www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 17(1), Jan-April 2026, 378-385. ISSN: 2007-1558. https://doi.org/10.61467/2007.1558.2026.v17i1.1181

Design of a Predictive Model Using AI and GIS for the Management and Optimization of Resources in Forest Fires in Eastern Michoacán

Samuel Efren Viñas Alvarez[1], Claudia Teresa González Ramírez[1], Gabriel Mora Patiño[1], José Luis Martínez Alcantar[1]

[1]Tecnológico Nacional de México / I.T. Zitácuaro

E-mails: samuel.va@zitacuaro.tecnm.mx , claudia.gr@zitacuaro.tecnm.mx, l21650366@zitacuaro.tecnm.mx, jose.ma@zitacuaro.tecnm.mx

Abstract. This study describes a design and validation proposal for an artificial intelligence–based predictive model to estimate the area affected by forest fires and support the optimisation of operational resource use in the eastern region of Michoacán, Mexico. Using a dataset comprising 930 historical records (2015–2024), three supervised learning algorithms were evaluated: Multilayer Perceptron (MLP), Random Forest (RF), and XGBoost. The MLP model, optimised using L2 regularisation, Dropout, and cross-validation, achieved the highest performance, with a coefficient of determination (R^2) of 0.76, compared with RF (0.57), XGBoost (0.51), and a weighted average ensemble model (0.60). The results were integrated into interactive cartographic platforms using QGIS and Leaflet JS, enabling the visualisation of areas with higher fire susceptibility and the prioritisation of interventions. A replicable tool with low computational requirements is put forward as suitable for institutional contexts with limited resources.

Keywords: forest fires, predictive models, multilayer perceptron, geographic information systems

Article Info
Received July 13, 2025
Accepted October 13, 2025

1 Introduction

Forest fires have increased in frequency and severity in recent decades as a result of climate change, rising global temperatures, and growing anthropogenic pressure on natural ecosystems (Neger et al., 2022; González R. & Ortiz P., 2022). In regions such as eastern Michoacán, this situation is exacerbated by practices such as illegal logging, extensive cattle grazing, and seasonal agriculture, creating critical conditions of ecological and social vulnerability (España Boquera et al., 2024; Bassaber Zuñiga et al., 2024).

Institutional responses to these events are often limited by the lack of data-driven predictive tools that would allow for the anticipation of critical scenarios and the efficient allocation of firefighting resources. In this context, the use of Artificial Intelligence (AI) models combined with Geographic Information Systems (GIS) has proven to be an effective strategy for predicting environmental risks (Flores Garnica, 2021; González Gutiérrez et al., 2023). Various international studies have validated the application of algorithms such as Random Forest (RF), Support Vector Machines (SVM), and deep neural networks (CNNs, RNNs) to estimate the occurrence and extent of forest fires based on multivariate data (Alkhatib et al., 2023; Moghim & Mehrabi, 2024; Andrianarivony & Akhloufi, 2024; Pati et al., 2024).

In Latin America, approaches based on the use of multitemporal satellite imagery and biophysical variables through platforms like Google Earth Engine have been developed, enhancing real-time monitoring and predictive capabilities (Santelices Moya et al., 2022; Martínez Saucedo & Ezequiel Inchausti, 2023). In Mexico, studies by Ibarra Montoya & Huerta Martínez (2016), Román Cuesta & Martínez Vilalta (2006), and Vilchis-Francés et al. (2015) have demonstrated that integrating ecological, climatic, and land management factors can strengthen mitigation strategies.

Beyond natural factors, forest fires are also strongly influenced by social conditions such as rural poverty and the lack of sustainable economic alternatives, which increase the prevalence of agricultural burning (Huerta Silva et al., 2024; Cruz Núñez

& Bulnes Aquino, 2019). This highlights the need for comprehensive predictive models that not only identify high-susceptibility areas but also help optimize the allocation of operational resources in contexts with limited capacities.

