VV	horizontal sight range	1.107968
U	daily average relative wetness	1.693530
Р	barometric pressure	2.125026
WW_code	daily weather events	1.200167
distance_to_road	the distance to road and railway	3.118877
distance_to_river	the distance to river	1.036912
distance_to_set	the distance to settlement	3.218824
elevation	elevation	1.881964
slope	slope	1.856331
aspect	aspect	1.011978
vegetation	vegetation cover	1.102114

Variable	Description	VIF
Т	daily average temperature	1.132697
Ff	daily average wind speed	1.199376
RRR	daily precipitation slope	1.173968
VV	horizontal sight range	1.105191
U	daily average relative wetness	1.646322
WW_code	daily weather events	1.180479
distance_to_road	the distance to road and railway	1.122223
distance_to_river	the distance to river	1.027408
elevation	elevation	1.101271
aspect	aspect	1.006388
vegetation	vegetation cover	1.093775

Table 11. Results of assessing collinearity of the independent model variables.

Thus, the results confirm the independence of the model variables.

3.3. Classification

To train the model, the data were divided into training and test samples in a ratio of 70% and 30%, respectively. Moreover, in each sample, the classes were balanced and had a ratio identical to that of the original set, namely, 44% to 56%.

Two important parameters affect the performance of RF: the number of trees in the forest and the number of random variables per the split node. Using the Randomized-SearchCV() method and the input data (a Random Forest Classifier model with default parameters, number of iterations, and number of cross-validations), the hyper parameters of the model were optimized and selected. In our work, the following parameters were used: number of trees (n_estimators)—600, number of values in a tree node (min_samples_leaf)—2, minimum number of samples required to split an internal node (min_samples_split)—23, maximum tree depth (max_depth)—15.

Tables 12 and 13 show the importance of the classification factors delivered by the RF method.

Variable	Description	Gini-Importance
Т	daily average temperature	0.177997
distance_to_set	the distance to settlement	0.147473
distance_to_road	the distance to road and railway	0.119130
WW_code	daily weather events	0.096867
vegetation	vegetation cover	0.087858
U	daily average relative wetness	0.087460
elevation	elevation	0.060333
VV	horizontal sight range	0.052372
RRR	daily precipitation	0.046682
Р	barometric pressure	0.036404
slope	slope	0.028554
distance_to_river	the distance to river	0.021861
Ff	daily average wind speed	0.020577
aspect	aspect	0.016433

Table 12. The importance of all variables (factors) of the model.

Variable	Description	Gini-Importance
Т	daily average temperature	0.201188
distance_to_road	the distance to road and railway	0.200019
WW_code	daily weather events	0.120854
vegetation	vegetation cover	0.119191
Ū	daily average relative wetness	0.097161
elevation	elevation	0.078460
VV	horizontal sight range	0.061406
RRR	daily precipitation	0.044311
distance_to_river	the distance to river	0.030183
Ff	daily average wind speed	0.025377
aspect	aspect	0.021849

The most important factors that influence the occurrence of a wildfire are climatic factors (especially the average daily temperature and current weather), social factors (distance to populated areas, roads), and the class of vegetation cover. Topographic indicators, humidity, precipitation, and distance to water bodies have less influence in the final model.

The average probability of a wildfire occurrence (belonging to the "fire" class) by month is reflected in the graph (Figure 12) and shows dynamics similar to those of the actual number of wildfires.



Figure 12. Comparison of the probability and number of wildfires by month.

The distribution of the classes "fire" and "no fire" by climatic categorical variables is shown in Figure 13. The graphs show the values of weather conditions that can cause wildfires. To a greater extent, wildfires were recorded at temperatures above 0 $^{\circ}$ C; a positive trend in temperature and the number of wildfires was observed, especially in the temperature range of more than 20 $^{\circ}$ C. Wildfires are more closely associated with such weather phenomena as haze, fog, thunderstorms, and dust, lack of precipitation, and lower relative humidity, which characterize dry climatic conditions and the occurrence of possible causes of wildfires, for example, dry thunderstorms.



Figure 13. Distribution of the number of records by classes "fire" and "no fire" by categories of climate variables: (a) temperature; (b) weather events; (c) relative humidity; (d) horizontal sight range; (e) precipitation; (f) wind speed.

