Journal of Mining and Earth Sciences

Website: https://jmes.humg.edu.vn

Optimizing machine learning models for enhanced forest fire susceptibility mapping in Gia Lai province

Hung Van Le [1,*], Duc Anh Hoang [2], Giang Truong Tran [2]

[1] *Thuyloi University, Hanoi, Vietnam*

[2] *Hanoi University of Mining and Geology, Hanoi Vietnam*

ARTICLE INFO

Article history:
Received 24th Oct. 2024
Revised 14th Jan. 2025
Accepted 29th Jan. 2025

Keywords:
Forest fire,
Gia Lai,
Machine learning,
Modeling,
Optimization.

ABSTRACT

Forest fires pose significant risks to ecosystems, biodiversity, human health, and the economy, with escalating global impacts. In Vietnam, particularly during the dry season, the rising threat of forest fires necessitates accurate predictive models for effective prevention and management. This study advances forest fire susceptibility mapping in Gia Lai province by leveraging optimized machine learning models. We evaluated five models - Deep Neural Networks (DNN), Random Forest (RF), Gradient Boosting (GB), Logistic Regression (LR), and Support Vector Machines (SVM) - using a dataset of 2,827 fire incidents (2007÷2021), an equal number of non-fire points, and 12 influencing factors: slope, aspect, elevation, curvature, land use, NDVI (Normalized Difference Vegetation Index), NDWI (Normalized Difference Water Index), NDMI (Normalized Difference Moisture Index), temperature, wind speed, relative humidity, and rainfall. Among the models, RF outperformed others and was further optimized using Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Bayesian Optimization (BO). The Acc-GA-Opt-RF model (Accuracy-Optimized Random Forest using GA) achieved the best performance, with 84.4% accuracy, an AUC (Area Under the ROC Curve) of 0.9083, PPV (Positive Predictive Value) of 88.2%, NPV (Negative Predictive Value) of 81.2%, sensitivity of 79.3%, specificity of 89.4%, F-score of 0.8354, and Kappa of 0.687, demonstrating significant improvements over the unoptimized RF model. Factor importance analysis, employing Average Impurity Decrease (AID) and Permutation Feature Importance (PFI), identified NDVI and NDWI as key predictors, highlighting the critical role of vegetation indices in forest fire susceptibility. The optimized RF model was utilized to generate a forest fire susceptibility map categorizing the region into six risk levels, providing actionable insights for targeted fire prevention and management in Gia Lai province.

*Corresponding author
E - mail: hungvle@tlu.edu.vn
DOI: 10.46326/JMES.2025.66(2).02

1. Introduction

Forest fires are highly destructive natural disasters that cause ecosystem damage, biodiversity loss, forest degradation, and greenhouse gas emissions, posing significant threats to human health and the economy (Anandaram et al., 2023). In Vietnam, particularly during dry seasons, these fires are often triggered by extreme weather and human activities such as slash-and-burn agriculture, leading to the loss of over 7,500 ha of forest in the past five years (VietNamNet Global, 2022). The rising frequency and severity of forest fires, intensified by climate change, highlight the urgent need for accurate predictive models to reduce environmental and economic impacts and protect human lives (Flannigan et al., 2009).

Machine learning (ML) models are crucial for predicting forest fire susceptibility, utilizing extensive datasets on weather, topography, vegetation, and historical fire data (Abid, 2021; Bui et al., 2018; Le et al., 2020). High-performance models like DNN, RF, and GB have proven effective in forest fire prediction (Le et al., 2021; Sathishkumar et al., 2023). Ensemble models such as RF and GB excel due to their ability to manage complex data and enhance prediction accuracy (Jain et al., 2020; Sarkar et al., 2024).

Optimizing hyperparameters is essential for enhancing ML model performance, especially in forest fire prediction (Al-Shabeeb et al., 2023; Bui et al., 2017; Islam et al., 2023). This study focuses on optimizing ML models to improve forest fire susceptibility mapping in Gia Lai province, Vietnam. We evaluated DNN, RF, GB, and benchmark models like LR and SVM. The RF model performed best and was further optimized using GA, PSO, and BO. The Acc-GA-Opt-RF model achieved superior performance with 84.4% accuracy, an AUC of 0.9083, and marked improvements in PPV, NPV, sensitivity, and specificity over the unoptimized RF model.

Feature importance was assessed using AID and PFI, with NDVI and NDWI identified as the most influential predictors of forest fire susceptibility. NDVI was the top factor, with importance values of 0.221 (AID) and 0.256 (PFI), highlighting the critical role of vegetation indices in fire risk prediction.

The optimized RF model was used to generate a forest fire susceptibility map for Gia Lai, categorizing the region into six risk levels, providing essential insights for targeted fire prevention and management. The study demonstrates the effectiveness of optimized ML models in enhancing predictive accuracy and supporting fire risk mitigation in high-risk areas.

The paper is structured as follows: Section 2 reviews the algorithms and optimization methods. Section 3 describes the study area and GIS database. Section 4 outlines the modeling methodology. Section 5 presents results and discusses model performance and factor significance. Section 6 concludes with key findings.

2. Background of the Algorithms Used

2.1. Benchmark Models

Benchmark models play a crucial role in developing and refining machine learning models by providing a baseline for comparison, helping to determine if new models outperform existing methods. In this study, LR, SVM, and DNN are used as benchmark models. LR offers a straightforward baseline for binary classifications, including forest fire susceptibility (Chang et al., 2013). SVM is effective for high-dimensional data, utilizing kernel functions to adapt to various data structures (Singh et al., 2021). DNN excels in capturing complex patterns through multiple hidden layers, addressing challenges beyond simpler models (Le et al., 2021).

2.2. Ensemble Learning

Ensemble learning models combine simpler models into a composite, providing higher accuracy and reducing variance and bias, thereby minimizing overfitting (Russell & Norvig, 2021). Their enhanced performance makes them preferred for assessing forest fire susceptibility (Hoang et al., 2023; Singh & Jeganathan, 2024).

2.2.1. Random Forest

RF combines multiple decision trees to create a more accurate and robust model, reducing overfitting and enhancing prediction accuracy, making it effective for large, complex datasets (Breiman, 2001). Its versatility and reliability make RF a preferred choice for predicting forest fires (Gao et al., 2023; Singh & Jeganathan, 2024). RF's capability to assess feature importance also aids in improving prediction accuracy.

Optimizing key RF hyperparameters is crucial for maximizing classification performance, particularly in forest fire susceptibility prediction (Bar et al., 2023). Key hyperparameters include (Breiman, 2001) : (1) n_estimators, which determines the number of trees and affects accuracy and overfitting risk; (2) max_depth, which controls tree complexity; (3) max_features, impacting model generalization; (4) min_samples_split, managing the minimum samples required to split nodes to prevent overfitting; and (5) min_samples_leaf, reducing overfitting by setting the minimum samples at leaf nodes. The careful tuning of these parameters enhances accuracy and model generalization on new data.