The present study builds upon the work developed by Martínez Alcántar et al. (2025), in which a Multilayer Perceptron (MLP) neural network was used to estimate the area affected by forest fires in Michoacán, achieving an R^2 of 0.84. In this new stage, the analysis is expanded by comparing three supervised learning algorithms (MLP, Random Forest, and XGBoost), and applying techniques such as cross-validation, dimensionality reduction, and model ensembling. Although the optimized MLP model achieved an R^2 of 0.76, this result represents an improvement in terms of generalization and stability compared to the more controlled conditions and lack of formal validation in the previous study.

The central objective of this work is to design a predictive model capable of accurately estimating the potentially affected surface area by forest fires and simultaneously optimizing the use of operational resources such as deployed personnel, combat hours, and associated costs. The model results were integrated into geospatial analysis tools using QGIS and an interactive viewer developed with Leaflet JS, enabling practical use by civil protection and environmental management institutions.

This approach aims to provide a technically sound and operationally viable solution that is replicable, accessible, and data-driven—particularly valuable for environmentally fragile regions with limited technological infrastructure.

2 Experimental procedures

The methodological development of this study was structured into five phases: data integration and preprocessing, variable selection, design of predictive models using three machine learning algorithms, performance evaluation, and the construction of a prototype for the optimization of operational resources.

2.1 Data Collection and Preprocessing

Se utilizó un conjunto de datos históricos que comprende 930 registros de incendios forestales ocurridos entre 2015 y 2024 en 18 municipios de la región oriente de Michoacán. Los datos fueron recopilados por instancias de protección civil y brigadas comunitarias, e incluyen variables operativas como duración del incendio (horas), costo operativo (MXN), número de personas desplegadas, tipo de impacto, causa general, causa específica, tipo de incendio, vegetación y coordenadas geográficas.

Las variables categóricas fueron codificadas numéricamente mediante técnicas de Label Encoding y One-Hot Encoding. Posteriormente, se aplicó una normalización tipo Min-Max Scaling para homogenizar las escalas y mejorar la estabilidad durante el entrenamiento de los modelos.

2.2 Variable Selection and Analysis

An exploratory data analysis (EDA) was conducted to identify patterns, outliers, and correlations among variables. Using the Pearson correlation matrix, potential redundancies were identified, and variables with correlation coefficients exceeding ±0.85 relative to others in the dataset were excluded. Although Principal Component Analysis (PCA) was explored as a dimensionality reduction technique, it was ultimately decided to retain the original variables due to their direct interpretability and individual contribution to the model.

2.3 Training of Predictive Models

Three supervised machine learning algorithms were implemented and compared: Multilayer Perceptron (MLP), Random Forest (RF), and XGBoost. The dataset was split into 70% for training and 30% for testing. All models were validated using K-Fold cross-validation (k=5) to ensure their generalization capability.

a) MLP Model
A feedforward neural network was used, with its architecture optimized through empirical testing. Although deeper models (up to four hidden layers) were initially explored, it was found that a more compact architecture with two hidden layers provided better generalization. The final architecture was configured as follows:

- Input layer: 6 variables

- 1st hidden layer: 64 neurons, ReLU activation
- 2nd hidden layer: 32 neurons, ReLU activation
- Output layer: 1 neuron, linear activation
- Loss function: Mean Squared Error (MSE)
- Optimizer: Adam with a tuned learning rate
- Regularization: L2 ($\lambda = 0.0005$) and Dropout (20% on the first hidden layer)
- Preprocessing: StandardScaler normalization

The model can be mathematically represented as follows:

$$\hat{y} = W^{(3)} \cdot \mathrm{ReLU}(W^{(2)} \cdot \mathrm{ReLU}(W^{(1)} \cdot x + b^{(1)}) + b^{(2)}) + b^{(3)} \quad (1)$$

Where:

- $x \in \mathbb{R}^6$ represents the input vector with normalized variables.
- $\hat{y}$ represents the estimated affected area in hectares.

b) Random Forest

The RF model was implemented as an ensemble of decision trees trained on random subsets of the dataset. This technique allows for modeling nonlinear relationships without the need for data normalization and is robust to outliers and noise. The model was tuned using cross-validation combined with a grid search (GridSearchCV) to determine the most effective configuration for the problem at hand.

