Kartiranje tveganja za gozdne požare s tehniko strojnega učenja na podlagi podatkov GIS in daljinskega zaznavanja: študija primera v provinci Dak Lak, Vietnam

Forest fire risk mapping using machine learning technique based on GIS and remote sensing data, a case study in Dak Lak province, Vietnam

Le Hung Trinh, Van Truong Vu, Quang Tu Le, Xuan Bien Tran

UDK: 004.85:528.8(597)
Klasifikacija prispevka po COBISS.SI: 1.01
Prispelo: 4. 3. 2026
Sprejeto: 17. 7. 2026

DOI: 10.15292/geodetski-vestnik.2026.03.421-436
SCIENTIFIC ARTICLE
Received: 4. 3. 2026
Accepted: 17. 7. 2026

SI | EN

IZVLEČEK

Vietnam je pogosto izpostavljen tveganju za gozdne požare, ki jih povzroča kombinacija ekstremnih vremenskih dogodkov in antropogenih dejavnosti. Njihovo zgodnje odkrivanje in napovedovanje požarnega tveganja ima ključno vlogo pri blaženju posledic in omogočanju učinkovitega odzivanja na takšne nesreče. Študija raziskuje uporabo naprednih modelov strojnega učenja, zlasti naključnega gozda (RF), konvolucijske nevronske mreže (CNN) in dolgega kratkoročnega spomina (LSTM), v povezavi z geoprostorskimi podatki za izdelavo karte napovedi tveganja gozdnih požarov za provinco Dak Lak v osrednjem višavju Vietnama. Vhodni nabor podatkov obsega 12 spremenljivk, pridobljenih z uporabo daljinskega zaznavanja in tehnologij GIS ter integrirani z zgodovinskimi točkami pojava požarov. Učinkovitost napovedovanja modelov je bila ocenjena z uporabo več metrik. Ugotovitve poudarjajo potencial uporabe naprednih tehnik strojnega učenja in analize prostorskih podatkov za povečanje natančnosti napovedovanja gozdnih požarov. Rezultati te študije bodo pomagali oblikovalcem politik in organom za upravljanje gozdov pri razvoju učinkovitejših strategij gospodarjenja z gozdovi in ukrepov za preprečevanje požarov, s čimer bodo prispevali k varovanju naravnih virov in blaginji lokalnih skupnosti.

KLJUČNE BESEDE

napovedovanje tveganja gozdnih požarov, strojno učenje, naključni gozd, CNN, LSTM, geoprostorski podatki, GIS, provinca Dak Lak

ABSTRACT

Vietnam is frequently exposed to the risk of forest fires, driven by a combination of extreme weather events and anthropogenic activities. Early detection and forecasting of forest fire risks play a pivotal role in mitigating the impact and facilitating effective responses to such disasters. This study investigates the application of advanced machine learning models, specifically Random Forest (RF), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM), in conjunction with geospatial data, to construct a forest fire risk prediction map for Dak Lak province in the Central Highlands of Vietnam. The input dataset comprises 12 variables, derived using remote sensing and GIS technologies, and integrated with historical fire occurrence points. The prediction performance of the models was evaluated using several metrics with a test dataset. The findings underscore the potential of leveraging advanced machine learning techniques and spatial data analysis to enhance the accuracy of forest fire risk prediction. Outcomes are expected to assist policymakers and forest management authorities in developing more effective forest management strategies and fire prevention measures, thereby contributing to the protection of natural resources and the welfare of local communities.

KEY WORDS

Forest fire risk prediction; Machine learning; Random Forest; CNN; LSTM; Geospatial data; GIS; Dak Lak province

1 Introduction

Forest fires are severe natural disasters that pose significant threats to human life, inflict extensive damage on forest ecosystems, and result in severe environmental pollution. These fires drive the decline of both forest coverage and quality, particularly in regions endowed with rich and diverse forest resources such as Vietnam (Tran Quang Bao et al., 2016; Doan et al., 2024). Forest fires release massive quantities of carbon dioxide into the atmosphere, which contributes to global warming, exacerbates the impacts of climate change, and adversely affects both the living environment and public health (Bowman et al., 2009; Chen et al., 2026; Sharma et al., 2026). Climate change has induced phenomena such as decreased precipitation, elevated temperatures, and prolonged dry seasons, all of which are primary catalysts for forest fires (Jolly et al., 2015). Additionally, anthropogenic activities in various regions worldwide have escalated the frequency of forest fires to alarming levels (Balch et al., 2017). Consequently, forest fire risk prediction has emerged as a critical research focus among scientists, aiming to provide early warnings that facilitate effective responses and mitigate damages caused by forest fires.

Traditional methods for forest fire research and forecasting predominantly rely on meteorological indices (Shetinsky, 1994; Arpaci et al., 2013; Venevsky, 2002). However, these approaches exhibit fundamental limitations as they primarily utilize ground level meteorological data. The relatively sparse distribution of local meteorological stations restricts their applicability over expansive geographical areas and compromises the accuracy of fire risk prediction models. These constraints can be overcome by integrating remote sensing data and Geographic Information Systems (GIS) with modern spatial analysis techniques based on artificial intelligence to forecast forest fire risks.

Recent global studies on forest fire risk prediction have predominantly focused on both traditional machine learning and deep learning models. Sarkar et al. (2024) employed five traditional machine learning models such as Random Forest (RF), Classification and Regression Trees (CART), Support Vector Machine (SVM), Boosted Regression Trees (BRT), and Multivariate Adaptive Regression Splines (MARS) to identify forest fire-prone areas in Northeast India. This study utilized 32 input variables encompassing topographical, climatic, and ecological factors. The results demonstrated that RF was the most suitable model for the study area, achieving an Area Under the Curve (AUC) value of 0.87, a metric derived from the receiver operating characteristic curve which is independent of arbitrary decision criteria (Fielding and Bell, 1997), and classifying 89.27% of the area as susceptible to forest fires. Similarly, the authors utilized traditional machine learning algorithms, specifically SVM and RF, integrated with fire point data verified via MODIS imagery (Mishra, Manoranjan, et al.,2024). The model inputs comprised 19 factors, including 3 topographical, 7 climatic, 4 ecological, and 5 anthropogenic variables. The findings revealed exceptionally high Precision scores for RF and SVM, at 0.94 and 0.89 respectively, for the study areas in Angul and Odisha states, India. Purnama et al. (2024) conducted a forest fire risk assessment in Turkey using a dataset with diverse conditioning factors such as precipitation, soil moisture, temperature, humidity, wind speed, land cover, elevation, aspect, slope, distance to roads/power grids, and population density. These were combined with traditional machine learning models including Decision Tree (DT), Naïve Bayes (NB), RF, Artificial Neural Network (ANN), and SVM. The study classified warning levels into four categories: very low, low, medium, and high. Based on evaluations using various metrics, the Kappa index, and Cross-Validation (CV), RF exhibited the highest performance (Accuracy: 0.80, F1-score:

0.78, Kappa: 0.71, Mean CV: 0.71). Jiang, Wenyu, et al. (2024) delved into the spatial and temporal distribution characteristics of forest fires, focusing on four critical fire-driving factors: topography, vegetation, weather, and human activities. The construction of a forest fire dataset comprising 11,507 incidents in Guangdong, south China, paved the way for implementing a risk assessment model based on a Convolutional Neural Network (CNN). This model leveraged its capability to extract high-level features from fire-inducing factors to predict the fire risk probability for specific spatial locations. Experimental comparisons highlighted the superior performance of the CNN model (Accuracy: 0.897, F1-score: 0.849, AUC: 0.962) over traditional machine learning methods. Peng, Yuwen, et al. (2024) presented a time-series prediction model based on a Long Short-Term Memory (LSTM) deep learning network and spatial interpolation to address forecasting challenges in Sichuan province, southwest China. The results indicated that LSTM was the most robust method, yielding highly accurate predictive outcomes.