2.2.2. Gradient Boosting

Boosting combines multiple weak learners sequentially, creating a strong model where each step corrects the previous errors. Gradient Boosting (GB) refines models by minimizing errors through gradient optimization. Key GB algorithms include AdaBoost, XGBoost, and CatBoost: AdaBoost enhances weak models' accuracy, XGBoost offers high performance and scalability, and CatBoost excels with categorical data (Russell & Norvig, 2021). GB algorithms are highly effective in classification tasks, including forest fire prediction (Koh, 2023).

2.3. Optimization Algorithms

Common optimization algorithms include BO, GA, and PSO. BO improves search efficiency by using past trial data to predict future outcomes (Islam et al., 2023). GA, inspired by evolutionary processes like selection, crossover, and mutation, identifies optimal hyperparameters (Al-Shabeeb et al., 2023). PSO, modeled on animal behavior, uses particles representing solutions that explore the search space through shared and individual experiences (Bui et al., 2017).

3. The Study Area and GIS Database

3.1. The Study Area

Gia Lai province (Figure 1) is situated in south-central Vietnam, covering 15,510 km^2. Its topography varies from 1,748 m at Kon Ka Kinh mountain in K'Bang district to 80 m in Krongpa district. In 2022, the province had a population of 1.591 million, with a density of 103 people/km^2. The economy relies heavily on agriculture, forestry, and fishing, contributing 22.2% to the GDP, with industry and construction at 18.96%, and retail and services at 58.84% (General Statistics Office, 2023).

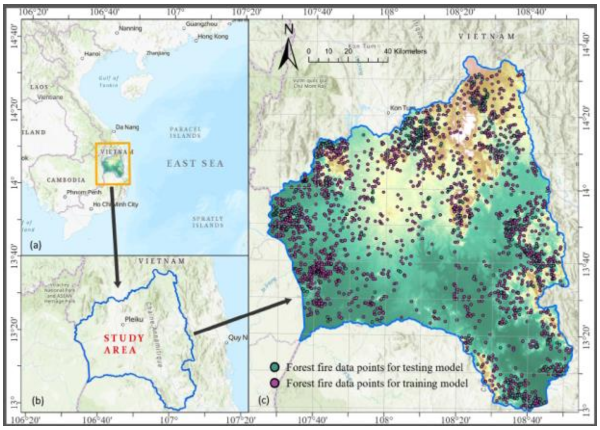

Figure 1. (a) and (b) Location of Gia Lai province, (c) Gia Lai province and forest fire locations map.

Agricultural and forested lands make up 90.14% of Gia Lai's area, with residential areas comprising 1.11%. The province has 648,300 ha of forest, including 478,800 ha of natural forest and 169,500 ha of planted forest (General Statistics Office, 2023). Frequent forest fires over the last decade have put 216,153 ha at high susceptibility, especially in planted, deciduous, and mixed bamboo forests (Nguyen, 2021).

Gia Lai experiences a tropical monsoon highland climate with high humidity and significant rainfall (Van et al., 2014). The rainy season spans from May to October, while the dry season runs from November to April. The average annual temperature ranges from 22 to 25°C, with annual rainfall between 2100 and 2200 mm (Le et al., 2021).

3.2. Forest Fire Inventory

This study utilizes a database of 2,827 forest fire incidents recorded from 2007 to 2021 (Figure 1c), originally compiled by Le et al. (2020) and subsequently updated with recent data. Fire locations were sourced from the Forest Protection Department's database (http://www.kiemlam.org.vn). The 2020-2021 dry season saw increased fire activity, with ten major fires affecting over 177 ha (Nguyen, 2021). Statistical analysis indicates that around 80% of fires occurred during the dry season, mainly between January and April. Severe fires were particularly noted in 2010, 2013, 2015, and 2016, largely driven by El Niño–Southern Oscillation events, causing droughts and a 12% reduction in rainfall (Sutton et al., 2019). In contrast, La Niña years, like 2011, experienced minimal fire activity (Le et al., 2021).

3.3. Influencing Factors

Forest fires result from ignition sources and various factors, including topography, vegetation, climate, and human activities (Cary et al., 2009). Identifying these influencing factors is essential for modeling forest fire susceptibility. This section outlines the factors considered in this study, with detailed descriptions available in (Le et al., 2021).

3.3.1. Topographical factors

Topography significantly influences forest fires through indirect and direct effects. Terrain variations create microclimates that affect temperature, vegetation cover, and tree species distribution, indirectly impacting fire occurrence and spread (Mermoz et al., 2005). Key factors like slope, aspect, elevation, and curvature directly influence fire spread and flammability. Slopes accelerate fire spread compared to flat areas (Dupuy & Maréchal, 2011), aspect affects solar radiation and vegetation moisture (Bennie et al., 2008), higher elevations with cooler temperatures and more precipitation reduce fire risk (Chen et al., 2018), and curvature alters soil conditions, affecting ignition probability (Hilton et al., 2016).. This study utilized a 30 m-resolution DEM of Gia Lai province to extract and analyze these factors (Figure 2) to evaluate their impact on forest fire behavior.

3.3.2. Human-Induced and Vegetation Factors

Human activities are a primary driver of forest fires globally, as population growth increases pressure on ecosystems, leading to deforestation and intensified land use, which elevate fire risks, especially in certain tree species (Viedma et al., 2017). Therefore, land use is a critical factor in forest fire prediction. In this study, we developed a land use map (Figure 3) with eleven categories based on district-level land use plans from Gia Lai province, provided by the People's Committee at a 1:50,000 scale.

For vegetation factors, we used the Normalized Difference Vegetation Index (NDVI) to assess vegetation health and fire fuel potential (Carlson & Ripley, 1997). Additionally, the Normalized Difference Water Index (NDWI) and Normalized Difference Moisture Index (NDMI)

