

Article

Performance Evaluation of Machine Learning Methods for Forest Fire Modeling and Prediction

Binh Thai Pham ¹, Abolfazl Jaafari ², Mohammadtaghi Avand ³, Nadhir Al-Ansari ^{4,*}, Tran Dinh Du ⁵, Hoang Phan Hai Yen ⁶, Tran Van Phong ⁷, Duy Huu Nguyen ⁸, Hiep Van Le ^{9,*}, Davood Mafi-Gholami ¹⁰, Indra Prakash ¹¹, Hoang Thi Thuy ¹² and Tran Thi Tuyen ^{12,*}

¹ University of Transport Technology, Hanoi 100000, Vietnam; binhpt@utt.edu.vn
² Research Institute of Forests and Rangelands, Agricultural Research, Education, and Extension Organization (AREEO), Tehran 64414-356, Iran; jaafari@rifr.ac.ir

³ Department of Watershed Management Engineering, College of Natural Resources, Tarbiat Modares University, Tehran 14115-111, Iran; Mt.avand@modares.ac.ir

⁴ Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, 97187 Lulea, Sweden

⁵ Department of Land Management, School of Agriculture and Resources, Vinh University, Nghe An 43000, Vietnam; dutd@vinhuni.edu.vn

⁶ Department of Geography, School of Social Education Vinh University, Nghe An 43000, Vietnam; hoangphanhaiyen@vinhuni.edu.vn

⁷ Institute of Geological Sciences, Vietnam Academy of Sciences and Technology, 84 Chua Lang Street, Dong da, Hanoi 100000, Vietnam; tvphong@igsvn.vast.vn

⁸ Faculty of Geography, VNU University of Science, Vietnam National University, 334 Nguyen Trai, Ha Noi 100000, Vietnam; nguyenhuuduy@hus.edu.vn

⁹ Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

¹⁰ Department of Forest Sciences, Faculty of Natural Resources and Earth Sciences, Shahrood University, Shahrood 8818634141, Iran; d.mafigholami@nres.sku.ac.ir

¹¹ Department of Science & Technology, Bhaskarcharya Institute for Space Applications and Geo-Informatics (BISAG), Government of Gujarat, Gandhinagar 382002, India; info@bisag.gujarat.gov.in

¹² Department of Resource and Environment Management, School of Agriculture and Resources, Vinh University, Nghe An 43000, Vietnam; thuyhoangthi@vinhuni.edu.vn

* Correspondence: nadhir.alansari@ltu.se (N.A.-A.); levanhiep2@duytan.edu.vn (H.V.L.); tuyentt@vinhuni.edu.vn (T.T.T.)

Received: 6 May 2020; Accepted: 12 June 2020; Published: 17 June 2020



Abstract: Predicting and mapping fire susceptibility is a top research priority in fire-prone forests worldwide. This study evaluates the abilities of the Bayes Network (BN), Naïve Bayes (NB), Decision Tree (DT), and Multivariate Logistic Regression (MLR) machine learning methods for the prediction and mapping fire susceptibility across the Pu Mat National Park, Nghe An Province, Vietnam. The modeling methodology was formulated based on processing the information from the 57 historical fires and a set of nine spatially explicit explanatory variables, namely elevation, slope degree, aspect, average annual temperate, drought index, river density, land cover, and distance from roads and residential areas. Using the area under the receiver operating characteristic curve (AUC) and seven other performance metrics, the models were validated in terms of their abilities to elucidate the general fire behaviors in the Pu Mat National Park and to predict future fires. Despite a few differences between the AUC values, the BN model with an AUC value of 0.96 was dominant over the other models in predicting future fires. The second best was the DT model (AUC = 0.94), followed by the NB (AUC = 0.939), and MLR (AUC = 0.937) models. Our robust analysis demonstrated that these models are sufficiently robust in response to the training and validation datasets change. Further, the results revealed that moderate to high levels of fire susceptibilities are associated with ~19% of the Pu Mat National Park where human activities are numerous. This study and the resultant susceptibility

maps provide a basis for developing more efficient fire-fighting strategies and reorganizing policies in favor of sustainable management of forest resources.

Keywords: Bayes network; decision tree; multivariate logistic regression; Naïve Bayes; spatial modeling

1. Introduction

Fires are potentially the most destructive natural disaster in forested areas [1,2] that burn millions of hectares annually [3,4] and are responsible for the loss of biodiversity, soil quality, and CO₂ capture [5]. The susceptibility of the forests and their adjacent areas, i.e., human settlements and infrastructures, to fires is a major concern to the communities in many land ecosystems of the world [6–12]. Increased changes in socioeconomic processes and climate that induced extensive modification of natural environment [13,14] and prolonged drought periods [15–19] have placed strong demands on authorities and decision makers to temporally and spatially delineate the forested areas in terms of susceptibility to fires [6,11,20]. Identifying areas with high/very high fire susceptibility must be undertaken to successfully design fire management plans [21] and allocate firefighting resources [22–26]. To this end, robust approaches and tools are required to enable the managers and engineers to accurately estimate the time, location, and extent of future fires [8,10,12,27–29]. The improvements in techniques for predicting fire susceptibility and delineating the forested areas into different susceptibility levels can help forest managers and policy makers to achieve a better understanding of fires that facilitates the development of prevention measures for the fire-prone forests [4,30].

In forest fire prediction, however, it is difficult to compile sufficient amounts of spatially explicit geo-environmental data, particularly over large-scale forests, due to field survey difficulties and budgetary constraints. Over the past decade, machine learning methods have successfully reached the primacy as a replacement to the traditional field-survey methods for the prediction of forest fire susceptibility by elucidating the relationship between historic fire events and different explanatory variables in order to predict future fires [20]. Examples for machine learning methods suggested and used for forest fire prediction include decision tree based classifiers [31,32], artificial neural network (ANN) [33,34], neuro-fuzzy [27,35–37], and support vector machine [7,33,38].

Despite the widespread application of these methods, many regions of the world have not yet been delineated in terms of fire susceptibility. Further, no single model/method has been yet identified to capture fire behavior in all regions due to the variation of training data from different regions [6,8,10,11,36,39,40]. To fill this significant gap in fire prediction efforts, we aimed this study to develop a suite of predictive models based on the four machine learning methods, namely Bayes Network, Naïve Bayes, Decision Tree, and Multivariate Logistic Regression for the prediction of fire susceptibility in the Pu Mat National Park of Vietnam. Although these methods have been broadly investigated in environmental studies, particularly for the prediction of landslides and floods [41–45], their joint application and comparison have not yet been reported for forest fire prediction. The outcomes from this study allow researchers to determine if a particular predictive model derived from machine learning methods aligns with their objectives for modeling and mapping of forest fire susceptibility.

2. Study Area

The Pu Mat National Park is located in the Nghe An Province in the north-central coast region (18°46' north latitude and 104°24' east longitude) of Vietnam (Figure 1). This park was established on 8 November 2001 and is a part of the Western Nghe An Biosphere Reserve. This park with an area about 94,804 ha spreads in Tuong Duong, Con Cuong, and Anh Son of Nghe An. Out of the total land area, the strictly protected area encompasses about 89.5 ha, the ecological recovery area covers about

1.6 ha, and a buffer zone that comprises about 86.000 ha. The park is located in a region characterized by the tropical monsoon climate. The average annual rainfall is recorded to be 1800 mm. Topography highly controls temperature such that the average annual temperature is 20 °C on the coast, 15 °C in the areas with an altitude of 900 m, 12 °C in the areas with an altitude of 1800 m, and 5 °C in the areas with an altitude of 2700 m. The highest temperatures are recorded in August, often exceeding 35 °C during the day. This park is usually faced with a five-month drought period, which typically extends from April to August. In general, the Pu Mat National Park is a greatly bio-diversified area in Vietnam that has periodically suffered fire damage. To safeguard the biodiversity as well as human settlements from recurrent fire events and to make more informed decisions for fire suppression operations, systematic, and continuous studies, such as the one presented in this paper, are required.

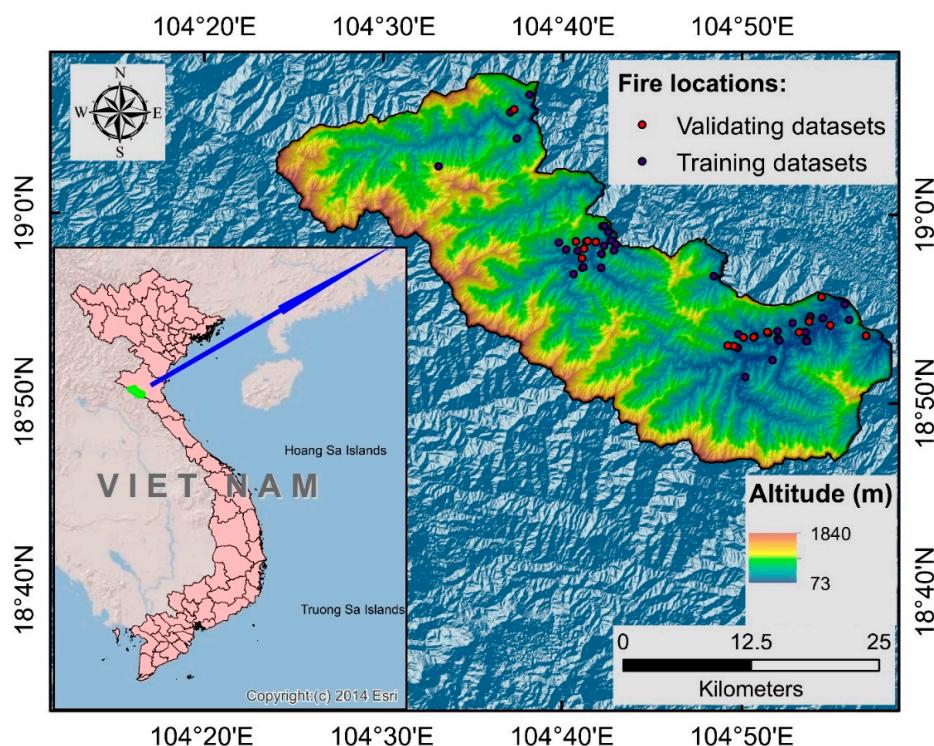


Figure 1. Location map of the Pu Mat National Park (study area) and historical fire events.

3. Data Preparation

3.1. Fire Inventory Map

An inventory map represents the historical fires occurred across the landscape. To prepare the inventory map of the Pu Mat National Park, we used 56 historical fires georeferenced perimeters from the period of 2014–2016. The records for these fires were obtained from the historical archives that were verified via multiple field surveys and observations. These fires usually occurred during the drought period. However, extensive human activities are supposed to intensify the occurrences.