In Vietnam, several studies have also integrated traditional machine learning and deep learning algorithms with geospatial data for forest fire risk prediction. Do et al. (2024) utilized a hybrid model combining an Adaptive Neuro Fuzzy Inference System with the Artificial Bee Colony algorithm (ABC ANFIS) to assess forest fire susceptibility in Dak Nong province. The results revealed that the majority of forests in the region faced high to dangerous fire risks, particularly in bamboo and dipterocarp forest areas. Pham et al. (2020) evaluated the capability of traditional machine learning methods such as NB, DT, and Multivariate Logistic Regression (MLR) to predict and map forest fire risk across Pu Mat National Park, Nghe An province. The modeling approach processed data from 57 historical fires and 9 input factors, specifically elevation, slope, aspect, drought index, river density, land cover, and distance from roads and residential areas. A recent study by Tran Xuan Truong et al. (2023) proposed a novel modeling approach based on TFDeepNN, which is an advanced deep learning network built upon the TensorFlow framework to enhance spatial feature extraction capabilities, and GIS for forest fire risk modeling. Focusing on the tropical forests of Phu Yen province, the model incorporated 306 historical forest fire locations (2019 to 2023) and 10 conditioning factors, using RF, SVM, and Logistic Regression (LR) as baseline models for comparison. The TFDeepNN model (AUC = 0.873) demonstrated superior predictive performance, outperforming the three baselines. For forest fire danger mapping, Tran Xuan Truong, et al. (2023) developed a novel ensemble learning model, referred to as HHO RSCDT, combining three distinct components: Random Subspace (RS), Credal Decision Tree (CDT), and Harris Hawks Optimizer (HHO). In this specific architecture, the Harris Hawks Optimizer metaheuristic algorithm is utilized to optimize the ensemble framework of RS and CDT to achieve better accuracy. Validated using the Phu Yen province dataset, the HHO-RSCDT model achieved an AUC value of 0.911, proving to be a highly effective tool for fire risk modeling. Bui et al. (2018) introduced a novel machine learning method named DFP-MnBpAnn for spatial modeling of forest fire risk in the tropical forests of Lam Dong province. This hybrid method applies a Differential Flower Pollination optimization algorithm to fine tune the weights of a Multilayer Perceptron Artificial Neural Network. Experimental results proved that the proposed DFP-MnBpAnn model outperformed other benchmark methods, achieving a classification accuracy of 88.43%.

While the aforementioned studies successfully utilized topographical, meteorological, ecological, and anthropogenic data layers to demonstrate the potential of various predictive algorithms, a substantial research gap remains in the methodological framework. Specifically, previous regional research frequently overlooks the critical evaluation of feature importance tailored to specific local fire inventory datasets. Furthermore, there is a pronounced lack of systematic approaches to rigorously optimize the internal parameters of the applied algorithms. Without rigorous feature selection and parameter optimization, predictive models often suffer from overfitting and reduced spatial generalization. To profoundly address this gap, our study proposes an advanced comprehensive framework. This framework meticulously integrates traditional machine learning and deep learning architectures with robust Feature Selection and Hyperparameter Tuning techniques. By systematically isolating the most influential fire driving factors and optimizing algorithmic configurations, this research aims to significantly elevate the accuracy, reliability, and practical applicability of forest fire risk forecasting using geospatial data.

2 Materials and methodology

2.1 Study area and Data

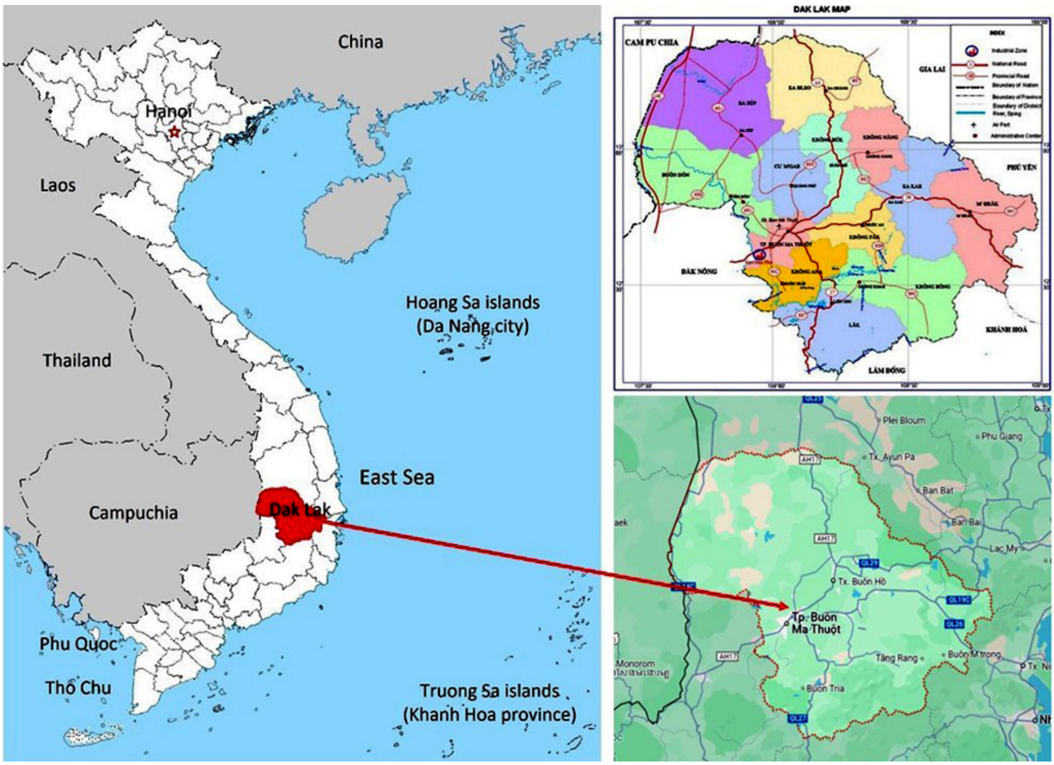

Figure 1: Map of the study area (Dak Lak province, Vietnam).

Study area. Dak Lak is one of the five provinces constituting the Central Highlands region of Vietnam. Located in the heart of this region, it lies between longitudes 107°28’57”E and 108°59’37”E, and latitudes 12°09’45”N and 13°25’06”N (Figure 1). Topographically, the province features a complex terrain characterized by alternating valleys and plateaus interspersed with high and medium-high mountains, gradually descending in elevation from the southeast to the northwest. The region experiences a humid tropical monsoon climate characteristic of the highlands, heavily influenced by the hot and dry Southwest monsoon (Trinh et al., 2025). It exhibits two distinct seasons, notably a prolonged dry season lasting from November to April. During this period, a significant drop in humidity, prevailing strong northeast winds, and high evaporation rates lead to severe drought conditions. Recently, exacerbated by deforestation and climate change, Dak Lak has emerged as a critical hotspot nationwide, with extensive forest areas frequently placed under Level IV (High) and Level V (Extremely Dangerous) fire weather warnings, indicating an imminent and severe risk of forest fires. Given these pronounced vulnerabilities, this study selected Dak Lak province as the primary research area.