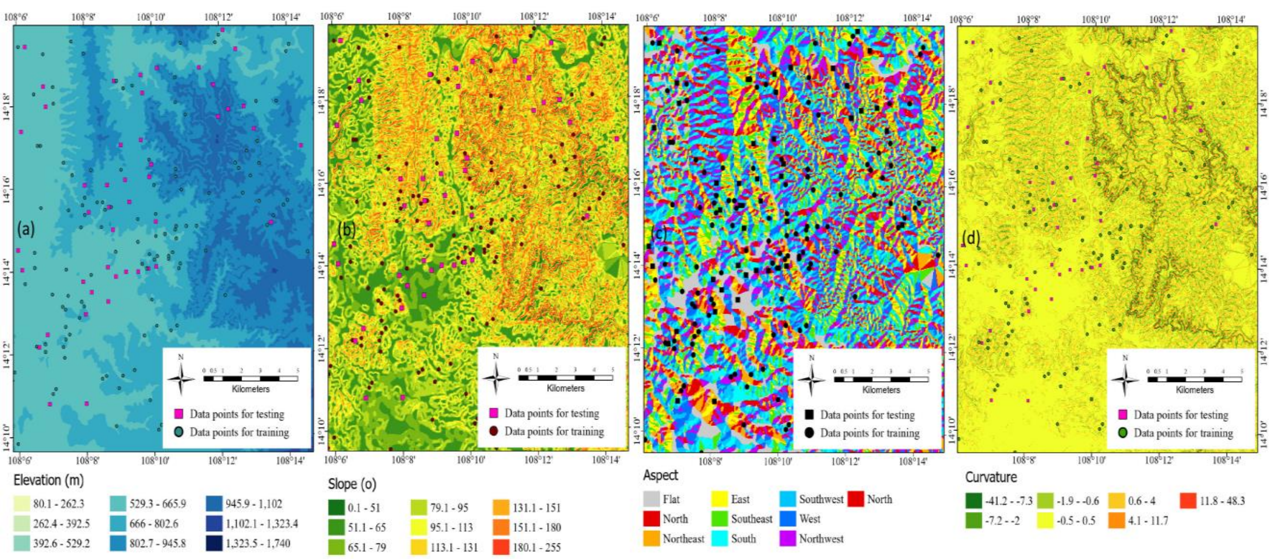

Figure 2. (a) Elevation map, (b) Slope map, (c) Aspect map, and (d) Curvature map.

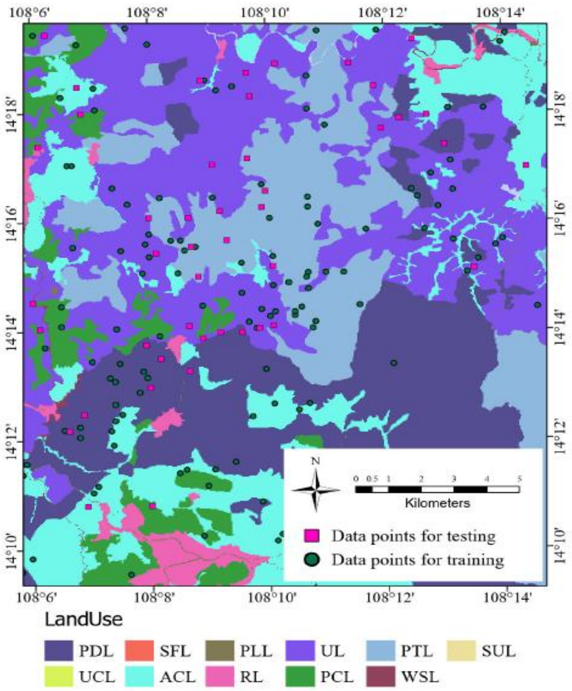

Figure 3. Land use map.

were used to evaluate vegetation water content and fuel moisture. These indices are crucial in predicting fire behavior due to their influence on fuel conditions. NDVI, NDWI, and NDMI were derived from 2021 Landsat-8 OLI satellite images with a 30 m resolution from the USGS EarthExplorer portal, following methods by (Tucker, 1979), (McFeeters, 1996), and (Wilson & Sader, 2002):

$$NDVI = \frac{NIR\ band - Red\ band}{NIR\ band + Red\ band}$$

$$NDWI = \frac{Green\ band - NIR\ band}{Green\ band + NIR\ band}$$

$$NDMI = \frac{NIR\ band - SWIR\ band}{NIR\ band + SWIR\ band}$$

Where NIR and SWIR represent the Near-Infrared and Short-Wave Infrared spectral bands, respectively. The maps of NDVI, NDWI, and NDMI are presented in Figure 4.

3.3.3. Meteorological Factors

Research has shown a strong link between climate change and forest fire patterns (Lacroix et al., 2020), highlighting the need to include climate-related factors in our analysis.

We selected four key climatic variables: temperature, wind speed, relative humidity, and rainfall (Figure 5), with data from 2007÷2021 sourced from https://www.ncdc.noaa.gov/. Temperature impacts soil moisture and directly influences plant combustion (Pourtaghi et al., 2016), and rising temperatures reduce vegetation moisture, elevating fire risk (Gillett et al., 2004). Wind speed affects fire spread by altering fuel moisture and supplying oxygen (Alexandridis et

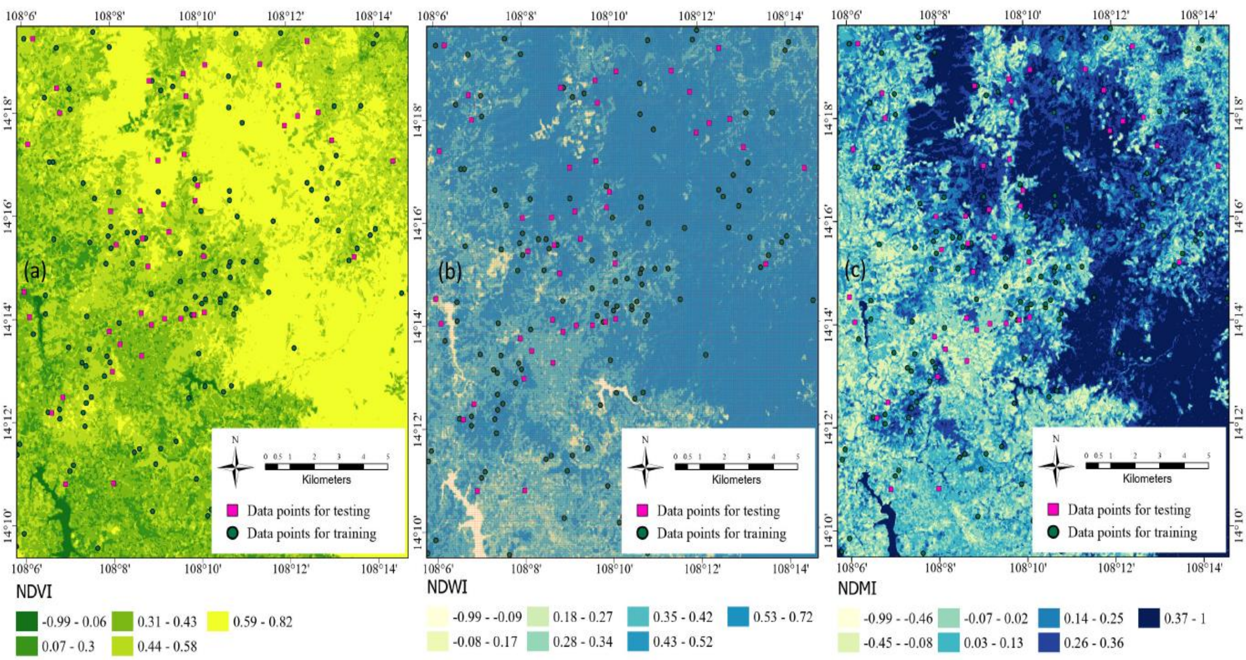

Figure 4. (a) NDVI map, (b) NDWI map, and (c) NDMI map.