Figure 14 shows the frequency of occurrence of the classes "fire" and "no fire" according to the characteristics of land cover vegetation, distance to roads, and relief height. The distribution of forest fires by vegetation class showed a higher frequency of occurrence in fields and open forest areas. A large proportion of wildfires were also characterized by a distance of less than 5 km from roads and a lower relief height. The distance to populated areas followed the distribution of distances to roads. The height of the relief was also associated with settlements and roads; they were located in lowlands. The nature of the distribution of wildfires, taking into account social factors, may indicate an anthropogenic cause of fires.

To build the fire hazard maps of Irkutsk Oblast using QGIS software (ver.3.34.2) [76], we made a grid of points with a distance of 1 km. For each point, the necessary values of input factors are determined, and the probability of belonging to the "fire" class is obtained using the developed model, which is further referred to as the "risk" field. For these points, interpolation is set using the "risk" attribute, where the cell area is 150×150 km. Next, a reclassification method was applied, as a result of which each final cell of the map is classified into five categories [79]: I—very high (probability 0.8–1), II—high (probability 0.6–0.8), III—medium (probability 0.4–0.6), IV—low (probability 0.2–0.4), V—very low (probability 0–0.2), representing the fire hazard class. The results obtained for assessing susceptibility to wildfires using the Random Forest model are shown in Figure 15. A low value (blue) represents areas with the least likelihood of wildfires.



Figure 14. Distribution of the number of records by classes "fire" and "no fire" by categories of vegetation, distance to roads, and relief heights: (**a**) land cover; (**b**) distance to road (km); (**c**) elevation (m).



Figure 15. Map of fire hazard (risk) in Irkutsk Oblast based on average climatic indicators for the summer period: very high—probability 0.8–1, high—probability 0.6–0.8, medium—probability 0.4–0.6, low—probability 0.2–0.4, very low—probability 0–0.2.

The algorithms and software of our study can be found in [80].

3.4. Model Performance Evaluation

The effectiveness of the wildfire risk prediction model was assessed based on the following metrics: accuracy, f1-score, and AUC (Table 14).

Table 14. Results of assessing the accuracy of the model.

Metrics	Value
Accuracy	0.89
F1-score	0.88
AUC	0.96

The ROC curve shows the relationship between the proportion of correct predictions (True Positive Rate) and the proportion of erroneous predictions (False Positive Rate) (Figure 16).





Table 15 presents an error matrix that shows the distribution for the actual and predicted "fire" and "no fire" classes.

Table 15. Classification error matrix.

	Predicted	
Actual	"No Fire"	"Fire"
"no fire"	4332 (93%), TN ¹	312 (7%), FP
"fire"	534 (15%), FN	3079 (85%), TP

¹ TN—true negatives, FN—false negatives, FP—false positives, TP—true positives.

4. Discussion and Conclusions

The present study employed the Random Forest machine learning method to address the problem of mapping the fire hazards in the territory of Irkutsk Oblast using meteorological factors, data on relief and vegetation, and social factors. The novelty of the research is the analysis of the most complete set of factors influencing the occurrence of wildfires, coverage of the entire territory of Irkutsk Oblast, and comparison of the results using case-based reasoning and RF methods. We developed and justified algorithmic foundations

23 of 27

of the methodology and information technology for the wildfire prediction in a certain territory, taking into account its individual characteristics.

To apply machine learning methods, an analysis of the factors influencing the wildfire occurrence was carried out. Collinearity was assessed, and, as a result, the atmospheric pressure, distance to settlements, and slope variables were excluded from the model. It was revealed that the climatic factors are most significant, especially the average daily temperature and the current weather, including thunderstorm activity, social factors (distance to roads), as well as the class of vegetation cover.

As part of the study, we carried out a wildfire risk analysis using case-based reasoning (CBR) [22,81] and RF methods.

The accuracy evaluation of the CBR method used the ratio of high-risk events (P) to all events (N) as a quantitative estimate for the prediction of accuracy (Accuracy) [81]:

$$Accuracy = \sum P/N,$$
 (1)

where P = 1 if the number of cases with similarity assessment in the interval [0.8, 1] > 0, and P = 0 if the number of cases with similarity assessment in the interval [0.8, 1] = 0.

For the reviewed forest districts, the value of the prediction accuracy estimation on the developed case bases was 0.874.