Final configuration:

- Number of trees (*n_estimators*): 200
- Maximum depth (*max_depth*): 10
- Minimum samples per split (*min_samples_split*): 2
- Minimum samples per leaf (*min_samples_leaf*): 1
- Feature selection (*max_features*): sqrt

The model can be generally expressed as:

$$\hat{y} = (1/T) * \sum h_t(x), \text{ for to } t = 1 \text{ to } T \quad (2)$$

Where:

- $\hat{y}$ is the predicted value (estimated affected area in hectares),
- T is the total number of trees in the ensemble,
- h_t is the prediction of the ttt-th decision tree for the input vector x

b) XGBoost

The XGBoost model was used as a tree-based boosting technique, trained to sequentially reduce the residual prediction errors. It optimizes a regularized loss function through gradient boosting, where each new tree improves the fit of the previous ensemble. The hyperparameters were tuned using grid search with cross-validation.

Final configuration:

- Number of trees (*n_estimators*): 200
- Maximum depth (*max_depth*): 6
- Learning rate (*learning_rate*): 0.1
- Subsampling ratio (*subsample*): 0.8
- Loss function: *reg:squarederror*

The general representation of the XGBoost model is as follows:

$$\hat{y} = \sum f_k(x), \text{ for to } k = 1 \text{ to } K \quad (3)$$

Where:

- $\hat{y}$ is the predicted value,
- K is the total number of boosting rounds (trees),
- $f_k(x)$ is the prediction of the k-th regression tree for input x

2.4 Resource Optimization Model

Once the MLP model was trained, a procedure was formulated to optimize the allocation of operational resources—namely, personnel, combat hours, and cost—with the goal of minimizing the estimated affected surface $\hat{y}$. Let $x\_op = \{Personnel, Hours, Cost\} \subset x$ represent the subset of operational input variables used in the optimization process.

The optimization problem is defined as follows:

$$\min_{x\_op} \hat{y}(x\_op)$$

Subject to:

$$50 \leq \text{Personnel} \leq 500$$
$$10 \leq \text{Hours} \leq 200$$
$$100{,}000 \leq \text{Cost} \leq 1{,}000{,}000$$

The objective function $\hat{y}(x\_op)$ corresponds to the surface area estimated by the trained MLP model given a specific configuration of operational resources. To solve the optimization problem, a grid search approach was applied, evaluating all discrete combinations of the input variables over a pre-defined range with fixed intervals. Continuous optimization techniques were not employed due to the non-differentiable nature of the neural network-based model.

2.5 Geospatial Visualization

Fire events were georeferenced using the latitude and longitude coordinates included in the dataset. Point vector layers were generated in QGIS to represent the spatial distribution of the events and their corresponding predictions. These layers were integrated into an interactive viewer developed with Leaflet JS, allowing users to query the location, the estimated affected surface area, and the optimal resources suggested for each event. This tool provides an accessible interface to support strategic decision-making in civil protection and environmental management (see Figures 1 and 2).