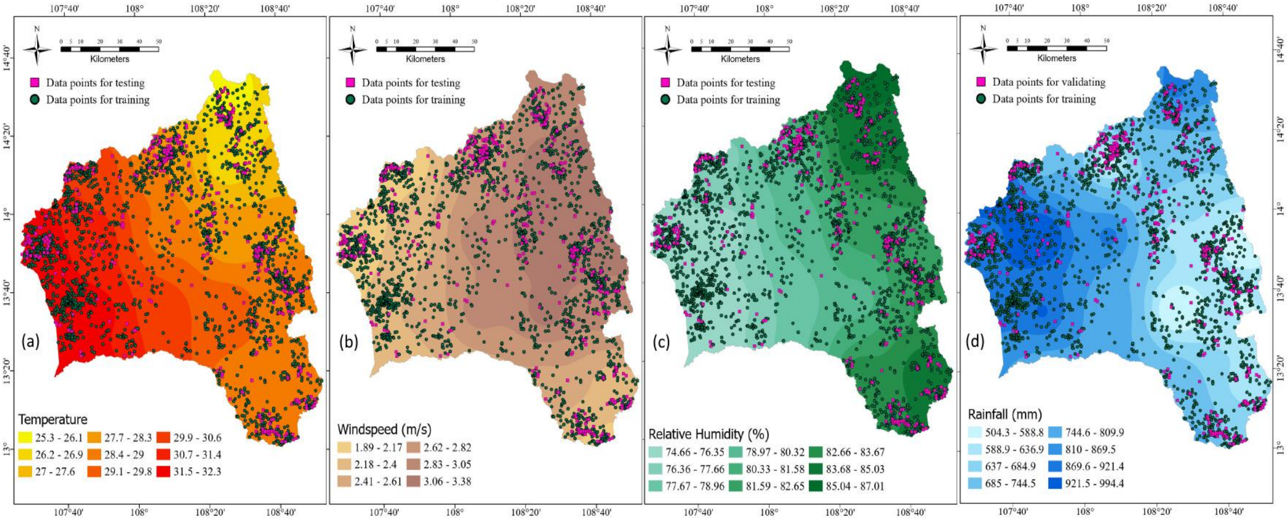

Figure 5. (a) Temperature, (b) Winspeed, (c) Humidity, and (d) Rainfall.

al., 2008). Relative humidity and precipitation significantly impact fuel moisture, which is crucial for fire ignition (Liu et al., 2013)

3.4. GIS Database

Developing machine learning models for forest fire susceptibility required a GIS database with a detailed fire inventory and 12 influencing factors, including topography, vegetation indices, and climatic variables. The database was built in ArcGIS 10.6 and stored in ESRI file geodatabase format. Due to the binary classification nature of forest fire prediction (Bui et al., 2017), the dataset includes both fire and non-fire occurrences, with 2,827 non-fire data points added, totaling 5,654 samples. These non-fire points were selected using the “Create Random Points” tool in ArcGIS, ensuring they were evenly distributed across the study area, distanced from fire points, and placed outside fire-prone regions, such as forested or flammable areas. The dataset was normalized by converting categorical data to a numerical scale from 0.01 to 0.99 (Le et al., 2020). It was then split into a training set of 4,240 samples (75%) and a test set of 1,414 samples (25%). The process of creating the GIS database, and the training and test sets, is illustrated in Figure 6.

4. Modeling Methodology

4.1. Modeling Process

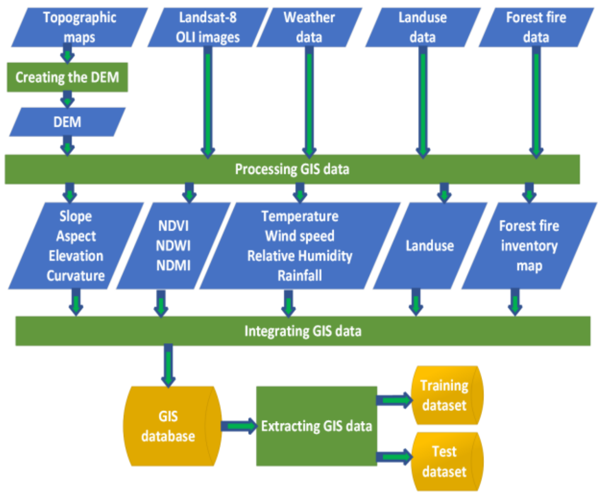

Figure 6. The creation of the GIS database.

We first built benchmark models using the training dataset and evaluated them on the test dataset to identify the most promising one for further optimization. Optimization was then performed using BO, PSO, and GA techniques. The model with the highest post-optimization performance was selected for forest fire susceptibility mapping in the study area. This process is shown in Figure 7.

4.2. Model Performance Assessment

We assessed the models using binary classification metrics, categorizing outcomes as true positives (TP), false negatives (FN), true negatives (TN), and false positives (FP). Key metrics included PPV, NPV, sensitivity (Sen), specificity (Spe), false positive rate (FPR),

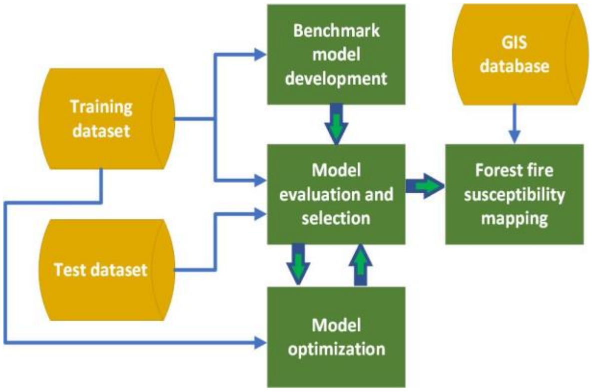

Figure 7. Modeling process.

accuracy (Acc), F1 Score (F-score), and Cohen's Kappa (Kappa), providing insights into the models' accuracy in classifying fire and non-fire events. The ROC curve and AUC were used to evaluate classification performance. Analyses were conducted on both training and test datasets to assess model fitting and predictive power (Le et al., 2020; Le et al., 2021). The following metrics were calculated (Powers, 2011):

$$PPV = \frac{TP}{TP + FP}; NPV = \frac{TN}{TN + FN}$$

$$Sen = \frac{TP}{TP + FN}; Spe = \frac{TN}{TN + FP}$$

$$FPR = \frac{FP}{FP + TN}; Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F - score = \frac{2 \times TP}{2 \times TP + FP + FN}$$

$$Kap = \frac{2 \times (TP \times TN - FN \times FP)}{(TP + FP) \times (FP + TN) + (TP + FN) \times (FN + TN)}$$

4.3. Construction of Benchmark Models

The benchmark models (LR, SVM, CatBoost, AdaBoost, XGBoost, RF, and DNN) were trained with default hyperparameters, except for the DNN, which was optimized using PSO. The DNN architecture, fine-tuned with 2÷5 hidden layers and 5÷50 neurons per layer, used ReLU activation and the Adam optimizer, while the output neuron used sigmoid activation. Five-fold cross-validation minimized overfitting, and the optimal architecture consisted of 5 hidden layers with 36, 50, 22, 46, and 5 neurons.