3.2. Explanatory Variables

Another important step in forest fire modeling and mapping is compiling a set of independent explanatory variables known as fire causative factors based on their potential relationship with the local characteristics of the area being investigated, historical fires, and data availability. In this study, we collected nine geo-environmental, climate, and human variables (i.e., elevation, slope degree, aspect, average annual temperature, drought index, river density, land cover, and distance from roads and residential areas) and converted each variable to categorized raster format with a cell size of 30 × 30 m (Figure 2).

Topography-related variables (elevation, slope degree, aspect) were selected due to their relevance to fire occurrence that has been widely demonstrated in the literature [12,24,28,29,36]. Terrain morphology heavily affects human accessibility, species density and composition, and fire behavior [9]. The topography-related variables considered in this study were derived from a 30-m digital elevation model (DEM) of the Pu Mat National Park.

As a hydrological variable, we used river density to quantify the amount of surface water and surrounding humidity within the study area and their influences on fire susceptibility across the study area. River density is the total length of rivers in a drainage basin divided by the total area of the drainage basin. In general, the region with a higher river density has lower sensitivity to fire occurrences [27].

Land cover was another variable that we used for modeling fire occurrence in the Pu Mat National Park. Land cover is a measure of forests, agriculture, wetlands, impervious surfaces, and other land types in a landscape. Land cover is typically used as a proxy for flammability of the landscape [46] for modeling fire probability [47,48]. The land cover map of the Pu Mat National Park was produced using the Landsat satellite images for the year 2016.

The climate-related variables selected for modeling fire susceptibility in the Pu Mat National Park were annual temperature and drought index. Temperature is an important variable for fire prediction because of its effect on the moisture content of the fuel, which in turn is a crucial parameter in fire ignition [49,50]. For this study, the meteorological data corresponding to the 2014–2016 period were used to develop a thematic map of average annual temperature for the Pu Mat National Park. Drought index was another climate-related variable that was used in this study because forest fires have the tandem of occurrence in the region most affected by drought periods [19,51]. Following Karnieli et al. [52], we computed the drought index of the Pu Mat National Park based on the relationship between the normalized difference vegetation index (NDVI) and land surface temperature (LST) as follows:

$$\text{Drought index} = (\text{temperature} - \text{LSTmin}) / (\text{LSTmax} - \text{LSTmin}) \quad (1)$$

$$\text{LSTmin} = -13.98 (\text{NDVI} + \text{smallest M temperature}) \quad (2)$$

$$\text{LSTmax} = -13.98 (\text{NDVI} + \text{maximum M temperature}) \quad (3)$$

$$\text{NDVI} = (\text{NIR} - \text{RED}) / (\text{NIR} + \text{RED}) \quad (4)$$

where NIR and RED are near-infrared and red spectral bands, respectively.

The reason to select this drought index was the availability of data required for calculating this index. Although other types of drought indices (e.g., standardized precipitation index, standardized precipitation evapotranspiration index, vegetation condition index, vegetation condition index, Palmer drought severity index, and temperature condition index) have been used for fire modeling [19], their application needs long-term precipitation and temperature data [15,16] that are unavailable for the Pu Mat National Park.

The literature identified human activities as a main cause of the majority of forest fires [7,39,53,54]. Activities such as picnic fires, shepherd fires, smoking, hunting, stubble burning, and arson have been repeatedly identified as the main cause of fire ignitions in forests worldwide [29]. Previous studies quantified the effects of human activities on fire probability using the proximity variables that incorporate the information related to distance from roads, railways, houses, industrial areas, and airports into the modeling process [24,55,56]. In this study, we elected to use two main proximity variables: distance from roads and distance from residential areas. The information for generating these layers was obtained from topographic maps at the scale of 1:100,000 obtained from the North Central Geological Federation of Vietnam. Table 1 details the main characteristics of the nine variables used in this study.

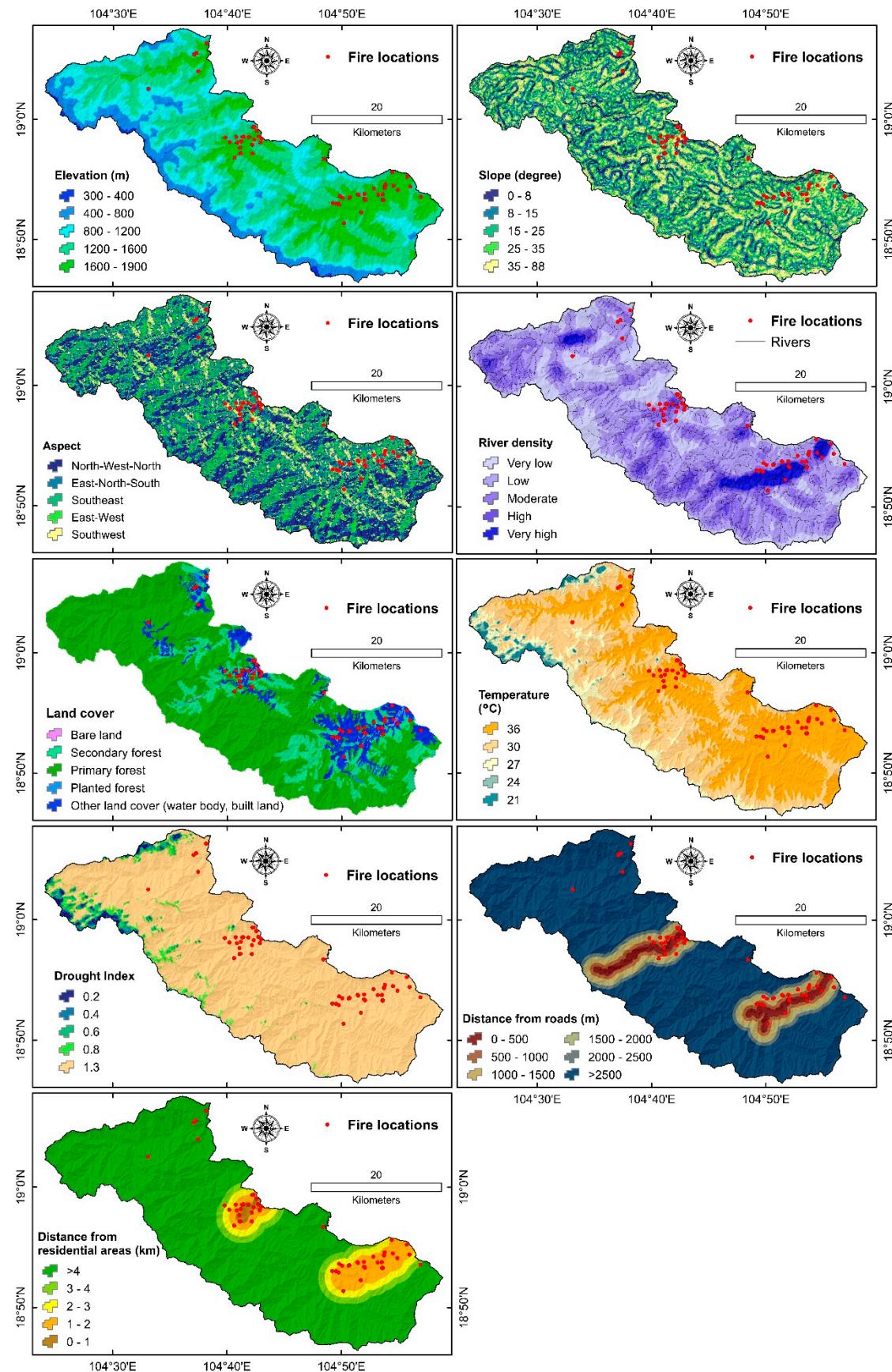


Figure 2. Explanatory variables used in this study.

Table 1. General characteristics of the variables used in this study.

Variable	Source	Scale	Access Date
Slope degree	USGS DEM	30 × 30 m	2015
Elevation (m)	USGS DEM	30 × 30 m	2015
Aspect	USGS DEM	30 × 30 m	2015
River density	USGS DEM	30 × 30 m	2015
Land cover	Landsat ETM+	30 × 30 m	2016
Annual temperature (°C)	VMO	-	2016
Drought index	NDVI and LST	30 × 30 m	2016
Distance from roads (m)	NCGFV and GEI	1:100,000	2015
Distance from residential areas (m)	NCGFV and GEI	1:100,000	2015

USGS DEM: digital elevation model obtained from the United States Geological Survey, VMO: Vietnam Meteorological Organization, NDVI: normalized difference vegetation index, LST: Land surface temperature, NCGFV: North Central Geological Federation of Vietnam, GEI: Google Earth images.

4. Methods

4.1. Relief-F Feature Selection Method

Selection of the most influential explanatory variables is a crucial step in a modeling task using machine learning that allows modelers to efficiently focus on those variables that better explain input–output interactions and contribute the most to the modeling process. Feature selection aims to remove irrelevant and redundant features toward introducing a small number of features that define the dataset better than the original set of features. In a modeling study using the machine learning methods, this can be achieved by measuring the importance of each variable for obtaining a higher accuracy in classification. One of the well-known methods in variable selection is the Relief-F method that was originally developed by Kira and Rendell [57] and then upgraded by Kononenko [58]. The original Relief algorithm can detect the conditional dependencies between attributes for feature selection, but its function is restricted to two class issues only. Moreover, it is not handling incomplete, noisy, and duplicate attributes in the dataset. However, the renewed Relief-F algorithm deals with multi class. The Relief-F method is a widely used feature selection method in the literature due to its simplicity and efficiency for variable ranking [59]. Theoretically, this algorithm ranks different features in terms of their utility for the problem being modeled and determines the most efficient features for the prediction task. In fire modeling, Relief-F measures the spatial associations between fire locations and different causative factors to calculate the average merit (AM) of each causative factor in separating fire-prone and fire-proof portions of the landscape.

4.2. Bayes Network (BN)

BN [60] is a probabilistic, statistical model that forms a set of random variables and their conditional dependencies (Bayes law) within an annotated directed acyclic graph. BN is a promising tool for explaining the relationships between an event and several possible explanatory variables. Structurally, the BN classifier is a directed acyclic graph where the arcs have a formal interpretation of probabilistic conditional independence. The quantitative part of this graph is a collection of conditional probability tables, each attached to a node that represents the probability of the variable at the node conditioned on its parents in the network. One of the important advantages of BN is that this method handles risk analysis and uncertainty assessment more accurately than the other models that only predict values. Managing missing values between input data, the ability to combine quantitative and qualitative data, and providing approximate solutions using simulation techniques or estimation methods in cases where a precise solution is not available are among the advantages of this method [61]. Bayes theory enables forward and backward computation, which means that in addition to predicting the target variable using the state of the input variables it is able to determine the effect of each of the input variables on the outputs of the model with the predicted variable status [62].