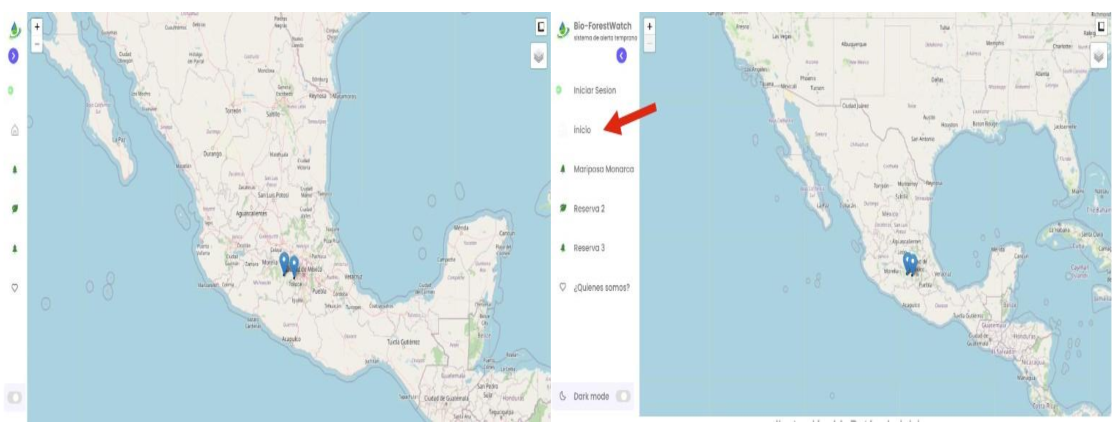

Fig. 1. Map showing the location of reserves in the eastern region of Michoacán

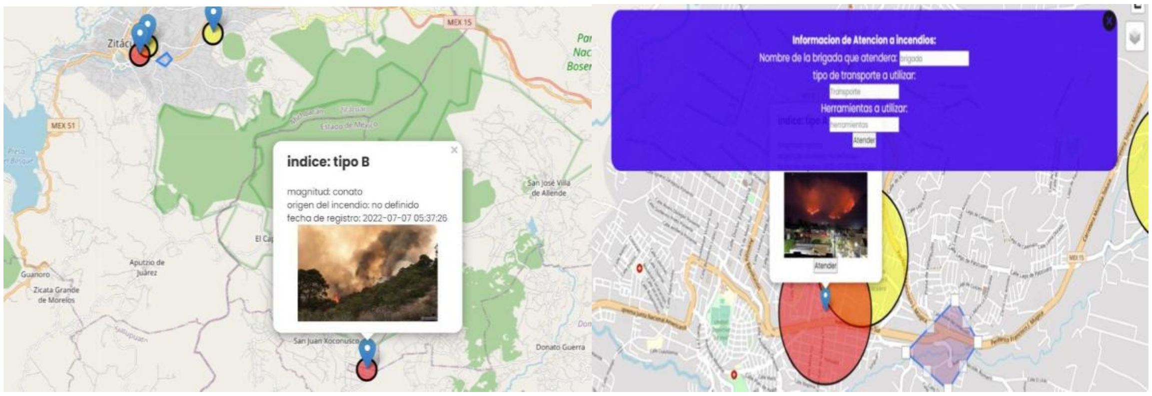

Fig 2. Fire Incident Data Capture

3 Results

The results were obtained after training, comparative evaluation, and validation of the predictive models, as well as the application of the resource optimization module. Standard regression metrics were used to assess model performance: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2).

3.1 Results by Model

The optimized MLP model achieved the best overall performance, with an R^2 of 0.76, outperforming the Random Forest (R^2 = 0.57) and XGBoost (R^2 = 0.51) models. The two-hidden-layer architecture, combined with regularization and cross-validation, improved generalization and prevented the overfitting observed in deeper architectures. Table 1 is presented below.

Table 1. Summary of Performance Metrics by Model

Model	MAE	MSE	RMSE	R^2
MLP	12.51	2154.07	46.41	0.76
Random Forest	17.82	3486.21	59.04	0.57
XGBoost	19.15	3781.44	61.49	0.51
Ensemble Model	13.74	2943.60	54.26	0.60

As observed, the MLP model clearly outperforms the others across all metrics, exhibiting lower absolute and squared errors, as well as greater explanatory power. The ensemble model was constructed using a weighted average of the individual predictions from the three models and achieved an improvement over RF and XGBoost, although it did not surpass the standalone performance of the MLP model.