The RF method allowed us to achieve the following accuracy scores: accuracy of 0.89, F1-score of 0.88, and AUC of 0.96.

At this stage, both methods show good results. Figure 17 shows a visualization of the results of applying the listed methods to build risk maps for the Kazachinsko-Lensky municipality for the fire-hazardous period June–August 2020. The created maps are the result of a preliminary forecast and can become the basis for the formation of control actions. In the future, during the fire-hazardous period, when receiving short-term forecast data on weather conditions, the results of the risk assessment should be clarified.



Figure 17. Results of constructing fire hazard maps for the Kazachinsko-Lena municipality for the fire-hazardous period June–August 2020: (a) CBR method [22]; (b) RF method. Hazard classes: very high—probability 0.8–1, high—probability 0.6–0.8, medium—probability 0.4–0.6, low—probability 0.2–0.4, very low—probability 0–0.2.

The CBR method is recommended to be used at the initial (preliminary) stage of research to identify areas most at risk of wildfires, especially when there are not enough data and preliminary research on the selection of important factors influencing the wildfire occurrence has not been carried out. The CBR assessment by the method is more "rough", as can be observed in Figure 17a, where there are no classes low and very low. As data on the territory accumulate, it will be possible to use other machine learning methods, in particular, RF and neural networks, especially their combination in the form of ensembles.

In the future, we plan to study additional factors to include in the model and obtain a more accurate solution, in particular, winter snowiness, holidays, gross domestic product, and more accurate data on thunderstorms. Subsequent research also implies building and training models that take into account individual characteristics of forest districts.

The results of the risk mapping of the territory showed that the southern most densely populated territories of the region, as well as territories located along the roads, are most at risk of wildfires, which confirmed the statement of government authorities that the human factor is the main cause of wildfires. However, it is necessary to clarify the model and detail the factors in which human activities (agriculture, woodworking, hunting, and fishing) are decisive.

The approach developed will become the basis for creating a wildfire prediction information technology designed for both scientific research and management. It is proposed, within the framework of the technology, to implement the functions of risk assessment in the territories of forest blocks, forest districts, municipalities, and other territorial structures. Wildfire risk maps will provide government authorities with additional information for making decisions on measures to reduce the risk and mitigate the consequences of wildfires.

Author Contributions: Conceptualization, O.N. and A.Y.; methodology, O.N. and J.P.; software, J.P.; validation, J.P.; formal analysis, O.N. and J.P.; investigation, O.N., A.Y. and J.P.; data curation, J.P.; writing—original draft preparation, O.N., A.Y. and J.P.; writing—review and editing, O.N. and A.Y.; visualization, J.P.; project administration, O.N. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Science and Higher Education of the Russian Federation, Grant No. 075-15-2020-787 for implementation of Major scientific projects on priority areas of scientific and technological development (the project "Fundamentals, methods and technologies for digital monitoring and forecasting of the environmental situation on the Baikal natural territory").

Data Availability Statement: Data on relief of Irkutsk Oblast are available in a publicly accessible repository that does not issue DOIs. Publicly available datasets were analyzed in this study. These data can be found here: https://www.worldclim.org/data/worldclim21.html (accessed on 27 August 2023). Data on average monthly weather indicators, including maximum temperature and precipitation, are available in a publicly accessible repository that does not issue DOIs. Publicly available datasets were analyzed in this study. These data can be found here: [https://www.worldclim.org/data/monthlywth.html] (accessed on 27 August 2023). Data on air temperature, atmospheric pressure, relative humidity, wind direction, wind speed, amount of precipitation, and weather phenomena are available in a publicly accessible repository that does not issue DOIs. Publicly available datasets were analyzed in this study. These data can be found here: [https://rp5.ru/IIoroga_B_MMpe] (accessed on 27 August 2023). Data on the classes of land cover are available in a publicly accessible repository that does not issue DOIs. Publicly accessible repository that does not issue DOIs. Publicly available datasets were analyzed in this study. These data can be found here: [https://rp5.ru/IIoroga_B_MMpe] (accessed on 27 August 2023). Data on the classes of land cover are available in a publicly accessible repository that does not issue DOIs. Publicly available datasets were analyzed in this study. These data can be found here: [https://geos.icc.ru/remotesensing] (accessed on 27 August 2023).