4.3. Naïve Bayes (NB)

NB classifier is a simple supervised function and is special form of discriminant analysis. NB is a member of the probability-based clustering family that utilizes the Bayesian theorem and assumes independence between variables to perform a classification task. Bayesian classification technique is typically used as a simple way to classify and label the objects or points. Although the NB classifier has some drawbacks (e.g., low performance or biased estimation of prior probability) due to its basic assumptions on variable relationships, this method has been proven to work efficiently for many real-world problems [63,64]. To apply NB for forest fire modeling, suppose $X = (x_1, x_2, \dots, x_n)$ is a vector of n properties that are independent explanatory variables. Thus, the probability of fire occurrence ($p(C_k|x_1, \dots, x_n)$) is represented as one of the states of the class of different events for different K s:

$$P(C_k|X) = \frac{P(C_k)P(X|C_k)}{P(X)} \quad (5)$$

NB can be used for both binary and multi-class classification problems. The NB classifier is very useful in high-dimensional problems.

4.4. Decision Tree (DT)

DT is a non-parametric, supervised learning method designed for classification and prediction problems. It is easy to interpret (due to the tree structure) and has a Boolean function (if each decision is binary, i.e., false or true). Decision trees extract predictive information in the form of human-understandable tree-rules (If/Then rules). Each decision in the tree can be seen as a feature. To make a prediction using decision trees, a tree-like structure is designed that first starts with all training samples and selects the variable that best fits the class and makes subdirectories [65]. The tree branches are the result of a test performed at each step by the algorithm on the middle nodes. Predictions also appear on the leaves of the tree. The split criterion in a node is based on the standard deviation of the output values that reach that node as a measure of error. By testing each attribute (parameter) in the node the expected decrease in error is calculated [66]. The reduction of standard deviation is calculated with the relation

$$SDR = \frac{m}{|T|} \times \beta(i) \times \left[sd(T) - \sum_{j \in (L,R)} \frac{|T_j|}{|T|} \times sd(T_j) \right] \quad (6)$$

where SDR is the decrease standard deviation, T is the series of samples that reach the node, m is the number of samples that have no missing values for this parameter, $\beta(i)$ is a correction factor, and TL and TR are sets created by dividing on this parameter.

4.5. Multivariate Logistic Regression (MLR)

In a regression model, an equation for predicting the values of the dependent variable based on one or more independent predictor variables is developed. Dependent variable (e.g., occurrence or non-occurrence of a fire) is a two-state qualitative variable that takes the value of 1 or 0. In fire probability modeling, the objective of MLR is to find the best model to describe the relationships between the occurrence or non-occurrence of a fire (i.e., dependent variable) and a set of independent variables known as fire influencing factors [47]. The general form of the logistic regression equation can be given as follows:

$$P = \frac{1}{1 + e^{-z}} \quad (7)$$

where P is the probability of fire occurrence and Z is the occurrence (1) or non-occurrence (0) of a fire event that is expressed by

$$Z = b_0 + b_1 \times x_1 + b_2 \times x_2 + \dots + b_n \times x_n \quad (8)$$

where b_0 is the intercept of the equation, b_i ($i = 0, 1, 2, \dots, n$) are the model coefficients, and x_i ($i = 0, 1, 2, \dots, n$) are the fire explanatory variables.

4.6. Validation Metrics

One of the most important steps after developing a model is to evaluate its training and predictive performance [18,67,68]. In this study, we used receiver operating characteristics (ROC) and several statistical measures (true positive (TP), true negative (TN), false positive (FP), false negative (FN), positive predictive value (PPV), negative predictive value (NPV), sensitivity (SST), specificity (SPF), accuracy (ACC), Kappa, and root mean square error (RMSE)) for the evaluation and comparison of the models developed for fire probability mapping. The following subsections provide a brief description of each metric.

4.6.1. Receiver Operating Characteristics (ROC)

The receiver operating characteristics (ROC) curve is one of the most important and widely used performance metrics for the evaluation of classification models in terms of their goodness-of-fit and generalizability [69–71]. This method is a probability-based curve that can measure models at different thresholds [72]. ROC curve represents a trade-off between sensitivity on the y -axis and 1-specificity on the x -axis. A model with an excellent performance achieves the area under the ROC curve (AUC) of >90 [73].

4.6.2. Statistical Metrics

The statistical metrics used for machine learning evaluation are categorized into three main groups of metrics, including specifically threshold, probability, and ranking metrics [74]. Threshold and ranking metrics are the most widely used metrics [75]. For this study, we opted to use the following five established and applicable threshold metrics to evaluate the BN, NB, DT, and MLP models: specificity (SPF), sensitivity (SST), accuracy (ACC), Kappa, and root mean square error (RMSE). Using these performance metrics, we investigated how well the different models used for the prediction of forest fire susceptibility captured the relationships between historical fires and different explanatory variables (i.e., goodness-of-fit with the training dataset) and made decisions when tested with the unseen validation dataset (i.e., generalization ability). We evaluated the goodness-of-fit and generalization ability of the models based on four components (i.e., true positive (TP), true negative (TN), false positive (FP), and false negative (FN)) of a 2×2 confusion matrix. TP and TN are the numbers of fires that are correctly classified as, respectively, fires and non-fires. FP and FN are the numbers of non-fires that are incorrectly classified as fires and non-fires [47,76]. The SPF, SST, ACC, Kappa, and RMSE are calculated as follows:

$$\text{SPF} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (9)$$

$$\text{SST} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad (10)$$

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (11)$$

$$\text{Kappa} = \frac{(\text{TP} + \text{TN}) - ((\text{TP} + \text{FN})(\text{TP} + \text{FP}) + (\text{FP} + \text{TN})(\text{FN} + \text{TN}))}{1 - ((\text{TP} + \text{FN})(\text{TP} + \text{FP}) + (\text{FP} + \text{TN})(\text{FN} + \text{TN}))} \quad (12)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\text{X}_{\text{obs}} - \text{X}_{\text{est}})^2}{N}} \quad (13)$$

where X_{obs} is the observations (i.e., validation dataset), and X_{est} is the estimated values by the forest fire predictive models.

5. Modeling Methodology

The flowchart of the methodology proposed for the prediction of forest fire susceptibility in the Pu Mat National Park is shown in Figure 3. The methodology starts by compiling a set of explanatory variables and generating an inventory map of the locations of historical fires. The historical fires were randomly allocated to two different sets: training dataset that contained 40 forest fire locations (70%) and validation dataset that included the remaining 17 forest fire locations (30%) [15,67,69,77,78]. To construct the final datasets, an equal number of non-fire points was randomly sampled from non-fire portions of the Pu Mat National Park. We coded the fire points as “1”, whereas the non-fire points were coded as “0” [79–82]. This process yielded training and validation datasets that consisted of 80 and 34 samples, respectively. Then, the dataset pre-processing was carried out using the Relief-F feature selection method to identify the variables with null predictive usefulness [31,68,70,83]. Through the modeling step, the machine learning methods were trained using the training dataset to develop forest fire predictive models. To check for the model robustness, a five-fold cross-validation procedure that produced five different folds of training and validation datasets was used [32,84,85]. In this procedure, one group out of the five groups was used as the validation dataset and the rest were used as the training dataset. Then, the models were trained using the training sets and validated using the validation dataset. This modeling process was repeated until each one of the five groups was used as the validation dataset. The ultimate outcomes of the modeling process were four distribution maps of forest fire susceptibility that were quantitatively analyzed [48] and compared to each other.

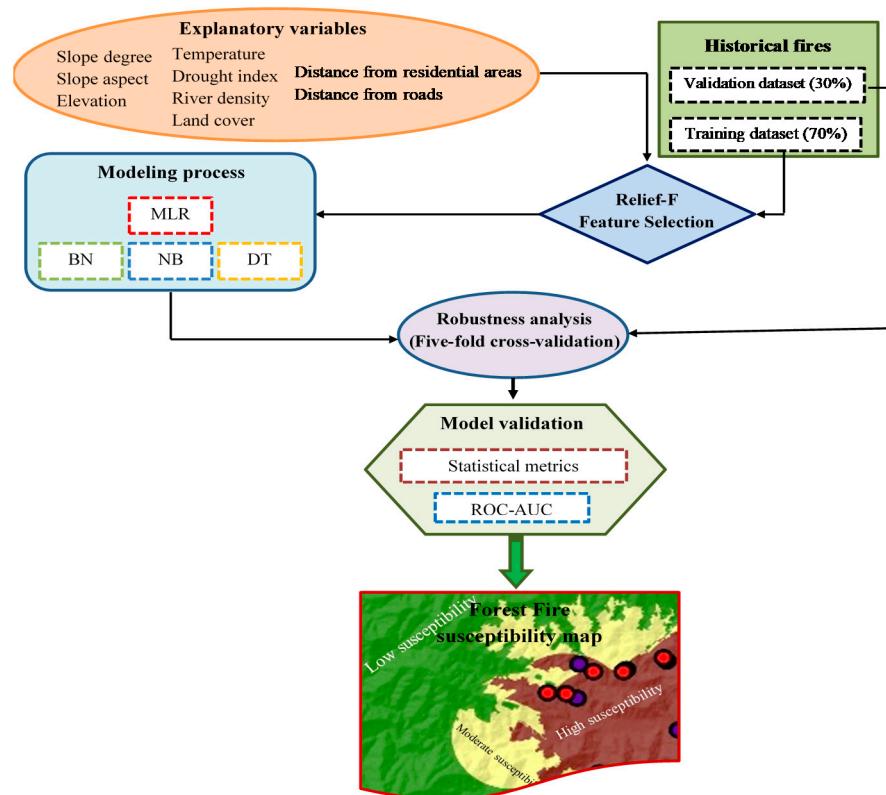


Figure 3. Flowchart of the methodology proposed for forest fire modeling and mapping.

6. Results and Discussions

6.1. Variable Importance

The results of the Relief-F ranking of fire explanatory variables based on their AM showed that distance from roads had the highest influence on fire occurrences in the Pu Mat National Park (Table 2). It means that greatest AM (85.9) was obtained for this variable, followed by distance from

residential areas (83.4), land cover (79.5), elevation (74.4), and annual temperature (71.8). Afterward, aspect, river density, slope degree, and drought index had the lowest AM equal to 56.5, 55.1, 53.8, and 48.7, respectively. These results revealed that the human-related variables (distance from roads and residential areas, and land cover) were the most influential variables, corroborating previous studies in Vietnam that reported on the significance of human activities on increasing the probability of fire occurrences [25,27,86]. Some other studies suggested that proximity to roads and residence areas intensifies the likelihood and frequency of fire ignitions, even in regions with a relatively low population density [20,87]. In a recent study, Elia, Giannico, Spano, Laforteza, and Sanesi [56] demonstrated the significance of distance to roads on increasing fire probabilities in more urbanized Mediterranean regions of southern Italy. In contrast, Gralewicz et al. [88] showed the declined probability of fire occurrence close to urbanized regions of Canada due to the policy of fully suppressing all wildfires.