3.2 Variable Importance Analysis

In the Random Forest model, the relative importance of the variables was evaluated based on impurity reduction. The three most impactful variables were:

1. Deployed personnel
2. Fire duration
3. Type of impact

This aligns with findings from previous studies, where human resources have been identified as a key factor in controlling the spread and severity of fires (Martínez Alcántar et al., 2025). The consistency in variable importance supported their inclusion as central components in both the MLP model and the optimization framework.

3.3 Optimization Model Results

Once the MLP model was trained, the optimization module was applied to identify resource combinations that minimized the expected affected surface area. A feasible domain was defined for personnel (50–500 individuals), combat hours (10–200 hours), and costs (100,000–1,000,000 MXN). The optimal results obtained were as follows:

- Estimated optimal cost: $1,000,000 MXN
- Optimal combat hours: 120 hours
- Optimal deployed personnel: 500 individuals
- Estimated affected area: 112.41 hectares

These values represent a maximum-response configuration within the defined operational domain, suggesting that the severity of the phenomenon in some scenarios requires aggressive intervention strategies. The results can be adjusted according to budgetary or resource availability constraints for specific scenarios.

3.4 Visual and Spatial Validation

The prediction and optimization results were integrated into a cartographic interface using QGIS and an interactive viewer developed with Leaflet JS. The visualization allows users to consult:

- Georeferenced events
- Estimated affected area per event
- Suggested optimal resources
- Thematic layers classified by severity

Figure 2 presents an example of the visualization of critical zones with high estimated impact, which can support the design of targeted intervention routes or differentiated contingency plans.

4 Conclusions

The results obtained in this study provide a foundation for reflecting on various aspects of the design and applicability of predictive models based on artificial intelligence for forest fire management. First, the optimized MLP model outperformed the Random Forest and XGBoost algorithms across all metrics, confirming the ability of neural networks to model nonlinear relationships in contexts involving structured and operational variables. Although the R^2 value of 0.76 was lower than the 0.84 reported by Martínez Alcántar et al. (2025) in a previous study using MLP, it is important to note that the earlier work did not apply cross-validation or formal regularization techniques, which may have resulted in overfitting to the training data. In contrast, the present proposal adopted a more rigorous approach that prioritized the model's ability to generalize to unseen data.
A key design aspect was the choice of a two-hidden-layer architecture for the MLP model, instead of the three layers used in the previous study. This decision was based on empirical testing through cross-validation, which revealed that a third hidden layer did not yield significant improvements and, in some cases, led to reduced generalization. Therefore, a more compact and efficient architecture was selected, maintaining high performance during validation while reducing error variance.

Another substantial advancement was the implementation of an operational resource optimization model, in which realistic constraints on personnel, combat time, and costs were defined. Unlike the previous study, where optimization was proposed only as a conceptual goal, this research developed a formal framework that enabled the identification of optimal intervention configurations to minimize the estimated affected area. This integration provides direct operational value for decision-makers and fire response teams.

Moreover, this analytical capability is strengthened through the incorporation of Geographic Information Systems (GIS). The georeferencing of fire events and their visualization in an interactive viewer using QGIS and Leaflet JS represent an essential component for transforming numerical predictions into actionable territorial knowledge. This functionality enables, for example, the rapid identification of high-susceptibility zones, the generation of estimated severity maps, and the targeted allocation of resources to priority areas. Previous experience documented in the literature highlights how such visual tools facilitate decision-making, even in institutional contexts with limited resources. This synergy between prediction, optimization, and visualization constitutes one of the main contributions of the present work. The developed system not only estimates the affected area, but also provides an integrated framework for territorial risk management in near-real time. Its modular design, low computational cost, and ease of adoption make it particularly suitable for institutions such as civil protection agencies, protected natural areas, and community fire management committees.