Acknowledgments: We would like to thank Evgeniy Kosogorov, master student of the Institute of Mathematics and Information Technologies of the Irkutsk State University, for technical support during data processing.

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

References

- 1. Eskandari, S.; Chuvieco, E. Fire danger assessment in Iran based on geospatial information. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, 42, 57–64. [CrossRef]
- Semeraro, T.; Mastroleo, G.; Aretano, R.; Facchinetti, G.; Zurlini, G.; Petrosillo, I. GIS Fuzzy Expert System for the assessment of ecosystems vulnerability to fire in managing Mediterranean natural protected areas. *J. Environ. Manag.* 2016, 168, 94–103. [CrossRef] [PubMed]
- 3. Jaafari, A.; Gholami, D.M.; Zenner, E.K. A Bayesian modeling of wildfire probability in the Zagros Mountains, Iran. *Ecol. Inf.* **2017**, *39*, 32–44. [CrossRef]
- 4. Oliveira, S.; Félix, F.; Nunes, A.; Lourenço, L.; Laneve, G.; Sebastián-López, A. Mapping wildfire vulnerability in Mediterranean Europe. testing a stepwise approach for operational purposes. *J. Environ. Manag.* **2018**, *206*, 158–169. [CrossRef]
- 5. Tutmez, B.; Ozdogan, M.G.; Boran, A. Mapping forest fires by nonparametric clustering analysis. J. For. Res. 2018, 29, 177–185. [CrossRef]
- 6. Hong, H.; Jaafari, A.; Zenner, E.K. Predicting spatial patterns of wildfire susceptibility in the Huichang County, China: An integrated model to analysis of landscape indicators. *Ecol. Indic.* **2019**, *101*, 878–891. [CrossRef]
- Neary, D.G.; Ryan, K.C.; DeBano, L.F. Wildland Fire in Ecosystems: Effects of Fire on Soils and Water, General Technical Report RMRSGTR-42-Volume 4; USDA, Forest Service, Rocky Mountain Research Station: Ogden, UT, USA, 2005; p. 250.
- 8. Rafaqat, W.; Iqbal, M.; Kanwal, R.; Song, W. Study of Driving Factors Using Machine Learning to Determine the Effect of Topography, Climate, and Fuel on Wildfire in Pakistan. *Remote Sens.* **2022**, *14*, 1918. [CrossRef]
- Kuklina, V.; Sizov, O.; Rasputina, E.; Bilichenko, I.; Krasnoshtanova, N.; Bogdanov, V.; Petrov, A. Fires on Ice: Emerging Permafrost Peatlands Fire Regimes in Russia's Subarctic Taiga. Land 2022, 11, 322. [CrossRef]
- 10. FerreirA-leiTe, F.; Lourenço, L.; Bento-Gonçalves, A. Large forest fires in mainland Portugal, brief characterization. *J. Mediterr. Geogr.* **2013**, *19*, 53–65. [CrossRef]
- 11. Xu, R.; Yu, P.; Abramson, M.J.; Johnston, F.H.; Samet, J.M.; Bell, M.L.; Haines, A.; Ebi, K.L.; Li, S.; Guo, Y. Wildfires, global climate change, and human health. *N. Engl. J. Med.* 2020, *383*, 2173–2181. [CrossRef]
- 12. Johnston, L.M.; Wang, X.; Erni, S.; Taylor, S.W.; McFayden, C.B.; Oliver, J.A.; Stockdale, C.; Christianson, A.; Boulanger, Y.; Gauthier, S. Wildland fire risk research in Canada. *Environ. Rev.* **2020**, *28*, 164–186. [CrossRef]