4.4. Benchmark Model Evaluation and Selection

The benchmark models were evaluated on the test dataset to establish a performance baseline, serving as a reference for subsequent optimizations. The RF model outperformed others in key metrics, including Acc, AUC, F-score, and Kappa (Table 1), and was thus chosen for further optimization.

Table 1. Key performance metrics of the benchmark models on the test dataset.

Model	Acc	AUC	F-score	Kappa
LR	61.0%	0.6328	0.6063	0.219
SVM	62.8%	0.7041	0.5731	0.256
CatBoost	77.0%	0.8574	0.7459	0.540
AdaBoost	81.0%	0.8497	0.8066	0.620
XGBoost	81.8%	0.8772	0.8103	0.636
RF	82.4%	0.8819	0.8178	0.648
DNN	78.9%	0.8619	0.7916	0.579

4.5. Model Optimization

The RF model was optimized using BO, PSO, and GA to refine hyperparameters, including n_estimators, max_depth, max_features, min_samples_split, and min_samples_leaf, aiming to enhance accuracy/AUC for effective fire prediction. Optimization was validated via five-fold cross-validation within the following constraints: n_estimators (10÷200), max_depth (5÷40), max_features (1÷12), min_samples_split (2÷5), and min_samples_leaf (1÷5), using an initial population of 20 and 50 iterations.

5. Results and Discussion

5.1. Performance Evaluation of Optimized RF Models

The RF model optimization aimed to improve accuracy (Acc)/AUC by fine-tuning the hyperparameters using BO, PSO, and GA. The optimized hyperparameter values are shown in

Table 2. The performance of these optimized models on both the training and test datasets is summarized in Table 3 and Table 4, along with the results of the unoptimized RF model.

Table 2. Hyperparameter values of the optimized RF models.

Model	max_depth	max_features	min_samples_leaf	min_samples_split	n_estimators
AUC-BO-Opt-RF	32	5	2	2	200
AUC-PSO-Opt-RF	30	3	1	3	123
AUC-GA-Opt-RF	35	3	1	2	96
Acc-BO-Opt-RF	36	4	2	3	55
Acc-PSO-Opt-RF	37	3	2	3	151
Acc-GA-Opt-RF	40	3	1	2	100

Table 3. Performance metrics of unoptimized and optimized RF models on the training dataset.

Model	TP	FN	FP	TN	Acc, %	AUC, %	PPV, %	NPV, %	Sen, %	Spe, %	F-score	Kappa
RF	2046	74	39	2081	97.3	0.9952	98.1	96.6	96.5	98.2	0.9731	0.947
Acc-BO-Opt-RF	2099	21	8	2112	99.3	0.9986	99.6	99.0	99.0	99.6	0.9931	0.986
Acc-PSO-Opt-RF	2104	16	13	2107	99.3	0.9987	99.4	98.4	98.3	99.4	0.9889	0.986
Acc-GA-Opt-RF	2103	17	12	2108	99.3	0.9994	99.4	99.2	99.2	99.4	0.9932	0.986
AUC-BO-Opt-RF	2078	42	9	2111	98.8	0.9994	99.6	98.0	98.0	99.6	0.9879	0.976
AUC-PSO-Opt-RF	2085	35	12	2108	98.9	0.9993	99.4	99.2	99.2	99.4	0.9932	0.978
AUC-GA-Opt-RF	2105	15	15	2105	99.3	0.9994	99.4	99.2	99.2	99.4	0.9932	0.986

Table 4. Performance metrics of unoptimized and optimized RF models on the test dataset

Model	TP	FN	FP	TN	Acc	AUC	PPV	NPV	Sen	Spe	F-score	Kappa
RF	559	148	101	606	82.4%	0.8819	84.7%	80.4%	79.1%	85.7%	0.8178	0.648
Acc-BO-Opt-RF	560	147	79	628	84.0%	0.9006	87.6%	81.0%	79.2%	88.8%	0.8321	0.680
Acc-PSO-Opt-RF	560	147	76	631	84.2%	0.9037	88.1%	81.1%	79.2%	89.3%	0.8340	0.685
Acc-GA-Opt-RF	561	146	75	632	84.4%	0.9083	88.2%	81.2%	79.3%	89.4%	0.8354	0.687
AUC-BO-Opt-RF	552	155	77	630	83.6%	0.9074	87.8%	80.3%	78.1%	89.1%	0.8263	0.672
AUC-PSO-Opt-RF	555	152	80	627	83.6%	0.9099	87.4%	80.5%	78.5%	88.7%	0.8271	0.672
AUC-GA-Opt-RF	559	148	77	630	84.1%	0.9085	87.9%	81.0%	79.1%	89.1%	0.8325	0.682

The training dataset evaluation revealed that all optimized models significantly outperformed the unoptimized RF model, with accuracies exceeding 98%, highlighting the effectiveness of BO, PSO, and GA optimization techniques. Specifically, Acc-GA-Opt-RF (GA optimized for accuracy) and AUC-GA-Opt-RF (GA optimized for AUC) achieved near-perfect AUC values and the highest accuracy of 99.3% (Table 3).

On the test dataset, the Acc-GA-Opt-RF model improved accuracy by 2% compared to the unoptimized RF, reaching 84.4%, while the AUC-GA-Opt-RF model increased the AUC by about 3% (from 0.8819÷0.9083) (Table 4). These results confirm that optimization significantly enhanced RF model performance. Although both models met their optimization goals, the Acc-GA-Opt-RF model slightly outperformed the other in accuracy and consistently maintained strong performance across all metrics, making it the preferred choice for forest fire susceptibility mapping.

The Acc-GA-Opt-RF model demonstrated robust performance on both the training dataset (Acc = 99.3%, AUC = 0.9994, F-score = 0.9932, Kappa = 0.986) and the test dataset (Acc = 84.4%, AUC = 0.9083, F-score = 0.8354, Kappa = 0.687), with the ROCs and AUC metrics shown in Figure 8. These results confirm its strong generalization capability and reliability, establishing it as the optimal model for mapping forest fire susceptibility in the study area.