Nonetheless, certain limitations must be acknowledged. The model was based exclusively on structured and operational variables, without incorporating dynamic factors such as real-time meteorological data or multitemporal satellite imagery.

Future research should integrate these data sources through APIs or remote sensing technologies, and explore evolutionary optimization techniques, hybrid neural network architectures, or geospatial inference methods.

Overall, this study not only validates the utility of the MLP approach in predicting forest fire-affected surface area but significantly expands its applicability by incorporating operational criteria and spatial analysis. The result is a predictive, analytical, and visual tool that can be effectively implemented in environmentally vulnerable and operationally complex regions. A relevant contribution of this research is the inclusion of an optimization module, which made it possible to estimate optimal operational configurations—in terms of personnel, combat hours, and costs—to minimize the area affected by fire. This operational component transforms the predictive model into a strategic planning tool for both institutional and community settings.

In addition, the visualization of results through GIS platforms such as QGIS and Leaflet JS provided an additional layer of spatial interpretation, facilitating the identification of critical zones and territorial prioritization. This georeferenced representation capability extends the model's utility beyond technical analysis, making it accessible to decision-makers in resource-constrained contexts.

Finally, the model could benefit from future extensions by incorporating dynamic variables such as real-time weather conditions, satellite imagery, and socioeconomic data. Future work may also include the use of hybrid techniques and optimization based on evolutionary algorithms. In summary, this study presents a scalable, low-cost tool to support decision-making in the prevention, mitigation, and management of forest fires, with potential for implementation in various regions with high environmental vulnerability.

References

Alkhatib, R., Sahwan, W., Alkhatieb, A., & Schütt, B. (2023). A brief review of machine learning algorithms in forest fires science. *Applied Sciences, 13*(14), 8275. https://doi.org/10.3390/app13148275

Andrianarivony, H. S., & Akhloufi, M. A. (2024). Machine learning and deep learning for wildfire spread prediction: A review. *Fire, 7*(12), 482. https://doi.org/10.3390/fire7120482

Bassaber Zuñiga, A. I., Maldonado Villegas, M. M., Curiel Guevara, S., Rojas Arias, M. L., & Valdez Rosas, J. S. (2024). Revisión de causas, consecuencias y medidas de respuesta frente a los incendios forestales: Un enfoque en el estado de Jalisco. *Revista Científica e-CUCBA, 21*, 100–114. https://doi.org/10.32870/e-cucba.vi21.327

Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO). (n.d.). *Áreas protegidas*. Retrieved March 15, 2025, from https://www.biodiversidad.gob.mx/region/areasprot

CONAFOR. (2024). Reporte semanal de incendios. *Comisión Nacional Forestal*. https://www.gob.mx/conafor/documentos/reporte-semanal-de-incendios

Cruz Núñez, X., & Bulnes Aquino, E. (2019). Emission impact of wildfires: El Tepozteco 2016. *Atmósfera, 32*(2), 85–93. https://doi.org/10.20937/ATM.2019.32.02.01

España Boquera, M., Champo Jiménez, O., & Uribe Salas, M. (2024). Extensión y severidad de incendios forestales en Michoacán en 2021 a partir de imágenes Sentinel-2. *Polibotánica, 57*, 125–144. https://doi.org/10.18387/polibotanica.57.7

Flores Garnica, J. G. (2021). Antecedentes y perspectivas de la investigación en incendios forestales en el INIFAP. *Revista Mexicana de Ciencias Forestales, 12*(Especial), e981. https://doi.org/10.29298/rmcf.v12iEspecial-1.981

González Gutiérrez, I., Farfán Gutiérrez, M., Morales Manilla, L. M., Pérez Salicrup, D. R., Garza Saldaña, J. J., & Medina Puente, A. (2023). Índices y modelos para la predicción de la ocurrencia de incendios forestales: Una revisión para México. *Revista Geográfica de América Central, 71*(2), 193–219. https://doi.org/10.15359/rgac.71-2.7