Data. Forest fires are complex phenomena triggered by a multitude of interconnected factors, primarily driven by topographical characteristics, meteorological conditions, land cover types and anthropogenic activities. Based on the specific environmental and socio-economic conditions of the study area, and building upon insights derived from previous literature, this study selected 12 conditioning factors, represented by 12 corresponding thematic map layers. These include: vegetation cover, land surface temperature (LST), mean monthly precipitation, wind speed, evapotranspiration, population density, land use status, elevation, slope, aspect, distance to roads, and distance to water bodies.

These spatial data layers can be systematically categorized into the following groups:

Remote Sensing Data Group:

The remote sensing data utilized in this study comprised Landsat 8/9 optical satellite imagery, which was acquired and processed directly on the Google Earth Engine (GEE) platform. Landsat 8/9 images captured between November 1, 2021, and April 30, 2022, underwent rigorous cloud and cloud-shadow filtering. This preprocessing workflow was executed using the Google CloudScore algorithm for cloud masking and the Temporal Dark Outlier Mask (TDOM) method for identifying and masking cloud shadows. Designed to eliminate cloud contamination, this entire procedure was conducted within the GEE environment. The resulting high-quality, cloud-free Landsat composites were subsequently used to compute key input variables for the machine learning models: the Normalized Difference Vegetation Index (NDVI) - representing vegetation cover, and Land Surface Temperature (LST) (Figure 2). Specifically, the Normalized Difference Vegetation Index (NDVI) was calculated to quantify vegetation density and health using the normalized mathematical ratio between the Near Infrared (Band 5) and Red (Band 4) spectral bands of the Landsat sensors (Tucker, 1979). The formula involves dividing the difference between the Near Infrared and Red bands by their sum. The Land Surface Temperature (LST) estimation involved a rigorous radiometric procedure as described by Hung (2014). Initially, the Thermal Infrared Sensor (Band 10) digital numbers were converted into Top of Atmosphere spectral radiance. This radiance was subsequently transformed into satellite brightness temperature. Finally, the brightness temperature was corrected utilizing the land surface emissivity, which was dynamically estimated based on the vegetation proportion derived from the NDVI, to retrieve the accurate and final LST values. All acquired data layers were resampled to a consistent spatial resolution of 30 m to ensure spatial uniformity across the dataset.

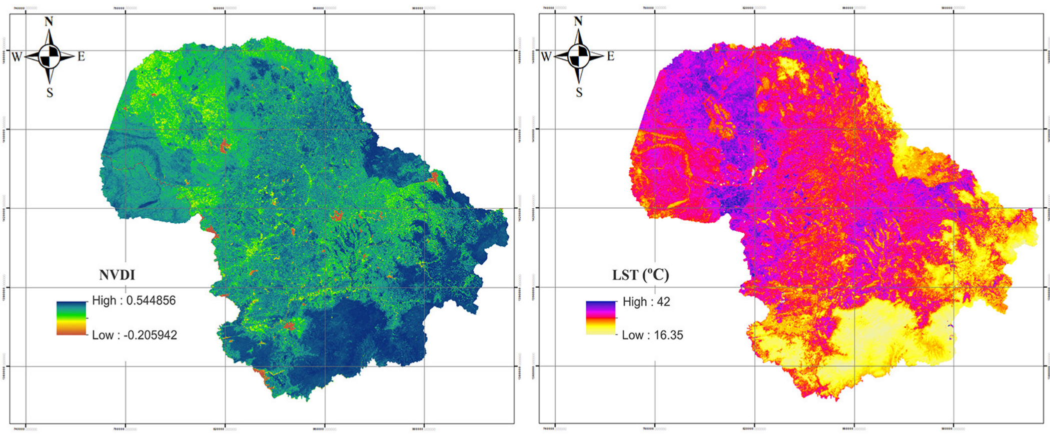

Figure 2: Spatial distribution of Vegetation Cover (NDVI) and Land Surface Temperature (LST) within the study area.

Topographical Data Group:

The Digital Elevation Model (DEM) for the study area was acquired and processed from the Shuttle Radar Topography Mission (SRTM) dataset, accessed via the GEE data catalog. The DEM provides critical insights into topographical variations, yielding essential parameters that significantly influence and govern forest fire behavior. From the raw DEM data, specific spatial input layers namely elevation, slope, and aspect, were extracted for the entire study area (Figure 3). Consistent with the remote sensing dataset, these extracted topographical layers were subsequently resampled to a uniform spatial resolution of 30 m.

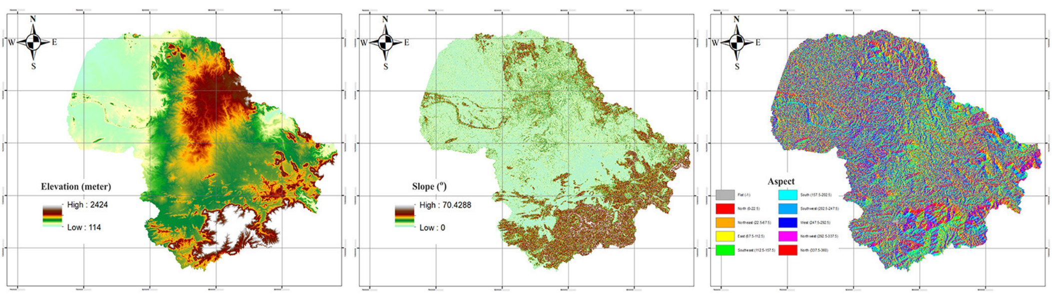

Figure 3: Topographical data layers (Elevation, Slope, and Aspect) of the study area.

Meteorological Data Group:

Meteorological variables are crucial for comprehending the drivers and dynamics of forest fire behavior. This study acquired data related to wind speed and mean monthly precipitation (Figure 4) from the WorldClim database (Fick and Hijmans, 2017). Upon acquisition, these meteorological data layers were subsequently resampled to a uniform spatial resolution of 30 m.

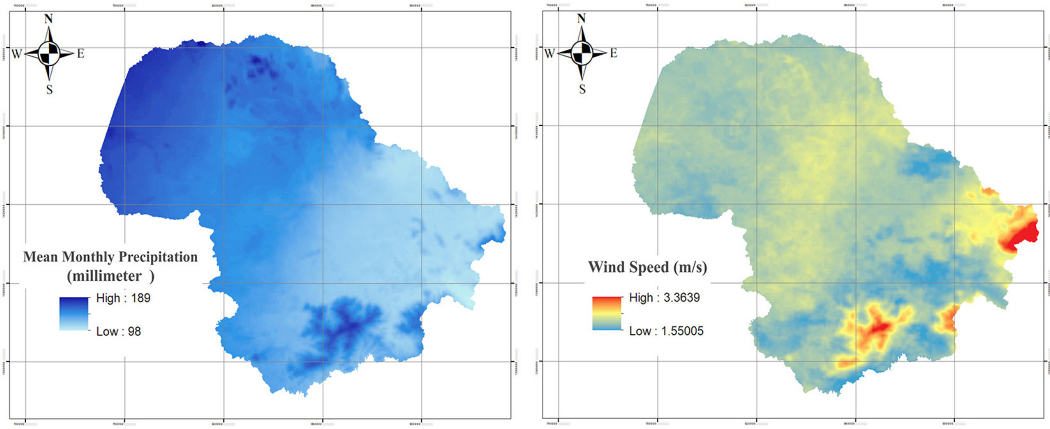

Figure 4: Spatial distribution of wind speed and mean monthly precipitation within the study area.