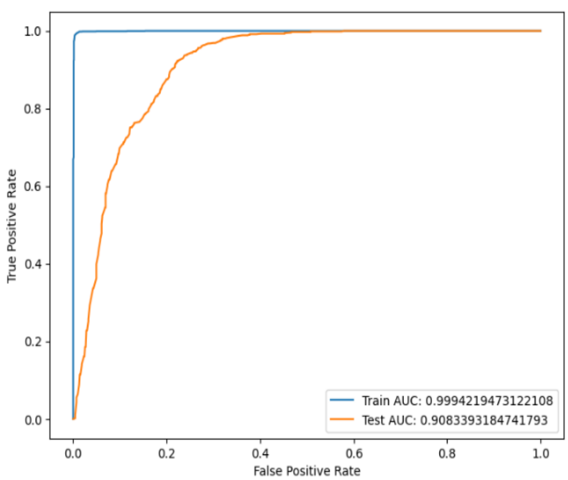

Figure 8. ROCs and AUC Metrics for the Acc-GA-Opt-RF model.

Although the Kappa value of 0.69 on the test dataset is considered "fairly good" (Landis & Koch, 1977), the discrepancy between the training and test Kappa values suggests a lack of model generalization, likely caused by its complexity and noise in the training data.

5.2. Importance of Influencing Factors

To determine key factors affecting forest fire susceptibility, we evaluated their predictive importance using two methods: (i) Average Impurity Decrease (AID), which uses the Gini impurity metric in RF to measure a feature's contribution by its ability to reduce classification uncertainty (Breiman, 2001), and (ii) Permutation Feature Importance (PFI), which quantifies feature importance by shuffling feature values and assessing the impact on model performance, with significant performance drops indicating higher importance (Molnar, 2022).

5.2.1. Factor Importance Using AID

The evaluation of factor importance in forest fires using the AID method in the RF model is summarized in Table 5 and Figure 9, ranked by decreasing importance. All factors have importance values above 0, confirming their relevance in predicting forest fires. NDVI is the most influential factor (0.221159), followed by NDWI (0.106616) and Aspect (0.097736). Other significant factors include NDMI, Slope, Curvature, and Elevation, with values of 0.091682, 0.078159, 0.074920, and 0.069838, respectively. Landuse, Rainfall, Windspeed, Humidity, and Temperature also contribute to the model's predictive accuracy.

Table 5. Importance of factors using AID.

No.	Factor	Importance
1	NDVI	0.221159
2	NDWI	0.106616
3	Aspect	0.097736
4	NDMI	0.091682
5	Slope	0.078159
6	Curvature	0.074920
7	Elevation	0.069838
8	Landuse	0.059607
9	Rainfall	0.055011
10	Windspeed	0.050356
11	Humidity	0.048165
12	Temperature	0.046750

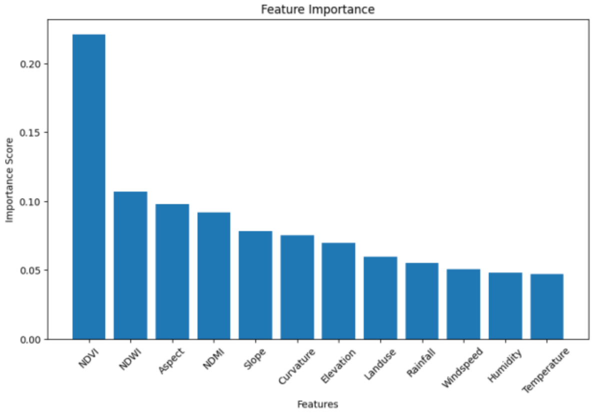

Figure 9. Importance of factors using AID.

5.2.2. Factor Importance Using PFI

PFI results from the Acc-GA-Opt-RF model ranked NDVI as the most critical factor (0.256295), followed by NDWI (0.095150) and Aspect (0.091439) (Table 6 and Figure 10). NDMI, Slope, and Elevation were also significant, emphasizing the roles of moisture and topography in fire behavior. Climatic factors like Rainfall, Humidity, Windspeed, and Temperature showed moderate importance, reflecting the influence of weather on fire susceptibility.

Table 6. Importance of factors using PFI.

No.	Factor	Importance
1	NDVI	0.256295
2	NDWI	0.095150
3	Aspect	0.091439
4	NDMI	0.084813
5	Slope	0.073946
6	Elevation	0.071826
7	Rainfall	0.064140
8	Humidity	0.058044
9	Curvature	0.054333
10	Landuse	0.054333
11	Windspeed	0.052213
12	Temperature	0.043467

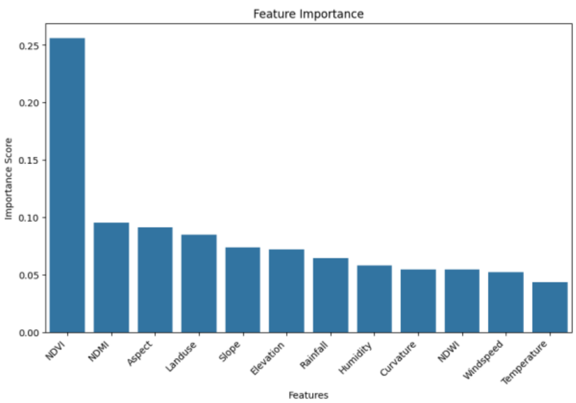

Figure 10. Importance of factors using PFI .

5.2.3. Discussion of Factor Importance

Both AID and PFI methods identified NDVI as the most important factor (0.221159 and 0.256295, respectively), aligning with Bui et al.. (2018). NDWI, Aspect, NDMI, Slope, and Elevation also showed significant importance, highlighting the roles of vegetation, topography, and elevation. Although Rainfall, Humidity, Curvature, Land Use, Windspeed, and Temperature ranked lower, they still contributed to improving predictive accuracy, emphasizing the combined impact of vegetation, topography, and climatic factors on forest fire susceptibility mapping.

The low importance of temperature may be due to Gia Lai's stable dry-season temperature (18÷25^{0}C), resulting in minimal influence on forest fire risk (Bui et al., 2018). Similarly, studies in other Central Highlands provinces, such as Bui et al. (2018) and Le et al. (2020), ranked temperature as moderately important, placing it sixth among factors. In contrast, Luu et al. (2014) in Dak Lak considered temperature significant, likely due to greater variability and its effect on vegetation drying. This study used objective methods (AID and PFI), whereas the Dak Lak study relied on expert judgment.