Land cover was ranked among the most influential factors causing fire occurrence in the Pu Mat National Park. Several studies have documented how a specific type (e.g., grasslands) of land cover is closely related to fire events, whereas some others (e.g., farmlands and orchards) are negatively related [89,90]. Nunes et al. [91] demonstrated that fires are selective for land cover such that they prefer specific land cover types. While there was a marked preference for shrubland and forest cover types, farmlands were clearly avoided. Our results revealed that fires are highly correlated with those portions of the Pu Mat National Park that experienced afforestation and urbanization, while natural forests are obviously fire-proof.

Although the Pu Mat National Park suffered recurrent prolonged drought occurrences, it seems that human-related variables are much stronger than the climate-related variables for fire ignition.

Since the AM of all nine explanatory variables was greater than zero, the spatial modeling was performed using all factors [67,68,70,77].

Table 2. Variable importance measured using the Relief-F method.

Rank	Variable	AM
1	Distance from roads	85.9
2	Distance from residential areas	83.4
3	Land cover	79.5
4	Elevation	74.4
5	Annual temperature	71.8
6	Aspect	56.5
7	River density	55.1
8	Slope degree	53.8
9	Drought index	48.7

6.2. Model Validation and Comparison

To validate the models and compare them to each other in terms of training and validation performances, we computed several performance metrics processing both training and validation datasets (Table 3). Regarding the PPV metric that is the proportion of correctly classified fire samples out of all samples classified as fire samples, the BN model with $PPV_{training} = 89.74\%$ and $PPV_{validation} = 100\%$ performed the best. In terms of the NPV metric that is the proportion of samples that were correctly classified as non-fire, the DT and MLR models with the values equal to 100% were identified as the best models. Regarding the SST metric that measured the models' abilities to predict a proportion of all fire samples as fire (i.e., true positives), the DT and MLR models with the values equal to 100% were dominant over the other models. In terms of the SPF metric that measured the models' abilities to predict a proportion of all non-fire samples as non-fire (i.e., true negatives), the BN model with $PPV_{training} = 89.47\%$ and $PPV_{validation} = 100\%$ was the best model. In terms of the ACC metric that measured the overall models' efficiencies, the MLR ($ACC = 92.31\%$) and BN ($ACC = 94.12\%$) were the most efficient models in the training phase and validation phase, respectively. Regarding the Kappa index, the MLR ($Kappa = 0.846$) and BN ($Kappa = 0.884$) showed perfect agreement between observed fires and predicted fires in the training phase and validation phase, respectively. These variant training

and validation performances that have been also previously observed in different models used for different applications [24,67–69,71,77,78,83,92] can be attributed to the specific nature and structure the models applied to different datasets. These results underscore the conclusion drawn by Bui, Khosravi, Tiefenbacher, Nguyen and Kazakis [84] that no model exists that always performs the best for all datasets from different sources.

Table 3. Model performance in the training and validation datasets.

Metric	Training Dataset				Validation Dataset			
	BN	DT	MLR	NB	BN	DT	MLR	NB
PPV (%)	89.74	82.05	84.62	87.18	100.00	64.71	76.47	94.12
NPV (%)	87.18	100.00	100.00	87.18	88.24	100.00	100.00	94.12
SST (%)	87.50	100.00	100.00	87.18	89.47	100.00	100.00	94.12
SPF (%)	89.47	84.78	86.67	87.18	100.00	73.91	80.95	94.12
ACC (%)	88.46	91.03	92.31	87.18	94.12	82.35	88.24	94.12
Kappa	0.769	0.821	0.846	0.744	0.884	0.647	0.765	0.882

In the matter of the magnitude of the modeling error, the four models exhibited training error that ranged from 0.255 (MLR) to 0.339 (NB) and validation error that ranged from 0.192 (BN) to 0.306 (DT) (Figure 4). Again, we are inclined to attribute these asymmetric performances of a model in training and validation phases to its computational algorithm when tested with different datasets.

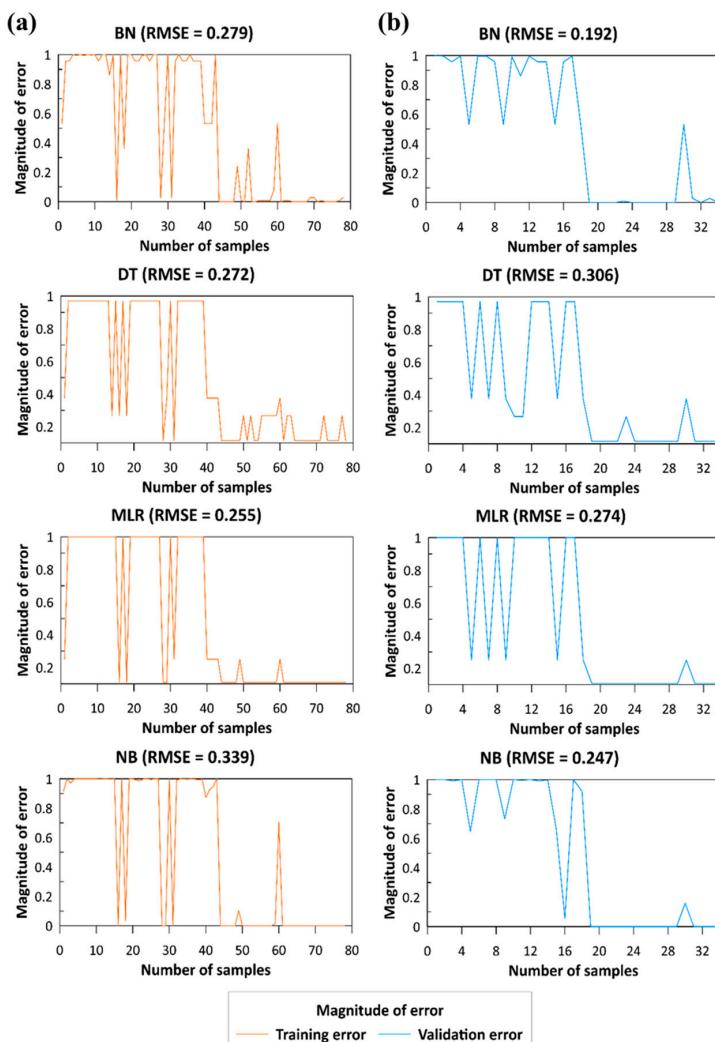


Figure 4. Magnitude of training error (a) and validation error (b) for the models.

The AUC values for the BN, DT, MLR, and NB models obtained from the training phase processing the training dataset were 0.99, 0.969, 0.986, and 0.983, respectively (Figure 5a). Based on these values, all four models performed excellently in distinguishing between training samples (fires and non-fires) with respect to the explanatory variables, although the BN model performed slightly better than the others.

The AUC values obtained from the validation phase that processed the validation dataset exhibited values of 0.96, 0.94, 0.937, and 0.939 for the BN, DT, MLR, and NB models, respectively (Figure 5b). Based on these results and the interpretation of AUC ranges given in the literature [73,93], we can say all four models developed in this study have a very high ability to predict future fire occurrences. In contrast to our results, previous studies have found that the LR models are often outperformed by other models for fire prediction [7,34,87,94].

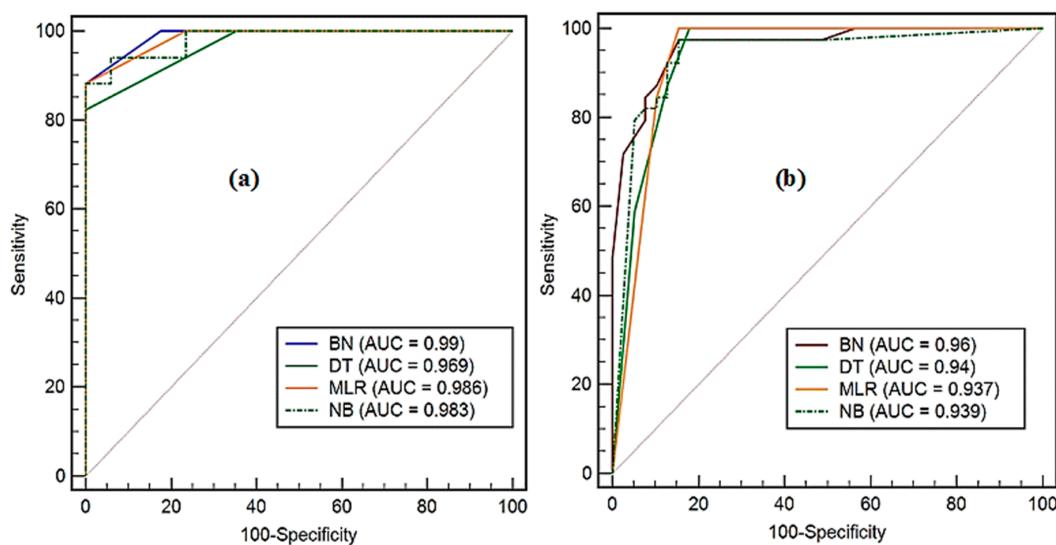


Figure 5. Receiver operating characteristics (ROC) curves and area under the receiver operating characteristic curve (AUC) values of the models in the (a) training and (b) validation datasets.

6.3. Robustness Analysis

The analysis of the model robustness based on the five different datasets (Fold 1–5) and three performance metrics (ACC, RMSE, and AUC) showed that the models were very stable, and their performance changed in a narrow range (Table 4). For example, the training phase of the BN model ranged ACC from 87.17% to 88.46% (mean = 87.44% and standard deviation = 0.57%), ranged RMSE from 0.279 to 0.301 (mean = 0.29 and standard deviation = 0.01), and ranged AUC from 0.98 to 0.99 (mean = 0.98 and standard deviation = 0.00). Further, the validation phase of this model ranged ACC from 99.85 to 100% (mean = 99.90% and standard deviation = 0.06%), ranged RMSE from 0.192 to 0.31 (mean = 0.28 and standard deviation = 0.05), and ranged AUC from 0.941 to 0.96 (mean = 0.96 and standard deviation = 0.01). Overall, these results revealed that four models used in this study are reliable and robust in response to training and validation data sets change. Our results are supported by previous works that reported on the reliability and robustness of machine learning methods for environmental studies [32,35,36,59,68,95] as well as for other real-world problem [96,97].

Table 4. Robustness analysis using five-fold cross-validation.