González R., A., & Ortiz P., C. F. (2022). Superficie forestal afectada por incendios en México: Apuntes iniciales hacia un modelo de manejo preventivo. *Revista de Ciencias Ambientales, 56*(1), 1–27. https://doi.org/10.15359/rca.56-1.1

Huerta Silva, M. H., Parra Vázquez, M. R., & Ianni, E. (2024). Relación entre pobreza y conservación del capital natural desde los medios de vida sustentables en Villaflores, Chiapas. *Revista de Ciencias Ambientales, 58*(1), 19352. https://doi.org/10.15359/rca.58-1.9

Ibarra Montoya, J. L., & Huerta Martínez, F. M. (2016). Cambio climático y predicción de incendios al 2050 en el Bosque La Primavera, Jalisco. *Revista Mexicana de Ciencias Forestales, 7*(37), 39–50. https://doi.org/10.29298/rmcf.v7i37.50

Latorre-Cárdenas, M. C., González-Rodríguez, A., Godínez-Gómez, O., Arima, E. Y., Young, K. R., Denvir, A., … Ghilardi, A. (2023). Estimación de patrones de fragmentación y conectividad del bosque templado en un paisaje dominado por aguacates. *Tierra, 12*(3), 631. https://www.redalyc.org/journal/2818/281857987010/

Martínez Alcántar, J. L., Ruiz Garduño, J. K., Mora Patiño, G., & Viñas Álvarez, S. E. (2025). *Modelo predictivo para la gestión de incendios forestales en Michoacán con redes neuronales* [Unpublishing]. 2° Congreso Nacional de Ingeniería Industrial y Sistemas.

Martínez Saucedo, A., & Ezequiel Inchausti, P. (2023). Predicción de incendios forestales mediante modelos de machine learning. En *Actas del XXVIII Congreso Argentino de Ciencias de la Computación (CACIC)*. http://sedici.unlp.edu.ar/handle/10915/149568

Moghim, S., & Mehrabi, M. (2024). Wildfire assessment using machine learning algorithms in different regions. *Fire Ecology, 20*, 104. https://doi.org/10.1186/s42408-024-00335-2

Neger, C., León Cruz, J. F., Galicia Sarmiento, L., & Manzo Delgado, L. D. (2022). Dinámica espaciotemporal, causas y efectos de los megaincendios forestales en México. *Madera y Bosques, 28*(2), e2822453. https://doi.org/10.21829/myb.2022.2822453

Pati, A., Rout, R. R., & Tripathy, R. K. (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*, 170048. https://doi.org/10.1016/j.scitotenv.2024.170048

Ribeiro, L., Gama, J., & Gonçalves, R. (2018). Fireman: A GIS-based decision support system for wildland fire management. *International Journal of Geographical Information Science, 32*(11), 2191–2213. https://doi.org/10.1080/13658816.2018.1472396

Román Cuesta, R. M., & Martínez Vilalta, J. (2006). Effectiveness of protected areas in mitigating fire: Case study of Chiapas. *Conservation Biology, 20*(4), 1074–1086. https://doi.org/10.1111/j.1523-1739.2006.00478.x

Santelices Moya, R., Cabrera Ariza, A., Silva Flores, P., & Navarro Cerrillo, R. M. (2022). Assessment of a wildfire in *Nothofagus alessandrii* forests based on Sentinel-2. *International Journal of Agriculture and Natural Resources, 49*(2), 85–96. https://doi.org/10.7764/ijanr.v49i2.2337

Ye, J., Wang, N., Sun, M., Liu, Q., Ding, N., & Li, M. (2023). A new method for rapid detection of forest fire disturbance events using GEE and the VCT algorithm: A case study in China. *Remote Sensing, 15*(2), 413. https://doi.org/10.3390/rs15020413