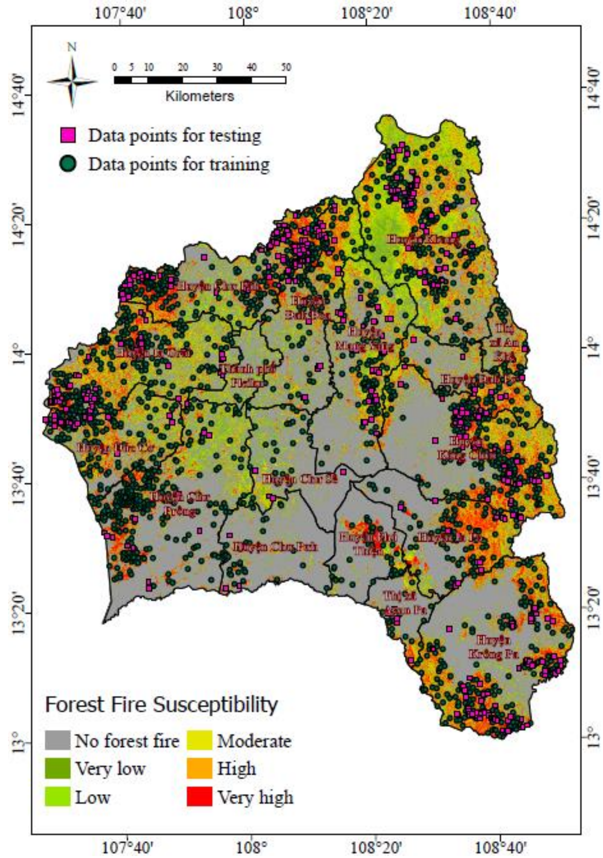

Figure 11. Forest fire susceptibility map.

5.3. Creation of a forest fire susceptibility map

Owing to its superior performance, the Acc-GA-Opt-RF model was utilized to compute forest fire susceptibility probabilities across Gia Lai province, which were subsequently converted into raster format for integration with ArcGIS.

The forest fire susceptibility map in Figure 11 classifies the region into six levels: No Forest Fire (0÷0.080), Very Low (0.081÷0.292), Low (0.293÷0.488), Moderate (0.489÷0.695), High (0.696÷0.905), and Very High (0.906÷1), using the Natural Breaks method in ArcGIS 10.6. This method was chosen for its effectiveness in identifying natural data groupings, minimizing within-class variance while maximizing between-class variance. In this study, the region was categorized into six levels, aligning with the data's natural structure.

The map highlights high-risk zones in Chu Pah, Mang Yang, Kong Chro, Ia Grai, and Dak Doa, which faced over 177 ha of forest loss during the 2020-2021 dry season, including two major fires in Chu Pah damaging 14.29 ha of planted forests (Nguyen, 2021). In contrast, districts like Ayun Pa and An Khe showed lower risk due to reforestation reducing fuel accumulation. These findings emphasize the need for targeted fire management strategies, including early warning systems, controlled burns, firebreaks, and community engagement, to mitigate risks and protect ecosystems and local communities.

6. Concluding remarks

This study demonstrates the pivotal role of optimized machine learning models in predicting forest fire susceptibility, particularly in high-risk regions like Gia Lai province. We aimed to identify and optimize the most effective model among commonly used high-performance models for forest fire prediction. Among the evaluated models - DNN, RF, GB, LR, and SVM - the RF model showed superior accuracy and predictive capabilities.

Hyperparameter optimization using GA, PSO, and BO significantly improved the RF model's performance, boosting accuracy by 2% and AUC by about 3% compared to the non-optimized version. The Acc-GA-Opt-RF model (Accuracy-Optimized RF using GA) achieved the best results, with an accuracy of 84.4%, an AUC of 0.9083, an F-score of 0.8354, and a Kappa of 0.687. These enhanced metrics enable more reliable identification of high-risk fire zones, enhancing predictive capabilities in real-world applications.

Feature importance analysis using AID and PFI identified NDVI and NDWI as the most influential factors, underscoring the critical role of vegetation indices in predicting fire risk. The optimized model was used to generate a forest fire susceptibility map for Gia Lai, categorizing areas into six risk levels. This map offers crucial guidance for policymakers and local authorities, aiding in targeted fire prevention and management strategies.

Overall, this study highlights that optimized machine learning models, supported by insights into key influencing factors, can significantly enhance predictive accuracy and serve as valuable tools for mitigating fire risks in vulnerable regions.

Contributions of authors

Hung Van Le - methodology, writing, review, editing, and supervision; Duc Hoang Anh - data collection and processing, model building, writing, and editing; Giang Truong Tran - data collection and processing, map creation, writing, and editing.

References

Abid, F. (2021). A Survey of Machine Learning Algorithms Based Forest Fires Prediction and Detection Systems. *Fire Technology*, 57(2), 559-590.

Alexandridis, A., Vakalis, D., Siettos, C. I., & Bafas, G. V. (2008). A cellular automata model for forest fire spread prediction: The case of the wildfire that swept through Spetses Island in 1990. *Applied Mathematics and Computation*, 204(1), 191-201.

Al-Shabeeb, A. R., Hamdan, I., Meimandi Parizi, S., Al-Fugara, A. k., Odat, S. a., Elkhrachy, I., Hu, T., & Sammen, S. S. (2023). A Comparative Study of Genetic Algorithm-Based Ensemble Models and Knowledge-Based Models for Wildfire Susceptibility Mapping. *Sustainability*, 15(21).

Anandaram, H., M, N., Cosio Borda, R. F., K, K., & S, Y. (2023). Forest fire management using machine learning techniques. *Measurement: Sensors*, 25, 100659.

Bar, S., Parida, B. R., Pandey, A. C., Shankar, B. U., Kumar, P., Panda, S. K., & Behera, M. D. (2023). Modeling and prediction of fire occurrences along an elevational gradient in Western Himalayas. *Applied Geography*, 151, 102867.

Bennie, J., Huntley, B., Wiltshire, A., Hill, M. O., & Baxter, R. (2008). Slope, aspect and climate: Spatially explicit and implicit models of topographic microclimate in chalk grassland. *Ecological Modelling*, 216(1), 47-59.

Breiman, L. (2001). Random Forests. *Machine Learning,* 45(1), 5-32.

Bui, D. T., Bui, Q.-T., Nguyen, Q.-P., Pradhan, B., Nampak, H., & Trinh, P. T. (2017). 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. *Agricultural and Forest Meteorology*, 233, 32-44.

Bui, T. D., Le, H. V., & Hoang, N.-D. (2018). GIS-based spatial prediction of tropical forest fire danger using a new hybrid machine learning method. *Ecological Informatics*, 48, 104-116.

Carlson, T. N., & Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment*, 62(3), 241-252.

Chang, Y., Zhu, Z., Bu, R., Chen, H., Feng, Y., Li, Y., Hu, Y., & Wang, Z. (2013). Predicting fire occurrence patterns with logistic regression in Heilongjiang Province, China. *Landscape Ecology*, 28(10), 1989-2004.