Model	Phase	Metric	Fold					Mean	SD
			1	2	3	4	5		
BN	Training	ACC	88.46	87.18	87.18	87.18	87.18	87.44	0.57
		RMSE	0.279	0.287	0.285	0.299	0.301	0.29	0.01
		AUC	0.99	0.984	0.98	0.98	0.98	0.98	0.00
	Validation	ACC	100	99.88	99.88	99.88	99.85	99.90	0.06
		RMSE	0.192	0.31	0.291	0.296	0.286	0.28	0.05
		AUC	0.96	0.954	0.965	0.941	0.956	0.96	0.01
DT	Training	ACC	91.03	89.99	90.87	89.62	89.9	90.28	0.63
		RMSE	0.272	0.306	0.267	0.325	0.321	0.30	0.03
		AUC	0.969	0.953	0.958	0.947	0.949	0.96	0.01
	Validation	ACC	94.12	94.12	93.18	93.01	93.18	93.52	0.55
		RMSE	0.306	0.307	0.296	0.302	0.298	0.30	0.00
		AUC	0.94	0.94	0.94	0.934	0.94	0.94	0.00
MLR	Training	ACC	92.31	91.9	92.9	89.9	89.18	91.24	1.61
		RMSE	0.255	0.35	0.344	0.352	0.34	0.33	0.04
		AUC	0.986	0.96	0.97	0.974	0.959	0.97	0.01
	Validation	ACC	88.24	87.06	90.18	88.14	88.14	88.35	1.13
		RMSE	0.274	0.203	0.295	0.306	0.299	0.28	0.04
		AUC	0.937	0.935	0.93	0.933	0.938	0.93	0.00
NB	Training	ACC	87.18	87.18	87.18	87.18	87.18	87.18	0.00
		RMSE	0.339	0.339	0.335	0.351	0.347	0.34	0.01
		AUC	0.983	0.983	0.979	0.979	0.979	0.98	0.00
	Validation	ACC	94.12	93.18	93.24	93.18	93.18	93.38	0.41
		RMSE	0.274	0.299	0.315	0.297	0.256	0.29	0.02
		AUC	0.939	0.937	0.932	0.933	0.932	0.93	0.00

6.4. Forest Fire Susceptibility Maps

Four maps were generated to depict the forest fire susceptibility predicted by BN, DT, MLR, and NB models (Figure 6). For each map, the probability of fire occurrence was classified into three levels of susceptibility from low to high using the natural breaks classification method. A quantitative analysis of the produced maps revealed that the lowest susceptibility level output from the NB model applied to 81.28% of the land area, whereas the moderate and high covered 4.52% and 14.2% of the area, respectively (Figure 7a). The MLR model classified the study at 81.54, 5.93, and 12.53% for the low, moderate, and high susceptibilities, respectively. The DT model classified 77.99% low susceptibility, 16.68% moderate susceptibility, and 5.33% high susceptibility. The BN model classifications were 82.68%, 8.51%, and 8.81% for the low, moderate, and high susceptibilities, respectively. On average, the models classified 80.9%, 8.9%, and 10.2% of the Pu Mat Natural Park into low, moderate, and high susceptibility to fire occurrences. Given the centralized moderate and high susceptibility classes around areas where human stakes are numerous (i.e., roads and residential areas), we can conclude that anthropogenic pressures transformed 19.1% of the study area into a susceptible zone to the future fires. The reliability of the forest fire susceptibility maps was assessed using frequency ratio analysis (Figure 7b,c), which donates the ratio between the percentage of actual fires and the percentage of the entire area for each susceptibility zone. In all four maps, the highest frequency ratio values belonged to the high susceptibility classes, followed by moderate and low classes for all produced maps. This indicates that all models performed well in delineating the Pu Mat National Park regarding the historical fire locations [47,48]. Despite these promising results, there might be several uncertainties

in such susceptibility maps. One possible source of uncertainty is the edge effect, which happens when some fires ignited outside the study area spread to the study area and may alter the level of susceptibility near the boundaries [98,99]. Since no information is available for fires spreading from the other areas to the Pu Mat National Park, we failed to analyze the potential edge effect in this study. When the required information is available, the application of the edge detection methods [98] can help researchers ensure that edge effected regions are identified and removed.

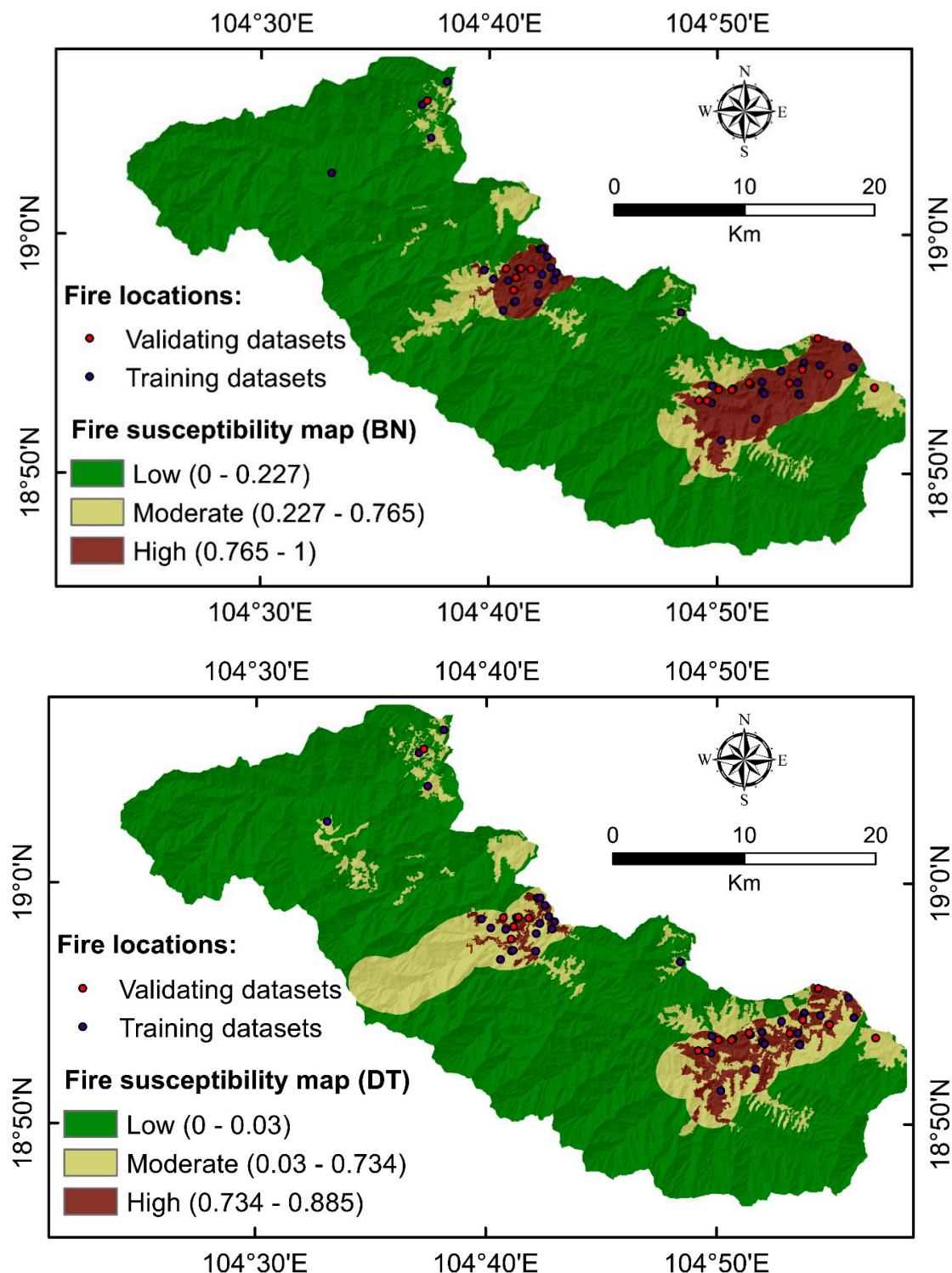


Figure 6. Cont.

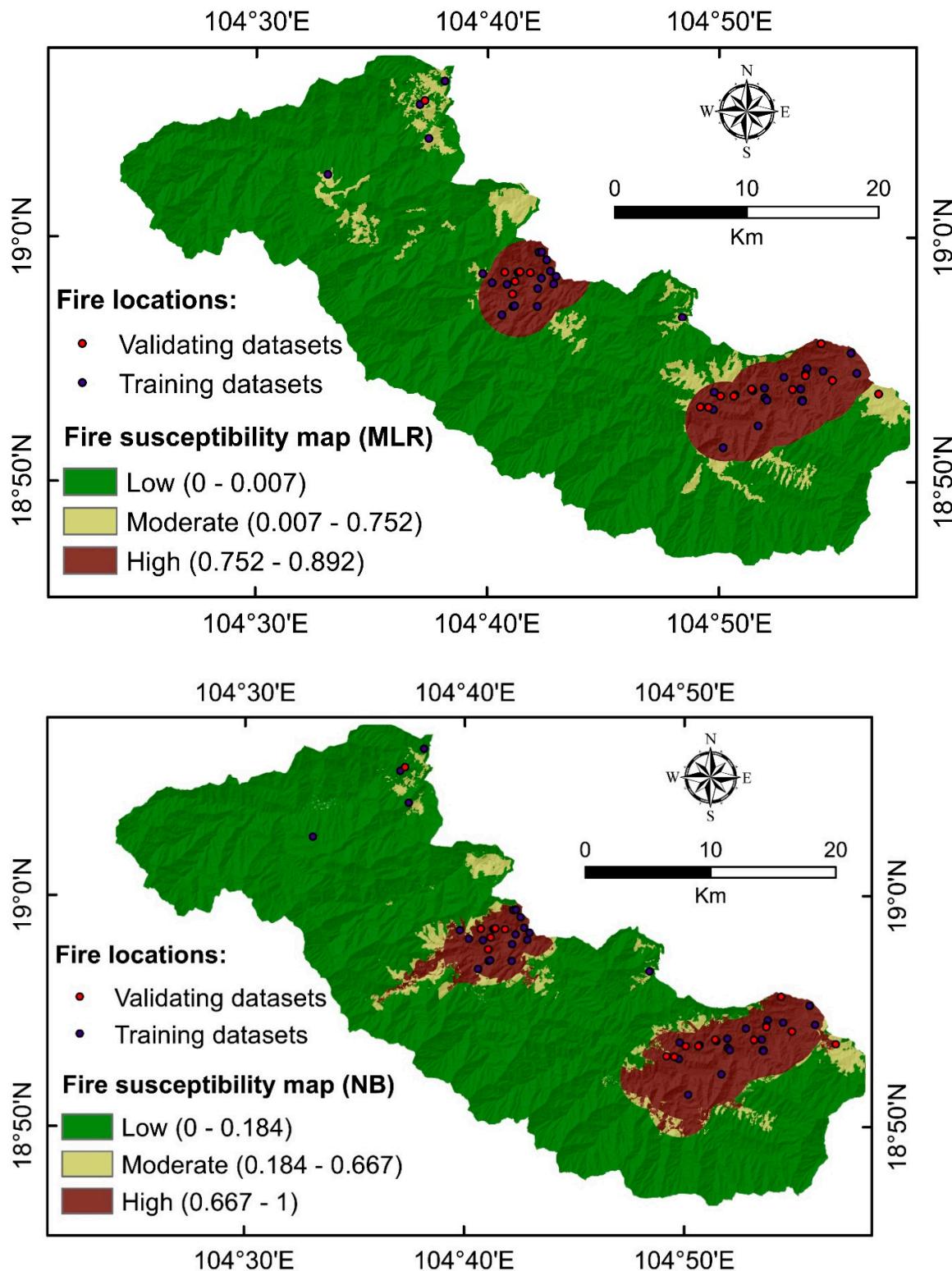


Figure 6. Fire susceptibility maps derived by the four models.