Multi-source Data Group:

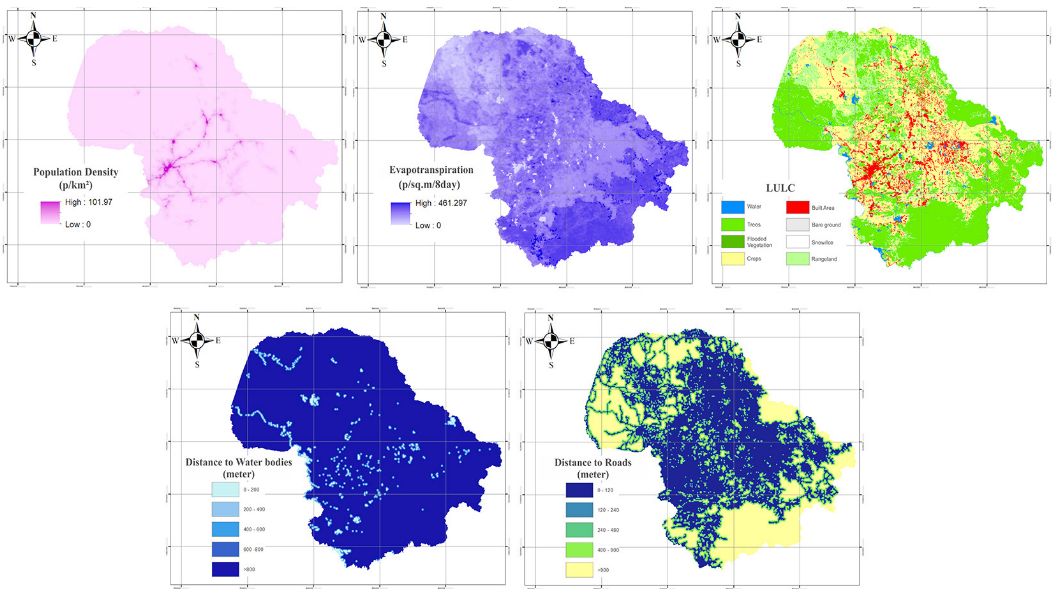

Figure 5: Data layers of population density, evapotranspiration, LULC, thematic layers of distance to roads and distance to water bodies within the study area.

Population density data was retrieved from the WorldPop dataset, available within the GEE data catalog. Surface evapotranspiration was derived from MODIS imagery, also accessed via the GEE platform. Meanwhile, Land Use and Land Cover (LULC) status data was obtained from ESRI Sentinel 2 10m Land Use and Land Cover Timeseries Downloader. The spatial distribution of these multi source data layers is visually presented in Figure 5.

Transportation (road network) and hydrological data for the study area were collected from OpenStreetMap (OpenStreetMap contributors, 2025), which is currently one of the most highly detailed, freely available geographic databases for these features. To evaluate their spatial influence on forest fire risk, the research team conducted a proximity analysis, categorizing the distances from these features into five distinct buffer classes. Specifically, the distance to roads layer was classified into five intervals: 0 to 120 m, 120 to 240 m, 240 to 480 m, 480 to 900 m, and greater than 900 m. These precise intervals were adopted from established regional forest fire risk assessments (Tran Xuan Truong et al., 2023), which empirically demonstrated that anthropogenic ignition probabilities are highly concentrated near transportation corridors and exponentially diminish across these specific spatial scales. Similarly, the distance to water bodies layer was divided into five intervals: 0 to 200 m, 200 to 400 m, 400 to 600 m, 600 to 800 m, and greater than 800 m. The selection of these specific thresholds was logically determined by the localized microclimatic impacts of hydrological features; specifically, the ambient cooling effects and elevated soil moisture gradients naturally dissipate over increments of 200 meters within typical tropical forest terrains, rendering proximity distances beyond 800 meters largely insignificant regarding fire deterrence. Following this spatial processing, these thematic layers were also resampled to a consistent spatial resolution of 30 m to align with the rest of the dataset.

Forest Fire and Non-Fire Inventory Data:

Historical forest fire occurrence data for this study were acquired from the open source portal of the Vietnam Forest Protection Department (Vietnam Forest Protection Department, 2025) and the Fire Information for Resource Management System of NASA (FIRMS, 2025). Data collection spanning the dry season from November 2021 to April 2022 recorded a total of 6,088 forest fire points within Dak Lak province. To establish a balanced dataset for the binary classification models, an equal number of absence points (non-fire points) were randomly generated across the unburned forest areas within the study region. The spatial distribution of both active fire locations and absence points is visualized in Figure 6. Subsequently, this comprehensive inventory was processed and partitioned into two distinct subsets: 70 percent of the data was allocated to the training dataset to calibrate the machine learning models, while the remaining 30 percent was reserved for the testing dataset to evaluate the predictive accuracy of the forest fire risk models in the study area. This 70 to 30 data splitting ratio was deliberately chosen because it is a highly established standard in machine learning research, ensuring that the algorithm receives a sufficiently large data volume for robust pattern recognition during the training phase while preserving an adequately sized, independent subset for rigorous performance validation.

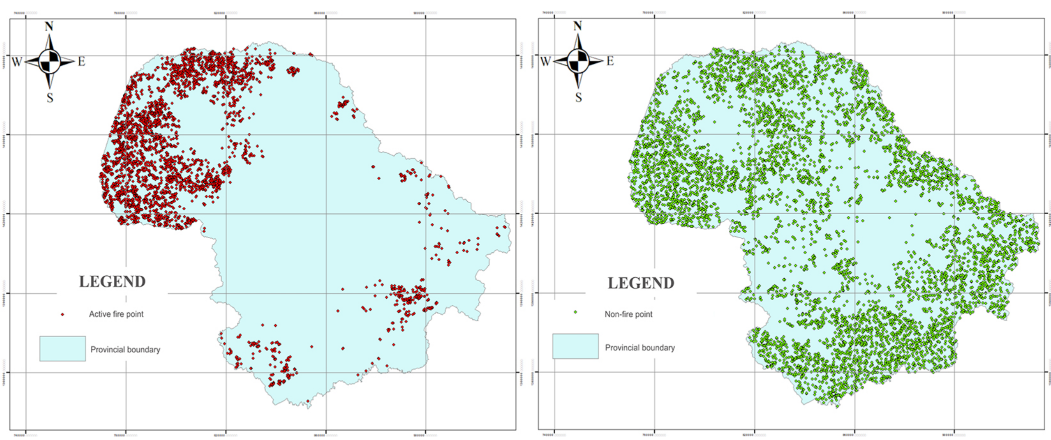

Figure 6: Spatial distribution of active fire points and non-fire points within the study area.

Theoretically, all the aforementioned conditioning factors influence forest fire susceptibility. However, their relative importance and impact magnitude vary significantly, heavily depending on the specific environmental and geographic characteristics of the study area. Consequently, this study proceeds to conduct a comprehensive feature importance analysis to evaluate the specific contribution of each factor, thereby selecting the most critical variables (Feature Selection) for the subsequent modeling phase.