- 13. Timofeeva, S.S.; Garmyshev, V.V. Ecological consequences of forest fires in the Irkutsk region. *Ecol. Ind. Russ.* 2017, 21, 46–49. (In Russian) [CrossRef]
- 14. Jain, P.; Coogan, S.C.P.; Subramanian, S.G.; Crowley, M.; Taylor, S.; Flannigan, M.D. A review of machine learning applications in wildfire science and management. *arXiv* 2020, arXiv:2003.00646v1. [CrossRef]
- 15. Gheshlaghi, H.A.; Feizizadeh, B.; Blaschke, T. GIS-based forest fire risk mapping using the analytical network process and fuzzy logic. *J. Environ. Plan. Manag.* **2019**, *63*, 481–499. [CrossRef]
- Carvalho, J.P.; Carola, M.; Tome, J.A.B. Forest fire modeling using rule-based fuzzy Cognitive maps and Voronoi based cellular automata. In Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society NAFIPS 2006, Montreal, QC, Canada, 3–6 June 2006; pp. 217–222.
- 17. Soto, M.E.C. The Identification and Assessment of Areas at Risk of Forest Fire Using Fuzzy Methodology. *Appl. Geogr.* **2012**, *35*, 199–207. [CrossRef]
- 18. Leal, B.E.Z.; Hirakawa, A.R.; Pereira, T.D. Onboard Fuzzy Logic Approach to Active Fire Detection in Brazilian Amazon Forest. *IEEE Trans. Aerosp. Electron. Syst.* **2016**, *52*, 883–890. [CrossRef]
- 19. Pourghasemi, H.; Beheshtirad, M.; Pradhan, B. A comparative assessment of prediction capabilities of modified analytical hierarchy process (M-AHP) and Mamdani fuzzy logic models using Netcad-GIS for forest fire susceptibility mapping. *Geomat. Nat. Hazards Risk* 2016, 7, 861–885. [CrossRef]
- 20. Bui, D.T.; Bui, Q.T.; Nguyen, Q.P.; Pradhan, B.; Nampak, H.; Trinh, P.T. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. *Agric. For. Meteorol.* **2017**, 233, 32–44.
- Jaafari, A.; Zenner, E.K.; Panahi, M.; Shahabi, H. Hybrid artificial intelligence models based on a neuro-fuzzy system and metaheuristic optimization algorithms for spatial prediction of wildfire probability. *Agric. For. Meteorol.* 2019, 266–267, 198–207. [CrossRef]
- 22. Dorodnykh, N.; Nikolaychuk, O.; Pestova, J.; Yurin, A. Forest Fire Risk Forecasting with the Aid of Case-Based Reasoning. *Appl. Sci.* 2022, 12, 8761. [CrossRef]
- 23. Pan, Y.; Birdsey, R.A.; Phillips, O.L.; Jackson, R.B. The Structure, Distribution, and Biomass of the World's Forests. *Annu. Rev. Ecol. Evol. Syst.* **2013**, *44*, 593–622. [CrossRef]
- 24. Liu, W.; Wang, S.; Zhou, Y.; Wang, L.; Zhu, J.; Wang, F. Lightning-caused forest fire risk rating assessment based on case-based reasoning: A case study in DaXingAn Mountains of China. *Nat. Hazards* **2016**, *81*, 347–363. [CrossRef]
- 25. Bot, K.; Borges, J.G. A Systematic Review of Applications of Machine Learning Techniques for Wildfire Management Decision Support. *Inventions* 2022, 7, 15. [CrossRef]
- 26. De Vasconcelos, M.P.; Silva, S.; Tome, M.; Alvim, M.; Pereira, J.C. Spatial prediction of fire ignition probabilities: Comparing logistic regression and neural networks. Photogramm. *Eng. Remote Sens.* **2001**, *67*, 73–81.