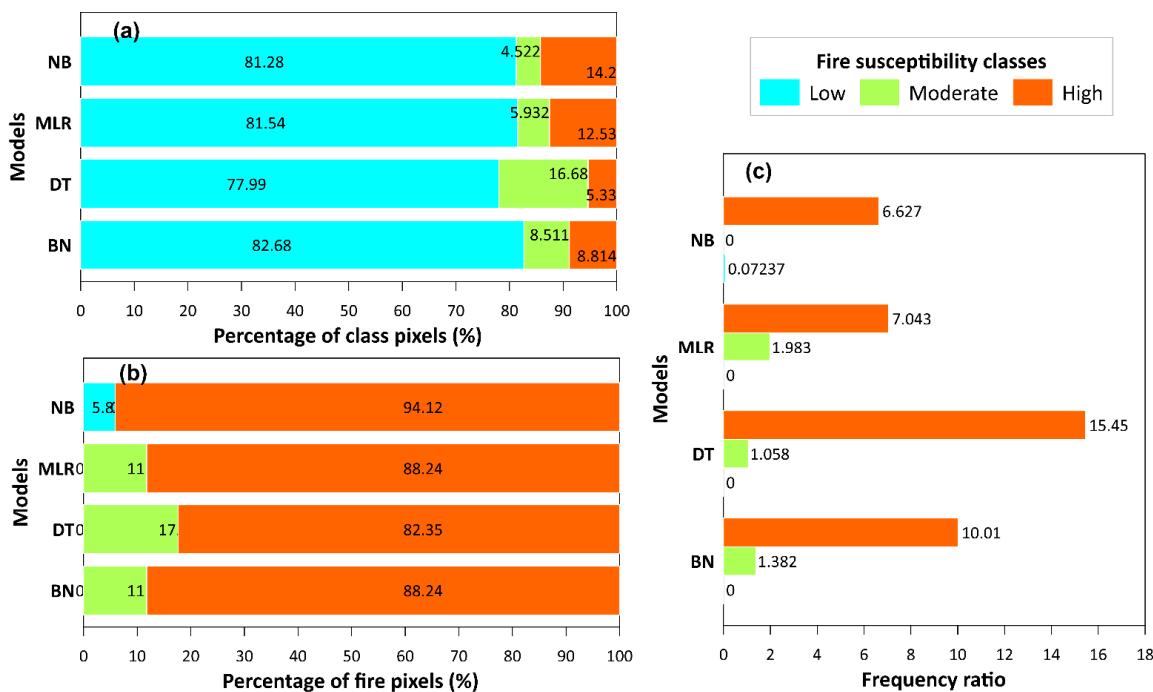


Figure 7. Quantitative analysis of the fire susceptibility maps: (a) Percentage of class pixels; (b) Percentage of fire pixels; (c) Frequency ratio analysis.

7. Conclusions

The accurate prediction of fire probability aids forest managers in drafting more efficient fire-fighting strategies and also helps to reorganize policies for sustainable management of forest resources. To achieve these, we evaluated and compared four fire predictive models derived from the BN, NB, DT, and MLR machine learning methods for predicting and mapping fire susceptibility in the Pu Mat National Park, Vietnam. We formulated our modeling methodology based on processing the information from the historical fires and a set of spatially explicit explanatory variables. The outcome of the ROC-AUC method and several other performance metrics revealed that all four models developed in this study had high accuracy in predicting future fire susceptibilities ($AUC > 0.90$) in the Pu Mat National Park, although the BN model performed slightly better than the others. Given the similar performance of these models, perhaps the most remarkable difference between these four models is in interpretability. Managers and decision makers prefer a model and value its outputs if they have some understanding of how the model yielded such the outputs, so the poor interpretability of a machine learning model may restrict its application in practice. From our experience, it is much easier to interpret the MLR model than the complex algorithmic-based Bayesian and DT models. Therefore, the outcomes from our study provide several implications for the selection of a specific model over the others. Although the explanatory variables of the models are not expected to change significantly over time and the results could thus be seen as a long-term prediction of fire susceptibility for the Pu Mat National Park, human activities that change land use patterns would render our long-term susceptibility estimates obsolete, making it necessary to regularly update the current susceptibility maps.

Our findings also demonstrated that fires affect ~19% of the study area, where human activities are numerous and fall within the moderate to high susceptibility to fire occurrence mainly because of the increased road developments and forest-human interfaces, underscoring the need to careful attention from the managers to avoid catastrophes in these portions of the area. Although we achieved a high level of prediction accuracy using the current dataset and four machine learning models, future studies could incorporate other explanatory variables (e.g., vegetation type and density, wind speed and direction, and different types of socio-economic factors and drought indices) into the modeling process in favor of better explaining fire behavior in the Pu Mat National Park.

Author Contributions: Conceptualization, B.T.P. and A.J.; Data curation, B.T.P., T.D.D., H.P.H.Y., T.V.P., D.H.N., H.V.L., H.T.T. and T.T.T.; Formal analysis, A.J.; Funding acquisition, N.A.-A., H.V.L. and T.T.T.; Methodology, B.T.P., A.J., T.D.D., H.P.H.Y., T.V.P., D.H.N., H.V.L., D.M.-G., H.T.T. and T.T.T.; Supervision, B.T.P., N.A.-A., H.V.L. and T.T.T.; Writing—original draft, all authors.; Writing—review and editing, B.T.P., A.J., M.A., H.V.L., D.M.-G., I.P. and H.T.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 105.08-2019.03.

Acknowledgments: We thank three anonymous reviewers for very thoughtful comments that helped improve this paper.

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

References

1. Bowman, D.M.; Balch, J.K.; Artaxo, P.; Bond, W.J.; Carlson, J.M.; Cochrane, M.A.; D’Antonio, C.M.; DeFries, R.S.; Doyle, J.C.; Harrison, S.P. Fire in the Earth system. *Science* **2009**, *324*, 481–484. [[CrossRef](#)]
2. Boer, M.M.; de Dios, V.R.; Bradstock, R.A. Unprecedented burn area of Australian mega forest fires. *Nat. Clim. Chang.* **2020**, *10*, 171–172. [[CrossRef](#)]
3. Meng, Y.; Deng, Y.; Shi, P. Mapping forest wildfire risk of the world. In *World Atlas of Natural Disaster Risk*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 261–275.
4. Stephens, S.L.; Burrows, N.; Buyantuyev, A.; Gray, R.W.; Keane, R.E.; Kubian, R.; Liu, S.; Seijo, F.; Shu, L.; Tolhurst, K.G. Temperate and boreal forest mega-fires: Characteristics and challenges. *Front. Ecol. Environ.* **2014**, *12*, 115–122. [[CrossRef](#)]
5. Bo, M.; Mercalli, L.; Pognant, F.; Berro, D.C.; Clerico, M. Urban air pollution, climate change and wildfires: The case study of an extended forest fire episode in northern Italy favoured by drought and warm weather conditions. *Energy Rep.* **2020**, *6*, 781–786. [[CrossRef](#)]
6. Adab, H.; Atabati, A.; Oliveira, S.; Gheslagh, A.M. Assessing fire hazard potential and its main drivers in Mazandaran province, Iran: A data-driven approach. *Environ. Monit. Assess.* **2018**, *190*, 670. [[CrossRef](#)] [[PubMed](#)]
7. Rodrigues, M.; de la Riva, J. An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environ. Model. Softw.* **2014**, *57*, 192–201. [[CrossRef](#)]
8. Satir, O.; Berberoglu, S.; Donmez, C. Mapping regional forest fire probability using artificial neural network model in a Mediterranean forest ecosystem. *Geomat. Nat. Hazards Risk* **2016**, *7*, 1645–1658. [[CrossRef](#)]
9. Syphard, A.D.; Radeloff, V.C.; Keuler, N.S.; Taylor, R.S.; Hawbaker, T.J.; Stewart, S.I.; Clayton, M.K. Predicting spatial patterns of fire on a southern California landscape. *Int. J. Wildland Fire* **2008**, *17*, 602–613. [[CrossRef](#)]
10. Tien Bui, D.; Le, H.V.; Hoang, N.D. GIS-based spatial prediction of tropical forest fire danger using a new hybrid machine learning method. *Ecol. Inform.* **2018**, *48*, 104–116. [[CrossRef](#)]
11. Viedma, O.; Urbeta, I.; Moreno, J. Wildfires and the role of their drivers are changing over time in a large rural area of west-central Spain. *Sci. Rep.* **2018**, *8*, 17797. [[CrossRef](#)]
12. 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](#)]
13. Mendoza-Ponce, A.; Corona-Núñez, R.O.; Galicia, L.; Kraxner, F. Identifying hotspots of land use cover change under socioeconomic and climate change scenarios in Mexico. *Ambio* **2019**, *48*, 336–349. [[CrossRef](#)] [[PubMed](#)]
14. Martinho, V.J.P.D. Socioeconomic Impacts of Forest Fires upon Portugal: An Analysis for the Agricultural and Forestry Sectors. *Sustainability* **2019**, *11*, 374. [[CrossRef](#)]
15. Mafi-Gholami, D.; Zenner, E.K.; Jaafari, A. Mangrove regional feedback to sea level rise and drought intensity at the end of the 21st century. *Ecol. Indic.* **2020**, *110*, 105972. [[CrossRef](#)]
16. Mafi-Gholami, D.; Zenner, E.K.; Jaafari, A.; Ward, R.D. Modeling multi-decadal mangrove leaf area index in response to drought along the semi-arid southern coasts of Iran. *Sci. Total Environ.* **2019**, *656*, 1326–1336. [[CrossRef](#)]
17. Mafi-Gholami, D.; Zenner, E.K.; Jaafari, A.; Bakhtiari, H.R.; Tien Bui, D. Multi-hazards vulnerability assessment of southern coasts of Iran. *J. Environ. Manag.* **2019**, *252*, 109628. [[CrossRef](#)]
18. Mafi-Gholami, D.; Zenner, E.K.; Jaafari, A.; Bui, D.T. Spatially explicit predictions of changes in the extent of mangroves of Iran at the end of the 21st century. *Estuar. Coast. Shelf Sci.* **2020**, *237*, 106644. [[CrossRef](#)]