2.2 Methodology

This capability allows it to identify underlying patterns and temporal trends associated with elevated forest fire risks. This study empirically investigates and compares the predictive performance of a robust traditional machine learning model, namely Random Forest (RF), against these advanced deep learning models (CNN and LSTM). Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks are the most prominently selected architectures among deep learning architectures. Specifically, LSTM - a specialized variant of RNN explicitly designed to overcome the vanishing gradient problem and capture long-term dependencies in sequential data is highly effective for analyzing continuous meteorological time-series data. By applying these algorithms to the experimental dataset of the study area, this research aims to identify the most optimal and accurate model for forest fire susceptibility mapping.

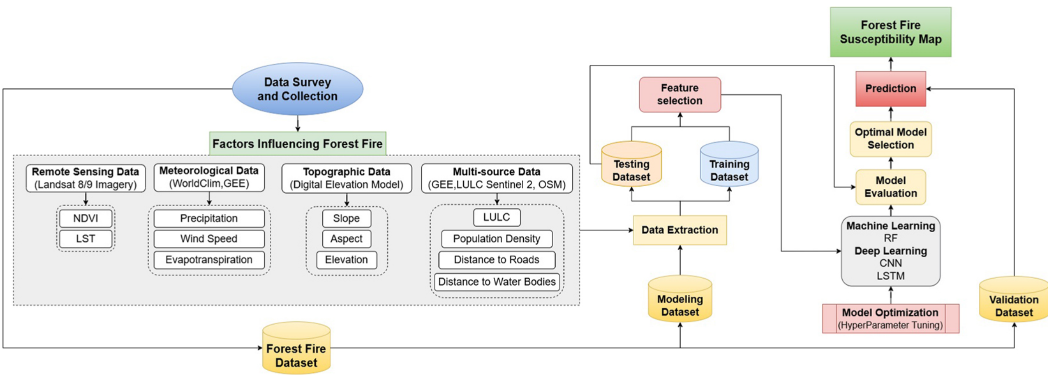

Figure 7: Flowchart of the proposed methodology.

The methodological framework consisted of seven sequential phases integrating geospatial data processing, feature engineering, machine learning optimization, and spatial prediction.

First, a comprehensive survey and analysis of the experimental area and available data sources were conducted, upon which the raw data were acquired. This included Landsat 8/9 satellite imagery, a Digital Elevation Model (DEM), meteorological datasets and other multi-source data necessary to construct the conditioning factors for the forest fire risk prediction models.

Following data acquisition and preprocessing, 12 spatial conditioning layers for the forest fire risk models were extracted and generated from the remote sensing and GIS datasets (Figure 7). All input layers were standardized to a 30 m raster resolution to ensure spatial consistency.

The third phase involved generating the machine learning training dataset. The historical forest fire inventory comprised both fire occurrence points and non-fire (absence) points. Based on the coordinates of these points, pixel values from the 12 conditioning layers were extracted, assigning a label of '1' to fire locations and '0' to non-fire locations. This process generated a dataset containing 12-dimensional feature matrix. Values were normalized to a [0, 1] range and randomly partitioned into a training subset (70%) and a testing subset (30%).

In the fourth faze the study conducted a comprehensive evaluation to determine which spatial layers to retain or discard using Feature Selection (FS) techniques. Based on the dataset's characteristics, three distinct FS approaches were implemented: the Filter method (utilizing Mutual Information [MI] and ANOVA), the Wrapper method (utilizing Recursive Feature Elimination [RFE]), and the Embedded method (utilizing Lasso, Decision Tree, and Extra Trees [ExRT]). Variables consistently demonstrating the lowest importance scores across these algorithms were eliminated, while features exhibiting a significant impact on forest fire occurrence were retained. This refined subset of critical features was then used to construct the finalized training dataset for the predictive models.

In the fifth faze Hyperparameter Tuning was employed to identify the optimal configurations for the models, a crucial step for significantly improving predictive performance. For traditional machine learning algorithms, the Grid Search method was utilized to determine the best parameters. For deep learning architectures (CNN and LSTM), the Keras Tuner tool was applied to optimize hyperparameters, such as the number of neurons and dropout rates.

During the training phase, the training dataset was fed into the respective models. For deep learning models (designed and constructed using TensorFlow/Keras), the training process yielded optimal weights and biases that represented the underlying features of the data. Using these learned parameters, forest fire probabilities were predicted for the entire Dak Lak province. Simultaneously, the models' predictive accuracy was evaluated on the testing dataset using statistical evaluation metrics, Receiver Operating Characteristic (ROC) curves, and Area Under the Curve (AUC) values.

Based on the accuracy assessment derived from the testing dataset, the most robust model for the study area was identified in the sixth faze and used to perform spatial predictions. The input for this prediction phase was a multi-band composite image, stacked in the order of the retained important features. This optimal model computed the fire occurrence probability across the validation dataset. The output is a predictive map containing probability values of fire risk for the study area. Furthermore, the actual fire and non-fire validation points were evaluated against the predicted probability thresholds to verify the superior performance of the chosen model. The overall predictive capabilities of the models were comprehensively assessed using Precision, Recall, F1-score, Accuracy, ROC curves, and AUC metrics.

Finally, the forest fire risk map for the study area was generated based on the optimal predictive model identified in faze 6. The predicted forest fire risk was reclassified into five susceptibility levels based on defined thresholds: Very Low, Low, Medium, High, and Dangerous (Extreme Danger). These thresholds can be adjusted depending on the predicted probability distribution and the specific natural conditions of the region. The final forest fire susceptibility map was symbolized and compiled using ArcMap software for publication and operational use.

3 Results and discussion

Spatial Data Analysis and Feature Selection:

The importance of the features within the dataset was evaluated using Feature Importance Scores and the F score derived from the Filter ANOVA method. The distance to water bodies feature, explicitly denoted as 'ThuyVan' in the subsequent evaluation charts, consistently exhibited very low importance scores across all six evaluated methods. Consequently, this feature was eliminated, and the dataset retained the remaining 11 best performing features to serve as inputs for the machine learning and deep learning models. These results are comprehensively illustrated in the graphs in Figure 8.

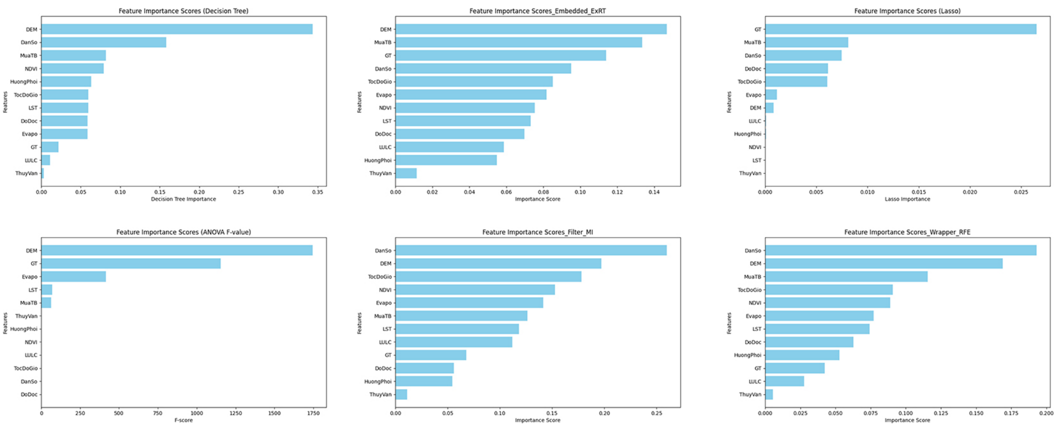

Figure 8: Feature importance evaluation of the training dataset.