- 27. Bisquert, M.; Caselles, E.; Sánchez, J.M.; Caselles, V. Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data. *Int. J. Wildl. Fire* **2012**, *21*, 1025. [CrossRef]
- 28. Safi, Y.; Bouroumi, A. Prediction of forest fires using artificial neural networks. Appl. Math. Sci. 2013, 7, 271–286. [CrossRef]
- 29. Goldarag, Y.J.; Mohammadzadeh, A.; Ardakani, A.S. Fire Risk Assessment Using Neural Network and Logistic Regression. J. Ind. Soc. Remote Sens. 2016, 44, 885–894. [CrossRef]
- 30. Adab, H. Landfire hazard assessment in the Caspian Hyrcanian forest ecoregion with the long-term MODIS active fire data. *Nat. Hazards* **2017**, *87*, 1807–1825. [CrossRef]
- Thach, N.N.; Ngo, D.B.T.; Xuan-Canh, P.; Hong-Thi, N.; Thi, B.H.; Nhat-Duc, H.; Dieu, T.B. Spatial pattern assessment of tropical forest fire danger at Thuan Chau area (Vietnam) using GIS-based advanced machine learning algorithms: A comparative study. *Ecol. Inform.* 2018, 46, 74–85. [CrossRef]
- 32. Zhang, G.; Wang, M.; Liu, K. Forest fire susceptibility modeling using a convolutional neural network for Yunnan province of China. *Int. J. Disaster Risk Sci.* 2019, *10*, 386–403. [CrossRef]
- Wu, Z.; Wang, B.; Li, M.; Tian, Y.; Quan, Y.; Liu, J. Simulation of forest fire spread based on artificial intelligence. *Ecol. Indic.* 2022, 136, 108653. [CrossRef]
- Oliveira, S.; Oehler, F.; San-Miguel-Ayanz, J.; Camia, A.; Pereira, J.M. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *For. Ecol. Manag.* 2012, 275, 117–129. [CrossRef]
- 35. Luo, R.; Dong, Y.; Gan, M.; Li, D.; Niu, S.; Oliver, A.; Wang, K.-L.; Luo, Y. Global analysis of influencing forces of fire activity: The threshold relationships between vegetation and fire. *Life Sci. J.* **2013**, *10*, 15–24.
- 36. Van Breugel, P.; Friis, I.; Demissew, S.; Lillesø, J.P.B.; Kindt, R. Current and future fire regimes and their influence on natural vegetation in Ethiopia. *Ecosystems* **2016**, *19*, 369–386. [CrossRef]
- 37. Guo, F.; Zhang, L.; Jin, S.; Tigabu, M.; Su, Z.; Wang, W. Modeling anthropogenic fire occurrence in the boreal forest of China using logistic regression and random forests. *Forests* **2016**, *7*, 250. [CrossRef]
- Cao, Y.; Wang, M.; Liu, K. Wildfire susceptibility assessment in southern China: A comparison of multiple methods. *Int. J. Disaster Risk Sci.* 2017, 8, 164–181. [CrossRef]
- Chaoxue, T.; Feng, Z. Mapping Forest Fire Risk Zones Using Machine Learning Algorithms in Hunan Province, China. Sustainability 2023, 15, 6292. [CrossRef]
- 40. Xu, X.; Huang, X.; Bian, H.; Wu, J.; Liang, C.; Cong, F. Total process of fault diagnosis for wind turbine gearbox, from the perspective of combination with feature extraction and machine learning: A review. *Energy AI* **2024**, *15*, 100318. [CrossRef]
- 41. Tsolaki, K.; Vafeiadis, T.; Nizamis, A.; Ioannidis, D.; Tzovaras, D. Utilizing machine learning on freight transportation and logistics applications: A review. *ICT Express* **2023**, *9*, 284–295. [CrossRef]