19. Parente, J.; Amraoui, M.; Menezes, I.; Pereira, M. Drought in Portugal: Current regime, comparison of indices and impacts on extreme wildfires. *Sci. Total Environ.* **2019**, *685*, 150–173. [[CrossRef](#)]
20. Jaafari, A.; Gholami, D.M.; Zenner, E.K. A Bayesian modeling of wildfire probability in the Zagros Mountains, Iran. *Ecol. Inform.* **2017**, *39*, 32–44. [[CrossRef](#)]
21. Sakellariou, S.; Tampeki, S.; Samara, F.; Flannigan, M.; Jaeger, D.; Christopoulou, O.; Sfougaris, A. Determination of fire risk to assist fire management for insular areas: The case of a small Greek island. *J. For. Res.* **2019**, *30*, 589–601. [[CrossRef](#)]
22. Dennison, P.E.; Brewer, S.C.; Arnold, J.D.; Moritz, M.A. Large wildfire trends in the western United States, 1984–2011. *Geophys. Res. Lett.* **2014**, *41*, 2928–2933. [[CrossRef](#)]
23. Goleiji, E.; Hosseini, S.M.; Khorasani, N.; Monavari, S.M. Forest fire risk assessment—an integrated approach based on multicriteria evaluation. *Environ. Monit. Assess.* **2017**, *189*, 612. [[CrossRef](#)] [[PubMed](#)]
24. Jaafari, A.; Mafi-Gholami, D.; Pham, B.T.; Tien Bui, D. Wildfire probability mapping: Bivariate vs. multivariate statistics. *Remote Sens.* **2019**, *11*, 618. [[CrossRef](#)]
25. Le, T.H.; Nguyen, T.N.T.; Lasko, K.; Ilavajhala, S.; Vadrevu, K.P.; Justice, C. Vegetation fires and air pollution in Vietnam. *Environ. Pollut.* **2014**, *195*, 267–275. [[CrossRef](#)] [[PubMed](#)]
26. 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](#)]
27. Tien Bui, D.; 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. [[CrossRef](#)]
28. Tien Bui, D.; Hoang, N.D.; Samui, P. Spatial pattern analysis and prediction of forest fire using new machine learning approach of Multivariate Adaptive Regression Splines and Differential Flower Pollination optimization: A case study at Lao Cai province (Viet Nam). *J. Environ. Manag.* **2019**, *237*, 476–487. [[CrossRef](#)]
29. Sevinc, V.; Kucuk, O.; Goltas, M. A Bayesian network model for prediction and analysis of possible forest fire causes. *For. Ecol. Manag.* **2020**, *457*, 117723. [[CrossRef](#)]
30. Mhawej, M.; Faour, G.; Adjizian-Gerard, J. A novel method to identify likely causes of wildfire. *Clim. Risk Manag.* **2017**, *16*, 120–132. [[CrossRef](#)]
31. Jaafari, A.; Zenner, E.K.; Pham, B.T. Wildfire spatial pattern analysis in the Zagros Mountains, Iran: A comparative study of decision tree based classifiers. *Ecol. Inform.* **2018**, *43*, 200–211. [[CrossRef](#)]
32. Gholamnia, K.; Gudiyangada Nachappa, T.; Ghorbanzadeh, O.; Blaschke, T. Comparisons of Diverse Machine Learning Approaches for Wildfire Susceptibility Mapping. *Symmetry* **2020**, *12*, 604. [[CrossRef](#)]
33. 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](#)]
34. Goldarag, Y.J.; Mohammadzadeh, A.; Ardakani, A. Fire risk assessment using neural network and logistic regression. *J. Indian Soc. Remote Sens.* **2016**, *44*, 885–894. [[CrossRef](#)]
35. Jaafari, A.; Razavi Termeh, S.V.; Bui, D.T. Genetic and firefly metaheuristic algorithms for an optimized neuro-fuzzy prediction modeling of wildfire probability. *J. Environ. Manag.* **2019**, *243*, 358–369. [[CrossRef](#)]
36. 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](#)]
37. Moayedi, H.; Mehrabi, M.; Bui, D.T.; Pradhan, B.; Foong, L.K. Fuzzy-metaheuristic ensembles for spatial assessment of forest fire susceptibility. *J. Environ. Manag.* **2020**, *260*, 109867. [[CrossRef](#)]
38. Jaafari, A.; Pourghasemi, H.R. Factors Influencing Regional-Scale Wildfire Probability in Iran: An Application of Random Forest and Support Vector Machine. In *Spatial Modeling in GIS and R for Earth and Environmental Sciences*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 607–619.
39. Vilar, L.; Woolford, D.G.; Martell, D.L.; Martín, M.P. A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain. *Int. J. Wildland Fire* **2010**, *19*, 325–337. [[CrossRef](#)]
40. Tehrany, M.S.; Jones, S.; Shabani, F.; Martínez-Álvarez, F.; Tien Bui, D. A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using LogitBoost machine learning classifier and multi-source geospatial data. *Theor. Appl. Climatol.* **2019**, *137*, 637–653. [[CrossRef](#)]

41. Pham, B.T.; Prakash, I.; Khosravi, K.; Chapi, K.; Trinh, P.T.; Ngo, T.Q.; Hosseini, S.V.; Bui, D.T. A comparison of Support Vector Machines and Bayesian algorithms for landslide susceptibility modelling. *Geocarto Int.* **2018**. [[CrossRef](#)]
42. Pham, B.T.; Bui, D.; Prakash, I.; Dholakia, M.B. Evaluation of predictive ability of support vector machines and naive Bayes trees methods for spatial prediction of landslides in Uttarakhand state (India) using GIS. *J. Geomat.* **2016**, *10*, 71–79.
43. Taheri, K.; Shahabi, H.; Chapi, K.; Shirzadi, A.; Gutiérrez, F.; Khosravi, K. Sinkhole susceptibility mapping: A comparison between Bayes-based machine learning algorithms. *Land Degrad. Dev.* **2019**, *30*, 730–745. [[CrossRef](#)]
44. Chen, W.; Zhang, S.; Li, R.; Shahabi, H. Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. *Sci. Total Environ.* **2018**, *644*, 1006–1018. [[CrossRef](#)] [[PubMed](#)]
45. Wu, Z.; Shen, Y.; Wang, H.; Wu, M. Assessing urban flood disaster risk using Bayesian network model and GIS applications. *Geomat. Nat. Hazards Risk* **2019**, *10*, 2163–2184. [[CrossRef](#)]
46. Vetrata, Y.; Cochrane, M.A. Fire Frequency and Related Land-Use and Land-Cover Changes in Indonesia’s Peatlands. *Remote Sens.* **2020**, *12*, 5. [[CrossRef](#)]
47. 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](#)]
48. Nami, M.H.; Jaafari, A.; Fallah, M.; Nabiuni, S. Spatial prediction of wildfire probability in the Hyrcanian ecoregion using evidential belief function model and GIS. *Int. J. Environ. Sci. Technol.* **2018**, *15*, 373–384. [[CrossRef](#)]
49. Chuvieco, E.; Caceres, D.; Riaño, D.; Martin, P.; Martínez-Vega, J.; de la Riva, J.; Pérez, F. Combining NDVI and surface temperature for the estimation of live fuel moisture content in forest fire danger rating. *Remote Sens. Environ.* **2004**, *92*, 322–331. [[CrossRef](#)]
50. Dimitrakopoulos, A.; Papaioannou, K.K. Flammability assessment of Mediterranean forest fuels. *Fire Technol.* **2001**, *37*, 143–152. [[CrossRef](#)]
51. Aragão, L.E.; Anderson, L.O.; Fonseca, M.G.; Rosan, T.M.; Vedovato, L.B.; Wagner, F.H.; Silva, C.V.; Junior, C.H.S.; Araujo, E.; Aguiar, A.P. 21st Century drought-related fires counteract the decline of Amazon deforestation carbon emissions. *Nat. Commun.* **2018**, *9*, 1–12. [[CrossRef](#)]
52. Karnieli, A.; Agam, N.; Pinker, R.T.; Anderson, M.; Imhoff, M.L.; Gutman, G.G.; Panov, N.; Goldberg, A. Use of NDVI and land surface temperature for drought assessment: Merits and limitations. *J. Clim.* **2010**, *23*, 618–633. [[CrossRef](#)]
53. Rodrigues, M.; Jiménez-Ruano, A.; Peña-Angulo, D.; de la Riva, J. A comprehensive spatial-temporal analysis of driving factors of human-caused wildfires in Spain using geographically weighted logistic regression. *J. Environ. Manag.* **2018**, *225*, 177–192. [[CrossRef](#)] [[PubMed](#)]
54. Parisien, M.-A.; Miller, C.; Parks, S.A.; DeLancey, E.R.; Robbin, F.-N.; Flannigan, M.D. The spatially varying influence of humans on fire probability in North America. *Environ. Res. Lett.* **2016**, *11*, 075005. [[CrossRef](#)]
55. Ricotta, C.; Di Vito, S. Modeling the landscape drivers of fire recurrence in Sardinia (Italy). *Environ. Manag.* **2014**, *53*, 1077–1084. [[CrossRef](#)] [[PubMed](#)]
56. Elia, M.; Giannico, V.; Spano, G.; Laforteza, R.; Sanesi, G. Likelihood and frequency of recurrent fire ignitions in highly urbanised Mediterranean landscapes. *Int. J. Wildland Fire* **2020**, *29*, 120–131. [[CrossRef](#)]
57. Kira, K.; Rendell, L.A. A practical approach to feature selection. In *Machine Learning Proceedings 1992*; Elsevier: Amsterdam, The Netherlands, 1992; pp. 249–256.
58. Kononenko, I. Estimating attributes: Analysis and extensions of RELIEF. In Proceedings of the European Conference on Machine Learning, Catania, Italy, 25–27 April 1995; pp. 171–182.
59. Dao, D.V.; Jaafari, A.; Bayat, M.; Mafi-Gholami, D.; Qi, C.; Moayed, H.; Phong, T.V.; Ly, H.-B.; Le, T.-T.; Trinh, P.T.; et al. A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *CATENA* **2020**, *188*, 104451. [[CrossRef](#)]
60. Friedman, N.; Geiger, D.; Goldszmidt, M. Bayesian network classifiers. *Mach. Learn.* **1997**, *29*, 131–163. [[CrossRef](#)]
61. Cheng, J.; Greiner, R. Comparing Bayesian network classifiers. In Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, San Francisco, CA, USA, 30 July–1 August 1999; pp. 101–108.