Hyperparameter Tuning Results:

The Random Forest (RF) model utilized n estimators (the number of decision trees) as its primary input hyperparameter. Meanwhile, the deep learning models, CNN and LSTM, employed the number of neurons and dropout rate as input hyperparameters and were trained over 200 epochs with a batch size of 4096. Table 1 presents the results of the hyperparameter tuning process for the traditional machine learning model, identifying its optimal hyperparameters. Correspondingly, Table 2 details the hyperparameter tuning results for the deep learning models. In both tables, the score represents the accuracy evaluation of each model on the testing dataset. These optimal parameters were subsequently selected to initialize the respective algorithms during the training phase.

Table 1: Hyperparameter tuning results for the traditional machine learning model.

No.	Model	n_estimators	Score
1	**RF**	**50**	**0.88**
		20	0.87
		10	0.86
		5	0.86

Table 2: Hyperparameter tuning results for the deep learning models.

No.	Model	Neuron	Dropout	Score
1	**LSTM**	**128**	**0.3**	**0.83**
		128	0.1	0.81
		64	0.2	0.8
		128	0.2	0.79
2	**CNN**	**128**	**0.2**	**0.8**
		128	0.3	0.79
		128	0.1	0.78
		64	0.2	0.78

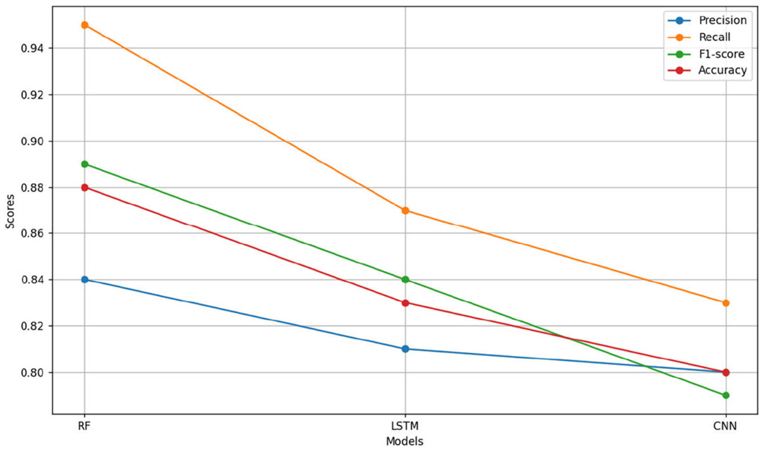

Figure 9: Performance metrics chart of the machine learning models.

To determine the most robust and suitable model for predicting forest fire risk in the study area, this study evaluated model performance using standard statistical metrics namely Precision, Recall, F1 score, and Accuracy alongside the ROC curve and AUC values. As observed from the results in Table 3 and the chart in Figure 9, using the inputted forest fire training dataset, the RF machine learning model achieved the highest Precision (0.84) and a remarkably high Recall (0.95), resulting in a high F1 score (0.89) and Accuracy (0.88). Thus, it demonstrated the best predictive performance among all evaluated models.

Table 3: Performance evaluation results of the predictive models.

No.	Model	P	R	F1	Accuracy	AUC
1	**RF**	0.84	0.95	0.89	0.88	0.95
2	**LSTM**	0.81	0.87	0.84	0.83	0.89
3	**CNN**	0.80	0.83	0.79	0.80	0.84

Based on the ROC curves and AUC values of the RF, LSTM, and CNN models presented in Figure 10, the RF model achieved an outstanding AUC of 0.95 for both classes (0: non-fire and 1: fire). The ROC curves for both classes closely approach the top-left corner of the plot, indicating the RF model's excellent capability to distinguish between classes, characterized by a high True Positive Rate (TPR) and a low False Positive Rate (FPR). This result proves that the RF model possesses Dangerous accuracy in classifying both non-fire and fire instances. Compared to LSTM and CNN, RF exhibits superior reliability and a more robust discriminative capacity between the classes. Through comprehensive performance evaluation using statistical metrics, ROC curves, and AUC values, this study selects the RF model as the optimal forest fire warning model. It is highly suitable for constructing the accurate forest fire susceptibility map for Dak Lak province.

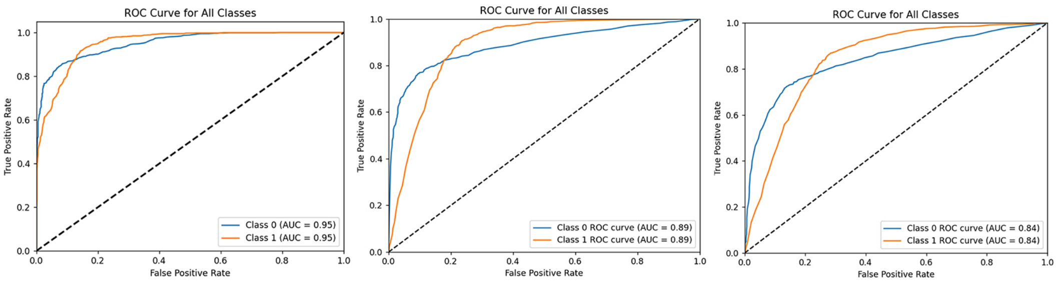

Figure 10: ROC curves and AUC values of the RF, LSTM, and CNN models.

Evaluation of Prediction Results using the Validation Dataset:

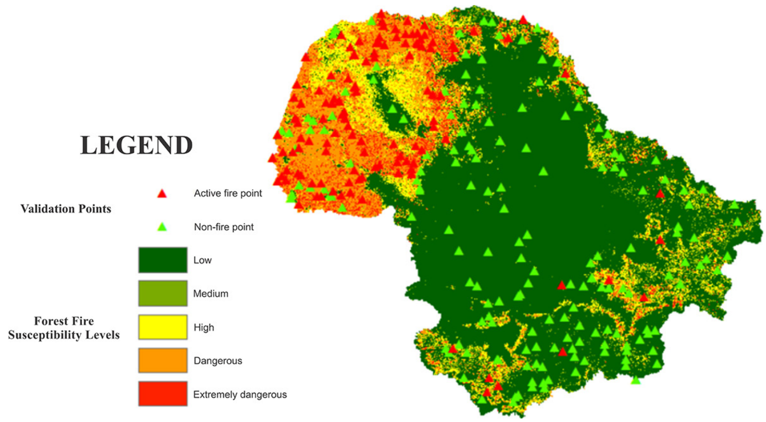

Figure 11: Fire risk prediction results derived from the RF model and the actual fire point validation dataset.

The optimal weights obtained during the training process of the Random Forest (RF) model were utilized to predict forest fire risk. The input for this predictive phase was a multidimensional composite raster generated from the optimal feature layers identified in the training dataset. The resulting output is a continuous raster image containing pixel values that represent the probability of fire occurrence. Subsequently, thresholding was applied using the specific probability values of 0.18, 0.29, 0.56, and 0.84. These precise threshold values were determined based on empirical spatial analysis and referenced from established forest fire danger mapping methodologies in the region (Tran Xuan Truong et al., 2023) to effectively reclassify the predicted continuous probabilities into five distinct susceptibility levels: Low, Medium, High, Dangerous, and Extremely Dangerous. The spatial distribution of these predicted risk levels, alongside the actual fire point validation dataset, is visually presented in Figure 11.