- 42. Rozante, J.R.; Ramirez, E.; Ramirez, D.; Rozante, G. Improved frost forecast using machine learning methods. *Artif. Intell. Geosci.* **2023**, *4*, 164–181. [CrossRef]
- 43. Liu, Z.; Ma, J.; Xia, D.; Jiang, S.; Ren, Z.; Tan, C.; Lei, D.; Guo, H. Toward the reliable prediction of reservoir landslide displacement using earthworm optimization algorithm-optimized support vector regression (EOA-SVR). *Nat. Hazards* **2023**. [CrossRef]
- 44. Ma, J.; Lei, D.; Ren, Z.; Tan, C.; Xia, D.; Guo, H. Automated Machine Learning-Based Landslide Susceptibility Mapping for the Three Gorges Reservoir Area, China. *Math. Geosci.* 2023. [CrossRef]
- Moskvichev, V.V.; Postnikova, U.S.; Taseiko, O.V. Information system for monitoring and managing the risks of development of Siberia and the Arctic regions. *Reliab. Theory Appl.* 2022, 17, 124–131.
- 46. Report on the Results of the Joint Control Event "Checking the Effectiveness of Planning and Spending Budget Funds Allocated for the Technical Equipment of the Constituent Entities of the Russian Federation with Forest Fire Equipment and Machinery". Available online: https://ach.gov.ru/upload/iblock/247/247d140a4eb4b0607c43b585a2a5e0ee.pdf (accessed on 15 January 2023).
- 47. Sofronova, T.M.; Volokitina, A.V.; Sofronov, M.A. Assessment of fire danger based on weather conditions in the mountain forests of the Southern Baikal region. *Geogr. Nat. Resour.* **2008**, *2*, 74–80.
- Eugenio, F.C.; Dos Santos, A.R.; Fiedler, N.C.; Ribeiro, G.A.; Da Silva, A.G.; Dos Santos, A.B.; Paneto, G.G.; Schettino, V.R. Applying GIS to Develop a Model for Forest Fire Risk: A Case Study in Espirito Santo, Brazil. *J. Environ. Manag.* 2016, 173, 65–71. [CrossRef] [PubMed]
- Tian, X.; Zhao, F.; Shu, L.; Wang, M. Distribution Characteristics and the Influence Factors of Forest Fires in China. *For. Ecol. Manag.* 2013, 310, 460–467. [CrossRef]
- 50. Bychkov, I.V.; Ruzhnikov, G.M.; Fedorov, R.K.; Khmelnov, A.E.; Popova, A.K. Organization of digital monitoring of the Baikal natural territory. *IOP Conf. Ser. Earth Environ. Sci.* 2021, 629, 012067. [CrossRef]
- Bashari, H.; Naghipour, A.A.; Khajeddin, S.J.; Sangoony, H.; Tahmasebi, P. Risk of fire occurrence in arid and semi-arid ecosystems of Iran: An investigation using Bayesian belief networks. *Environ. Monit. Assess* 2016, 188, 531. [CrossRef]
- 52. Sivrikaya, F.; Küçük, Ö. Modeling forest fire risk based on GIS-based analytical hierarchy process and statistical analysis in Mediterranean region. *Ecol. Inform.* **2022**, *68*, 101537. [CrossRef]
- 53. Mosadeghi, R.; Warnken, J.; Tomlinson, R.; Mirfenderesk, H. Uncertainty Analysis in the Application of Multi-Criteria Decision-Making Methods in Australian Strategic Environmental Decisions. J. Environ. Plan. Manag. 2013, 56, 10978. [CrossRef]
- 54. Feizizadeh, B.; Omrani, K.; Aghdam, F.B. Fuzzy Analytical Hierarchical Process and Spatially Explicit Uncertainty Analysis Approach for Multiple Forest Fire Risk Mapping. *J. Geogr. Inf. Sci.* **2015**, *3*, 72–80. [CrossRef]