62. Davies, P. *Bayesian Decision Networks for Management of High Conservation Assets (National Water Initiative—Australian Government Water Fund; Report 6/6 Report to the Conservation of Freshwater Ecosystem Values Project; Water Resources Division, Department of Primary Industries and Water: Hobart, Australia, 2007.*
63. Pham, B.T.; Pradhan, B.; Tien Bui, D.; Prakash, I.; Dholakia, M.B. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environ. Model. Softw.* **2016**, *84*, 240–250. [[CrossRef](#)]
64. Pham, B.T.; Prakash, I.; Jaafari, A.; Bui, D.T. Spatial Prediction of Rainfall-Induced Landslides Using Aggregating One-Dependence Estimators Classifier. *J. Indian Soc. Remote Sens.* **2018**, *46*, 1457–1470. [[CrossRef](#)]
65. Debeljak, M.; Džeroski, S. Decision trees in ecological modelling. In *Modelling Complex Ecological Dynamics*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 197–209.
66. Wang, Y.; Witten, I.; van Someren, M.; Widmer, G. Inducing models trees for continuous classes. In Proceedings of the Poster Papers of the European Conference on Machine Learning, Department of Computer Science, University of Waikato, Hamilton, New Zealand, 1997.
67. Nguyen, P.T.; Ha, D.H.; Jaafari, A.; Nguyen, H.D.; Van Phong, T.; Al-Ansari, N.; Prakash, I.; Le, H.V.; Pham, B.T. Groundwater Potential Mapping Combining Artificial Neural Network and Real AdaBoost Ensemble Technique: The DakNong Province Case-study, Vietnam. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2473. [[CrossRef](#)]
68. Nguyen, P.T.; Ha, D.H.; Avand, M.; Jaafari, A.; Nguyen, H.D.; Al-Ansari, N.; Phong, T.V.; Sharma, R.; Kumar, R.; Le, H.V.; et al. Soft Computing Ensemble Models Based on Logistic Regression for Groundwater Potential Mapping. *Appl. Sci.* **2020**, *10*, 2469. [[CrossRef](#)]
69. Jaafari, A. LiDAR-supported prediction of slope failures using an integrated ensemble weights-of-evidence and analytical hierarchy process. *Environ. Earth Sci.* **2018**, *77*, 42. [[CrossRef](#)]
70. Janizadeh, S.; Avand, M.; Jaafari, A.; Phong, T.V.; Bayat, M.; Ahmadisharaf, E.; Prakash, I.; Pham, B.T.; Lee, S. Prediction Success of Machine Learning Methods for Flash Flood Susceptibility Mapping in the Tafresh Watershed, Iran. *Sustainability* **2019**, *11*, 5426. [[CrossRef](#)]
71. Chen, W.; Hong, H.; Panahi, M.; Shahabi, H.; Wang, Y.; Shirzadi, A.; Pirasteh, S.; Alesheikh, A.A.; Khosravi, K.; Panahi, S.; et al. Spatial Prediction of Landslide Susceptibility Using GIS-Based Data Mining Techniques of ANFIS with Whale Optimization Algorithm (WOA) and Grey Wolf Optimizer (GWO). *Appl. Sci.* **2019**, *9*, 3755. [[CrossRef](#)]
72. Avand, M.; Janizadeh, S.; Tien Bui, D.; Pham, V.H.; Ngo, P.T.T.; Nhu, V.-H. A tree-based intelligence ensemble approach for spatial prediction of potential groundwater. *Int. J. Digit. Earth* **2020**, *1*–22. [[CrossRef](#)]
73. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2013; Volume 398.
74. Caruana, R.; Niculescu-Mizil, A. Data mining in metric space: An empirical analysis of supervised learning performance criteria. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WC, USA, 22–25 August 2004; pp. 69–78.
75. Lavesson, N.; Davidsson, P. Generic methods for multi-criteria evaluation. In Proceedings of the 2008 SIAM International Conference on Data Mining, Atlanta, GA, USA, 24–26 April 2008; pp. 541–546.
76. Nhu, V.-H.; Shirzadi, A.; Shahabi, H.; Chen, W.; Clague, J.J.; Geertsema, M.; Jaafari, A.; Avand, M.; Miraki, S.; Asl, D.T. Shallow Landslide Susceptibility Mapping by Random Forest Base Classifier and its Ensembles in a Semi-Arid Region of Iran. *Forests* **2020**, *11*, 421. [[CrossRef](#)]
77. Nhu, V.-H.; Janizadeh, S.; Avand, M.; Chen, W.; Farzin, M.; Omidvar, E.; Shirzadi, A.; Shahabi, H.; Clague, J.J.; Jaafari, A.; et al. GIS-Based Gully Erosion Susceptibility Mapping: A Comparison of Computational Ensemble Data Mining Models. *Appl. Sci.* **2020**, *10*, 2039. [[CrossRef](#)]
78. Bui, D.T.; Moayedi, H.; Gör, M.; Jaafari, A.; Foong, L.K. Predicting slope stability failure through machine learning paradigms. *Isprs Int. J. Geo Inf.* **2019**, *8*, 395. [[CrossRef](#)]
79. Rahmati, O.; Panahi, M.; Ghiasi, S.S.; Deo, R.C.; Tiefenbacher, J.P.; Pradhan, B.; Jahani, A.; Goshtasb, H.; Kornejady, A.; Shahabi, H.; et al. Hybridized neural fuzzy ensembles for dust source modeling and prediction. *Atmos. Environ.* **2020**, *224*, 117320. [[CrossRef](#)]
80. Falah, F.; Ghorbani Nejad, S.; Rahmati, O.; Daneshfar, M.; Zeinivand, H. Applicability of generalized additive model in groundwater potential modelling and comparison its performance by bivariate statistical methods. *Geocarto Int.* **2017**, *32*, 1069–1089. [[CrossRef](#)]

81. Rahmati, O.; Melesse, A.M. Application of Dempster–Shafer theory, spatial analysis and remote sensing for groundwater potentiality and nitrate pollution analysis in the semi-arid region of Khuzestan, Iran. *Sci. Total Environ.* **2016**, *568*, 1110–1123. [[CrossRef](#)]
82. Rahmati, O.; Haghizadeh, A.; Stefanidis, S. Assessing the Accuracy of GIS-Based Analytical Hierarchy Process for Watershed Prioritization; Gorganrood River Basin, Iran. *Water Resour. Manag.* **2016**, *30*, 1131–1150. [[CrossRef](#)]
83. Pham, B.T.; Jaafari, A.; Prakash, I.; Singh, S.K.; Quoc, N.K.; Bui, D.T. Hybrid computational intelligence models for groundwater potential mapping. *Catena* **2019**, *182*. [[CrossRef](#)]
84. Bui, D.T.; Khosravi, K.; Tiefenbacher, J.; Nguyen, H.; Kazakis, N. Improving prediction of water quality indices using novel hybrid machine-learning algorithms. *Sci. Total Environ.* **2020**, *721*, 137612. [[CrossRef](#)] [[PubMed](#)]
85. Rahmati, O.; Panahi, M.; Kalantari, Z.; Soltani, E.; Falah, F.; Dayal, K.S.; Mohammadi, F.; Deo, R.C.; Tiefenbacher, J.; Tien Bui, D. Capability and robustness of novel hybridized models used for drought hazard modeling in southeast Queensland, Australia. *Sci. Total Environ.* **2019**. [[CrossRef](#)] [[PubMed](#)]
86. Tien Bui, D.; Le, K.-T.T.; Nguyen, V.C.; Le, H.D.; Revhaug, I. Tropical forest fire susceptibility mapping at the Cat Ba National Park Area, Hai Phong City, Vietnam, using GIS-based kernel logistic regression. *Remote Sens.* **2016**, *8*, 347. [[CrossRef](#)]
87. Massada, A.B.; Syphard, A.D.; Stewart, S.I.; Radeloff, V.C. Wildfire ignition-distribution modelling: A comparative study in the Huron–Manistee National Forest, Michigan, USA. *Int. J. Wildland Fire* **2013**, *22*, 174–183. [[CrossRef](#)]
88. Gralewicz, N.J.; Nelson, T.A.; Wulder, M.A. Factors influencing national scale wildfire susceptibility in Canada. *For. Ecol. Manag.* **2012**, *265*, 20–29. [[CrossRef](#)]
89. Kocher, S.D.; Butsic, V. Governance of land use planning to reduce fire risk to homes Mediterranean France and California. *Land* **2017**, *6*, 24. [[CrossRef](#)]
90. Baeza, M.; De Luis, M.; Raventós, J.; Escarré, A. Factors influencing fire behaviour in shrublands of different stand ages and the implications for using prescribed burning to reduce wildfire risk. *J. Environ. Manag.* **2002**, *65*, 199–208. [[CrossRef](#)] [[PubMed](#)]
91. Nunes, M.C.; Vasconcelos, M.J.; Pereira, J.M.; Dasgupta, N.; Alldredge, R.J.; Rego, F.C. Land cover type and fire in Portugal: Do fires burn land cover selectively? *Landsc. Ecol.* **2005**, *20*, 661–673. [[CrossRef](#)]
92. Bayat, M.; Ghorbanpour, M.; Zare, R.; Jaafari, A.; Thai Pham, B. Application of artificial neural networks for predicting tree survival and mortality in the Hyrcanian forest of Iran. *Comput. Electron. Agric.* **2019**, *164*. [[CrossRef](#)]
93. McCune, B.; Grace, J.B.; Urban, D.L. *Analysis of Ecological Communities*; MjM Software Design: Gleneden Beach, OR, USA, 2002; Volume 28.
94. 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.
95. Jaafari, A.; Panahi, M.; Pham, B.T.; Shahabi, H.; Bui, D.T.; Rezaie, F.; Lee, S. Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility. *Catena* **2019**, *175*, 430–445. [[CrossRef](#)]
96. Zhao, X.; Gong, M.; Zuo, X.; Pan, L. Guest Editorial: Advances in Bio-inspired Heuristics for Computing. *Caai Trans. Intell. Technol.* **2019**, *4*, 127–128. [[CrossRef](#)]
97. Sakai, H.; Nakata, M.; Wu, W.-Z.; Miao, D.; Wang, G. Guest Editorial: Rough Sets and Data Mining. *Caai Trans. Intell. Technol.* **2019**, *4*, 201–202. [[CrossRef](#)]
98. Wei, X.; Larsen, C.P.S. Methods to Detect Edge Effected Reductions in Fire Frequency in Simulated Forest Landscapes. *ISPRS Int. J. Geo Inf.* **2019**, *8*, 277. [[CrossRef](#)]
99. Parkins, K.; York, A.; Di Stefano, J. Edge effects in fire-prone landscapes: Ecological importance and implications for fauna. *Ecol. Evol.* **2018**, *8*, 5937–5948. [[CrossRef](#)]