To assess the accuracy of the forest fire risk predictions generated by the selected model, this study employed a validation dataset extracted from the initially collected fire and non-fire inventory. This validation dataset, comprising 122 forest fire points and 170 non-fire points, was strictly utilized to evaluate the predictive accuracy of the model.

Table 4: Accuracy assessment using the historical fire validation dataset.

Model	Forest Fire Susceptibility Levels					
	Type of points	**Low**	**Medium**	**High**	**Dangerous**	**Extremely Dangerous**
RF	Fire points	0	1	6	45	70
	Non-fire points	110	30	15	12	3

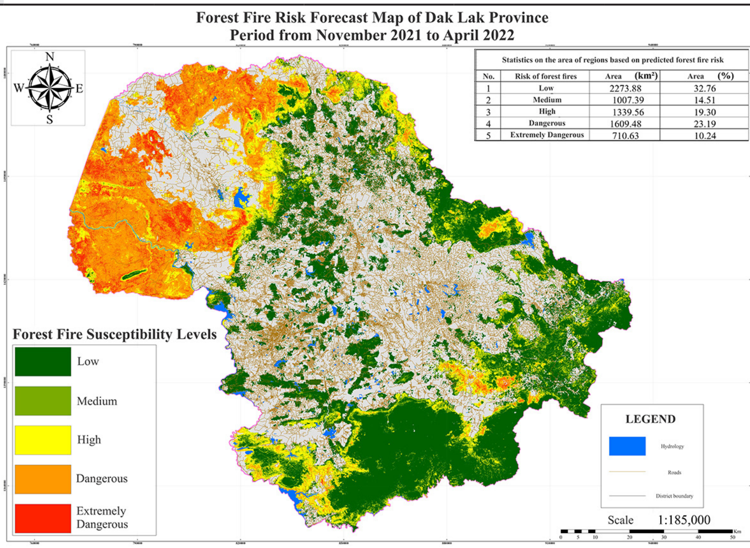

Figure 12: Forest fire susceptibility map of the study area.

An analysis of the evaluation result in Table 4 shows that 115 forest fire validation points (accounting for 94.26% of the total validation fire points) are located within areas predicted as "Dangerous" and "Extremely Dangerous" risk. Specifically, 45 fire points fall into the "Dangerous" category and 70 within the "Extremely Dangerous" category. The number of fire points in the High, Medium, and Low risk zones is negligible , representing only 0.82% of all fire points. For the non-fire locations, 140 points (82.35% of all non-fire points) fall within areas predicted as "Low" and "Medium" risk, with 110 points in the "Low" category and 30 points in the "Medium" category. In the High, Dangerous, and Extremely Dangerous risk zones, the proportion of non-fire points remains very low, reaching only 8.82% of the total non-fire points. These results indicate that the RF model exhibit exceptionally high accuracy and spatial reliability, proving highly suitable for the development and deployment of a robust forest fire susceptibility map for Dak Lak province, which is comprehensively visualized in Figure 12.

4 Conclussion

This study investigated the application of machine learning models integrated with geospatial data for forest fire risk prediction. The findings demonstrate that geospatial data can be effectively leveraged to construct robust thematic input layers for predictive modeling. Beyond the conventionally utilized variables in forest fire forecasting, this research incorporated and evaluated additional conditioning factors, specifically surface evapotranspiration, land use/land cover status, distance to roads and distance to water bodies. Furthermore, the study proposed a systematic approach to eliminate non-contributing features, alongside rigorous empirical testing of various machine learning architectures to identify the optimal hyperparameter configurations for training and spatial forecasting. Through comprehensive analysis and performance evaluation, the Random Forest (RF) model was identified as the most highly suitable and scalable algorithm for developing forest fire susceptibility maps across Dak Lak province. The outcomes of this study provide essential decision-support tools, empowering forest managers and relevant authorities to formulate more effective forest management and fire prevention strategies, thereby contributing significantly to the protection and conservation of forest resources and ecological habitats**.**

Literature and references:

Arpaci, A., Eastaugh, C. S., & Vacik, H. (2013). Selecting the best performing fire weather indices for Austrian ecoregions. Theoretical and Applied Climatology, 114, 393–406. https://doi.org/10.1007/s00704-013-0839-7.

Balch, J. K., Bradley, B. A., Abatzoglou, J. T., Nagy, R. C., Fusco, E. J., & Mahood, A. L. (2017). Human-started wildfires expand the fire niche across the United States. Proceedings of the National Academy of Sciences, 114(11), 2946-2951.

Bowman, D. M., Balch, J. K., Artaxo, P., Bond, W. J., Carlson, J. M., Cochrane, M. A., ... & Pyne, S. J. (2009). Fire in the Earth system. science, 324(5926), 481-484.

Bui, D. T., Van Le, H., & 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.

Chen, W., Zhang, Y., Zou, Y., Zhang Z. (2026). Climate feedback of forest fires amplified by atmospheric chemistry. Nature Geoscience 19, 402 - 405. https://doi.org/10.1038/s41561-026-01926-1

Do, T. A. T., Tran, H. D., & Do, A. N. T. (2024). Classifying forest cover and mapping forest fire susceptibility in Dak Nong province, Vietnam utilizing remote sensing and machine learning. Ecological Informatics, 79, 102392.

Doan, T.N.P., Trinh, L.H., Zablotskii, V.R., . . . & Le, V.P. (2024). Forest fire risk assessment and mapping using remote sensing and GIS techniques: a case stude in Nghe An province, Vietnam. Izvestiya, Atmospheric and Oceanic Physics, 60(9), 982 - 991.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International journal of climatology, 37(12), 4302-4315.

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental conservation, 24(1), 38-49.

Hung, T. L. (2014). Studies of land surface temperature distribution using multispectral image Landsat. Vietnam Journal of Earth sciences, 36, 4145.

Jiang, W., Qiao, Y., Zheng, X., Zhou, J., Jiang, J., Meng, Q., ... & Wang, F. (2024). Wildfire risk assessment using deep learning in Guangdong Province, China. International Journal of Applied Earth Observation and Geoinformation, 128, 103750.

Jolly, W. M., Cochrane, M. A., Freeborn, P. H., Holden, Z. A., Brown, T. J., Williamson, G. J., & Bowman, D. M. (2015). Climate-induced variations in global wildfire danger from 1979 to 2013. Nature communications, 6(1), 7537.

Mishra, M., Guria, R., Baraj, B., Nanda, A. P., Santos, C. A. G., da Silva, R. M., & Laksono, F. A. T. (2024). Spatial analysis and machine learning prediction of forest fire susceptibility: a comprehensive approach for effective management and mitigation. Science of the Total Environment, 926, 171713.

National Aeronautics and Space Administration (NASA), Fire Information for Resource Management System (FIRMS) (2025). https://firms.modaps.eosdis.nasa.gov, Accessed 12 August, 2025.

OpenStreetMap contributors (2025). Planet dump. https://planet.osm.org, Accessed 12 May, 2025.

Peng, Y., Su, H., Sun, M., & Li, M. (2024). Reconstructing historical forest fire risk in the non-satellite era using the improved forest fire danger index and long short-term memory deep learning-a case study in Sichuan Province, southwestern China. Forest Ecosystems, 11, 100170.

Pham, B. T., Jaafari, A., Avand, M., Al-Ansari, N., Dinh Du, T., Yen, H. P. H., ... & Tuyen, T. T. (2020). Performance evaluation of machine learning methods for forest fire modeling and prediction. Symmetry, 12(6), 1022.