Chen, B. X., Sun, Y. F., Zhang, H. B., Han, Z. H., Wang, J. S., Li, Y. K., & Yang, X. L. (2018). Temperature change along elevation and its effect on the alpine timberline tree growth in the southeast of the Tibetan Plateau. *Advances in Climate Change Research*, 9(3), 185-191.

Dupuy, J. L., & Maréchal, J. (2011). Slope effect on laboratory fire spread: contribution of radiation and convection to fuel bed preheating. *International Journal of Wildland Fire*, 20(2), 289-307.

Flannigan, M. D., Krawchuk, M. A., de Groot, W. J., Wotton, B. M., & Gowman, L. M. (2009). Implications of changing climate for global wildland fire. *International Journal of Wildland Fire*, 18(5), 483-507.

Gao, C., Lin, H., & Hu, H. (2023). Forest-Fire-Risk Prediction Based on Random Forest and Backpropagation Neural Network of Heihe Area in Heilongjiang Province, *China. Forests*, 14(2).

General Statistics Office. (2023). 2022 Statistical Yearbook of Viet Nam*. Statistical Publishing House.*

Gillett, N. P., Weaver, A. J., Zwiers, F. W., & Flannigan, M. D. (2004). Detecting the effect of climate change on Canadian forest fires. *Geophysical Research Letters*, 31(18).

Hilton, J. E., Miller, C., Sharples, J. J., & Sullivan, A. L. (2016). Curvature effects in the dynamic propagation of wildfires. *International Journal of Wildland Fire,* 25(12), 1238-1251.

Hoang, D. A., Le, H. V., Pham, D. V., Hoa, P. V., & Tien Bui, D. (2023). Hybrid BBO-DE Optimized SPAARCTree Ensemble for Landslide Susceptibility Mapping. *Remote Sensing*, 15(8).

https://baogialai.com.vn/tiem-an-nguy-co-chay-rung-post128443.html

https://vietnamnet.vn/en/vietnam-records-over-17-000-fires-in-five-years-2059848.html

Islam, A. M., Masud, F. B., Ahmed, M. R., Jafar, A. I., Ullah, J. R., Islam, S., Shatabda, S., & Islam, A. K. M. M. (2023). An Attention-Guided Deep-Learning-Based Network with Bayesian Optimization for Forest Fire Classification and Localization. *Forests*, 14(10).

Jain, P., Coogan, S. C. P., Subramanian, S. G., Crowley, M., Taylor, S., & Flannigan, M. D. (2020). A review of machine learning applications in wildfire science and management. *Environmental Reviews*, 28(4), 478-505.

Koh, J. (2023). Gradient boosting with extreme-value theory for wildfire prediction. *Extremes*, 26(2), 273-299.

Lacroix, K., Gifford, R., & Rush, J. (2020). Climate change beliefs shape the interpretation of forest fire events. *Climatic Change*, 159(1), 103-120.

Landis, J. R. & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174.

Le, H. V., Bui, Q. T., Tien Bui, D., Tran, H. H., & Hoang, N. D. (2020). A Hybrid Intelligence System Based on Relevance Vector Machines and Imperialist Competitive Optimization for Modelling Forest Fire Danger Using GIS. *Journal of Environmental Informatics*, 36(1).

Le, H. V., Hoang, D. A., Tran, C. T., Nguyen, P. Q., Tran, V. H. T., Hoang, N. D., Amiri, M., Ngo, T. P. T., Nhu, H. V., Hoang, T. V., & Tien Bui, D. (2021). A new approach of deep neural computing for spatial prediction of wildfire danger at tropical climate areas. *Ecological Informatics*, 63, 101300.

Liu, Y., Goodrick, S. L., & Stanturf, J. A. (2013). Future U.S. wildfire potential trends projected using a dynamically downscaled climate change scenario. *Forest Ecology and Management*, 294, 120-135.

Luu, T. A., Chan, A. T., Hoang, T. H. N., & Le, B. B. (2014). Application of Remote Sensing Imagery and GIS in Establishment of Forest Fire Hazard Map in Daklak Province. *Vietnam Journal of Earth Sciences*, 36(3), 5908

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425-1432.

Mermoz, M., Kitzberger, T., & Veblen, T. T. (2005). Landscape Influences on Occurrence and Spread of Wildfires in Patagonian Forests and Shrublands. *Ecology*, 86(10), 2705-2715.

Molnar, C. (2022). Interpretable Machine Learning (2nd ed.). *Independently published*

Nguyen, D. (2021, 24/12/2021). Potential Risk of Forest Fires. Gia Lai online.

Pourtaghi, Z. S., Pourghasemi, H. R., Aretano, R., & Semeraro, T. (2016). Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. *Ecological Indicators*, 64, 72-84.

Powers, D. M. W. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *International Journal of Machine Learning Technology*, 2(1), 37-63.

Russell, S., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4th ed.). *Pearson*.

Sarkar, M. S., Majhi, B. K., Pathak, B., Biswas, T., Mahapatra, S., Kumar, D., Bhatt, I. D., Kuniyal, J. C., & Nautiyal, S. (2024). Ensembling machine learning models to identify forest fire-susceptible zones in Northeast India. *Ecological Informatics*, 102598.

Sathishkumar, V. E., Cho, J., Subramanian, M., & Naren, O. S. (2023). Forest fire and smoke detection using deep learning-based learning without forgetting. *Fire Ecology*, 19(1), 9.

Singh, K. R., Neethu, K. P., Madhurekaa, K., Harita, A., & Mohan, P. (2021). Parallel SVM model for forest fire prediction. *Soft Computing Letters*, 3, 100014.

Singh, S. S., & Jeganathan, C. (2024). Using ensemble machine learning algorithm to predict forest fire occurrence probability in Madhya Pradesh and Chhattisgarh*, India. Advances in Space Research,* 73(6), 2969-2987.

Sutton, W. R., Srivastava, J. P., Koo, J., Vasileiou, I., & Pradesha, A. (2019). Striking a Balance: Managing El Niño and La Niña in Cambodia's Agriculture. *World Bank.*

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment,* 8(2), 127-150.

Van, N. K., Ly, P. T., & Hong, N. T. (2014). Bioclimatic map of Tay Nguyen at scale 1:250,000 for setting up sustainable ecological economic models. *Vietnam Journal of Earth Sciences,* 36(4), 504-514.

Viedma, O., Moreno, J. M., Güngöroglu, C., Cosgun, U., & Kavgacı, A. (2017). Recent land-use and land-cover changes and its driving factors in a fire-prone area of southwestern Turkey. *Journal of Environmental Management*, 197, 719-731.

VietNamNet Global. (2022, 14/09/2022). Vietnam records over 17,000 fires in five years. VietNamNet Global.

Wilson, E. H., & Sader, S. A. (2002). Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sensing of Environment,* 80(3), 385-396.