- 55. Ghorbanzadeh, O.; Valizadeh Kamran, K.; Blaschke, T.; Aryal, J.; Naboureh, A.; Einali, J.; Bian, J. Spatial prediction of wildfire susceptibility using field survey GPS data and machine learning approaches. *Fire* **2019**, *2*, 43. [CrossRef]
- 56. Rodrigues, M.; de la Riva, J. An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environ. Modell. Softw.* **2014**, *57*, 192–201. [CrossRef]
- 57. Gigović, L.; Pourghasemi, H.R.; Drobnjak, S.; Bai, S. Testing a new ensemble model based on SVM and random forest in forest fire susceptibility assessment and its mapping in Serbia's Tara National Park. *Forests* **2019**, *10*, 408. [CrossRef]
- 58. Tymstra, C.; Jain, P.; Flannigan, M.D. Characterisation of initial fire weather conditions for large spring wildfires in Alberta, Canada. *Int. J. Wildland Fire* **2021**, *30*, 823–835. [CrossRef]
- 59. Júnior, J.S.S.; Paulo, J.R.; Mendes, J.; Alves, D.; Ribeiro, L.M.; Viegas, C. Automatic forest fire danger rating calibration: Exploring clustering techniques for regionally customizable fire danger classification. *Expert Syst. Appl.* **2022**, 193, 116380. [CrossRef]
- Pourghasemi, H.R.; Gayen, A.; Lasaponara, R.; Tiefenbacher, J.P. Application of learning vector quantization and different machine learning techniques to assessing forest fire influence factors and spatial modeling. *Environ. Res.* 2020, 184, 109321. [CrossRef]
- 61. Breiman, L. Random forests. Mach. Learn. 2001, 45, 5–23. [CrossRef]
- 62. Harris, L.B.; Taylor, A.H.; Kassa, H.; Leta, S.; Powell, B. Humans and climate modulate fire activity across Ethiopia. *Fire Ecol.* 2023, 19, 15. [CrossRef]
- 63. Archibald, S.; Roy, D.P.; van Wilgen, B.W.; Scholes, R.J. What limits fire? An examination of drivers of burnt area in Southern Africa. *Glob. Chang. Biol.* 2009, *15*, 613–630. [CrossRef]
- 64. Latysheva, I.V.; Vologzhina, S.Z.; Loshchenko, K.A.; Makukhin, V.L. Influence of forest fires on pollution level territories of the Irkutsk region. In Proceedings of the LXXVI Herzen Readings—Geography: Development of Science and Education, St. Petersburg, Russia, 19 April 2023.
- 65. Tomshin, O.; Solovyev, V. Spatio-temporal patterns of wildfires in Siberia during 2001–2020. *Geocarto Int.* **2022**, *37*, 7339–7357. [CrossRef]
- 66. Bartalev, S.A.; Stytsenko, F.V.; Egorov, V.A.; Loupian, E.A. Satellite-based assessment of Russian forest fire mortality. *Russ. J. For. Sci.* 2015, *2*, 83–94.
- 67. Drozdova, T.I.; Sorokovikova, E.V. Analysis of forest fires in Irkutsk region for 2010–2019. XXI Century Technosphere Saf. 2021, 6, 29–41. (In Russian) [CrossRef]
- 68. Belousova, Y.P.; Latysheva, I.V.; Latyshev, S.V.; Loshchenko, K.A.; Shcheblykin, A.S. Natural factors of forest fires in Irkutsk oblast. *Biosfera* **2016**, *8*, 390–400.
- 69. Available online: https://38.mchs.gov.ru/deyatelnost/press-centr/operativnaya-informaciya/svodka-chs-i-proisshestviy/24 68755 (accessed on 27 August 2023).
- Vashchalova, T.V.; Garmyshev, V.V. Atmospheric pollution of the Irkutsk region as a result of natural fires and public health risk assessment. RUDN J. Ecol. Life Saf. 2020, 28, 252–262. [CrossRef]
- 71. Taschilin, M.; Yakovleva, I.; Sakerin, S.; Zorkaltseva, O.; Tatarnikov, A.; Scheglova, E. Spatiotemporal Variations of Aerosol Optical Depth in the Atmosphere over Baikal Region Based on MODIS Data. *Atmosphere* **2021**, *12*, 1706. [CrossRef]
- 72. Institute of Solar-Terrestrial Physics of Siberian Branch of Russian Academy of Sciences (ISTP SB RAS). Available online: http://en.iszf.irk.ru/Main_Page (accessed on 13 August 2022).
- 73. Irkutsk Department of Hydrometeorology and Environmental Monitoring. Available online: https://www.irmeteo.ru/ (accessed on 13 August 2023).
- 74. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 2017, 37, 4302–4315. [CrossRef]
- 75. V.B. Sochava Institute of Geography SB RAS. Available online: http://www.irigs.irk.ru/# (accessed on 13 August 2023).
- 76. QGIS. A Free and Open Source Geographic Information System. Available online: https://qgis.org (accessed on 27 August 2023).
- 77. Bychkov, I.V.; Ruzhnikov, G.M.; Fedorov, R.K.; Popova, A.K.; Avramenko, Y.V. On classification of Sentinel-2 satellite images by a neural network ResNet-50. *Comput. Opt.* **2023**, *47*, 474–481. [CrossRef]
- NumPy, SciPy, and Pandas: Correlation With Python. Available online: https://realpython.com/numpy-scipy-pandas-correlation-python/ (accessed on 27 August 2023).
- 79. GOST R 22.1.09-99 Safety in Emergency Situations. Monitoring and Forecasting of Forest Fires. Available online: https://docs.cntd.ru/document/1200025900 (accessed on 27 August 2023). (In Russian)
- GitHub—Lab42-Team/geoanalitics: An application to analyzing geo-data. Available online: https://github.com/Lab42-Team/geoanalytics (accessed on 15 October 2023).
- Dorodnykh, N.O.; Nikolaychuk, O.A.; Pestova, J.V.; Yurin, A.Y. Creation of Prototypes of Case-Based Knowledge Bases Using Transformations of Decision Tables to Predict the Risk of Forest Fires. *Pattern Recognit. Image Anal.* 2023, 33, 274–281. [CrossRef]

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.