Purnama, M. I., Jaya, I. N. S., Syaufina, L., Çoban, H. O., & Raihan, M. (2024, March). Predicting forest fire vulnerability using machine learning approaches in The Mediterranean Region: a case study of Türkiye. In IOP Conference Series: Earth and Environmental Science (Vol. 1315, No. 1, p. 012056). IOP Publishing.

Sarkar, M. S., Majhi, B. K., Pathak, B., Biswas, T., Mahapatra, S., Kumar, D., ... & Nautiyal, S. (2024). Ensembling machine learning models to identify forest fire-susceptible zones in Northeast India. Ecological Informatics, 81, 102598.

Sharma, D., Kashyap, M. P., Das, D., . . . & Taludar, N.C. (2026). Assessing fire-induced tree cover loss and its contribution to carbon emission in BRICS + nations. Discover Environment 4, 7 (2026). https://doi.org/10.1007/s44274-025-00477-3

Shetinsky, E. A. (1994). Protection of forests and forest pyrology. Ecology, Moscow, 209.

Tran, Q. B., Nguyen, T. C., & Mai, H. A. (2016). Research on developing automated software for early forest fire detection from ground observation stations. Journal of Forestry Science, (3), 4538-4546.

Trinh, L.H., Le, V.P. & Nguyen, T.T.N. (2025). Pre-classification method for detectiong vegetation cover changes in Dak Lak province (Vietnam). Bulletin of Geography. Physical Geography Series, 28, 5 - 16. http://doi.org/10.12775/

bgeo-2025-0001.

Truong, T. X., Nhu, V. H., Phuong, D. T. N., Nghi, L. T., Hung, N. N., Hoa, P. V., & Bui, D. T. (2023). A new approach based on TensorFlow deep neural networks with ADAM optimizer and GIS for spatial prediction of forest fire danger in tropical areas. Remote Sensing, 15(14), 3458.

Truong, T. X., Phuong, N. D. T., Nghi, L. T., & Nhu, V. H. (2023). A novel HHO-RSCDT ensemble learning approach for forest fire danger mapping using GIS. Vietnam Journal of Earth Sciences, 45, 338-356.

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

Venevsky, S., Thonicke, K., Sitch, S., & Cramer, W. (2002). Simulating fire regimes in human-dominated ecosystems: Iberian Peninsula case study. Global Change Biology, 8(10), 984-998.

Vietnam Forest Protection Department (2025). Forest fire warning and monitoring system. https://watch.pcccr.vn, Accessed 12 August, 2025.

Trinh L. H., Vu V. T., Le Q. T., Tran X. B. (2026). Forest fire risk mapping using machine learning technique based on GIS and remote sensing data, a case study in Dak Lak province, Vietnam. Geodetski vestnik, 70 (3), 421-436.
DOI: 10.15292/geodetski-vestnik.2026.03.421-436

Le Hung Trinh, Associate Prof., Ph.D.
Le Quy Don Technical University,
236 Hoang Quoc Viet Str., Nghia Do ward, Hanoi, Vietnam
e-mail: trinhlehung@lqdtu.edu.vn

Quang Tu Le, Postgraduate student
Le Quy Don Technical University,
236 Hoang Quoc Viet Str., Nghia Do ward, Hanoi, Vietnam
e-mail: tulequang114213@gmail.com

Van Truong Vu, Ph.D.
Le Quy Don Technical University,
236 Hoang Quoc Viet Str., Nghia Do ward, Hanoi, Vietnam
e-mail: truongvv@mta.edu.vn

Xuan Bien Tran, Associate Prof., Ph.D.
Hanoi University of Natural Resources and Environment,
Hanoi, Vietnam
e-mail: txbien.ph@hunre.edu.vn

RECENZIRANI ČLANKI | PEER-REVIEWED ARTICLES

GIS analiza dostopnosti in opremljenosti prostora ob tramvajski progi v mestu Oran (Alžirija)

Spatial Accessibility and Service Coverage of the Oran Tramway (Algeria): A GIS-Based Analysis

Karima Berghout, Naziha Lamri, Abdelhalim Bendib

UDK: 711.4:004.65GIS:624.21.095.325(65)
Klasifikacija prispevka po COBISS.SI: 1.01
Prispelo: 18. 2. 2026
Sprejeto: 12. 6. 2026

DOI: 10.15292/geodetski-vestnik.2026.03.437-459
SCIENTIFIC ARTICLE
Received: 18. 2. 2026
Accepted: 12. 6. 2026

SI | EN

IZVLEČEK

Študija ocenjuje prostorsko učinkovitost tramvajske proge v mestu Oran z uporabo GIS pristopa, ki vključuje analizo storitvenih funkcij, vpetost v prometni sistem, gostoto prebivalcev in dostopnost javnih objektov. Za identifikacijo dobro in premalo oskrbovanih območij v 2 kilometra dolgem koridorju vzdolž tramvajske proge je bila uporabljena analiza prometnega omrežja, interpolacija z inverznim uteževanjem razdalj (IDW), ocena gostote jeder in Ripleyjeva K-funkcija. Rezultati kažejo, da je 57 % preučevanega območja na oddaljenosti 750 metrov od tramvajskih postajališč, medtem ko je 25 % območja več kot 1000 metrov oddaljeno od tramvajskih postajališč, ki so obenem nezadostno opremljena z javnimi in storitvenimi funkcijami. Najboljšo pokritost imajo osrednje gosto naseljene soseske, s povprečno gostoto 101,32 prebivalca na 100 m², v oddaljenosti do 300 metrov od postajališč, v katerih je tudi približno 28 % javnih objektov, zlasti športnih, izobraževalnih in kulturnih, medtem ko so zdravstvene in upravne storitve premalo zastopane. Zmerno koncentracijo javnih objektov okrog tramvajskih postajališč potrjuje tudi Ripleyjeva K-funkcija. Ugotovitve poudarjajo potrebo po bolj uravnoteženem multimodalnem načrtovanju, zlasti na nezadostno oskrbovanih obrobnih območjih, s čimer bi izboljšali učinkovitost prometnega sistema in zagotovili uravnotežen prostorski razvoj.

ABSTRACT

This study assesses the spatial efficiency of the Oran tramway using an integrated GIS-based approach combining service area analysis, network accessibility, population density, and public facility distribution. Network analysis, Inverse Distance Weighted (IDW) interpolation, Kernel Density Estimation, and Ripley's K-function were applied to identify well-served and underserved areas within a 2 km corridor along the tramway line. Results show that 57% of the study area lies within 750 m of tramway stops, while 25% is located beyond 1,000 m and is relatively underserved. Central and densely populated neighbourhoods have the best coverage, with an average population density of 101.32 inhabitants per 100 m² within 300 m of stops. About 28% of public facilities are located in areas with very high service coverage (<300 m), particularly sports, educational, and cultural facilities, whereas healthcare and administrative services are underrepresented. Ripley's K-function confirms significant clustering of public facilities around tramway stops. The findings highlight the need for more equitable multimodal planning, particularly in peripheral underserved areas, to improve spatial justice and system performance.

KLJUČNE BESEDE

dostopnost s tramvajem, mesto Oran, načrtovanje urbane mobilnosti, uravnotežen prostorski razvoj, opremljenost z javnimi funkcijami

KEY WORDS

Tramway accessibility, Oran city, Urban mobility planning, Spatial equity, Public facility coverage