

Sofiane Bououden
Salim Ziani
Mohammed Chadli
Ivan Zelinka *Editors*

Proceedings of the 6th International Conference on Electrical Engineering and Control Applications – Volume 2

ICEECA 2024, 19–21 November,
Khenchela, Algeria



Deep Learning Prediction of Wildfire Probability Occurrence Using Meteorological Factors from Khenchela Province Forests

Mohamed Lamri^{1,2(✉)}, Abdelhakim Sahour^{1,2}, Farouk Boumehrez^{1,3},
and Abdelaali Bekhouche⁴

¹ Department of Electrical Engineering, Abbes Laghrour University, 40004 Khenchela, Algeria
mohamed.lamri@univ-khenchela.dz

² Laboratory of Systems and Applications of Information and Telecommunications Technologies (SATIT), Department of Electrical Engineering, ABBES Laghrour University, BP 1252, Route de Batna Khenchela ELhama 40004, Khenchela, Algeria

³ Telecommunications Laboratory (LT), Telecommunications Institute, 8 May 1945 University, 24000 Guelma, Algeria

⁴ Department of Computer Sciences, ICOSI Laboratory, Abbes Laghrour University, 40004 Khenchela, Algeria

Abstract. Wildfire prediction is crucial for mitigating ecological and economic losses during forest management. This study presents a three-phased AI-driven methodology for enhanced wildfire risk assessment in Khenchela Province. First, explainable AI (XAI, LIME, and SHAP) analyzes eight machine-learning models (CatBoost, XGBoost, SVM, RF, AdaBoost, LR, LightGBM, and DT) to identify 12 key wildfire predictors from a historical dataset. Second, deep learning models (LSTM, RNN, and Prophet) forecast meteorological conditions to improve temporal resolution. Finally, a Bayesian model integrates these forecasts with moisture codes (FFMC, DMC, and DC) and indices (ISI, BUI, and NDVI) to predict wildfire likelihood. Validation using Accuracy, Precision, Recall, F1-score, and AUC demonstrated superior performance, with CatBoost achieving a 95.93% accuracy. This integrated approach enhances wildfire prediction reliability, enables proactive mitigation strategies, and shows the synergistic benefits of combining machine learning, deep learning, and Bayesian modeling for improved wildfire risk management.

Keywords: Deep Learning · Machine Learning · Bayesian Modeling · Wildfire Prediction · XAI · Meteorological Factors · Moisture Codes · Indices · Khenchela Province · Risk Assessment

1 Introduction

The 2008 and 2012 wildfires in Algeria's Khenchela Province were the most intense over more than a decade. The Conservation of Forests of the Province of Khenchela (CFPK) reported that Forest Fires have increased significantly in the last few years,

causing human, ecological, and economic losses in the regions affected by these fires. Scoring more than 15719 hectares of forested and green land had burned, 73.55% of the total tree cover lost over the two decades [1].

With so many fires happening around the world on a near-annual basis, predictive modeling of forest fires is essential for a proactive approach in high-risk areas. The official reports prepared after the wildfires indicate that environmental conditions, such as extreme climatic conditions characterized by severe drought and unfavorable winds, coupled with intentional or unintentional human factors, caused devastating fires in Khenchela Province, which negatively impact the environment and play an important role in the carbon cycle and in mitigating the effects of climate change, including high average temperatures, increased winter precipitation, and drier summers (CFPK in 2018) [2].

However, to improve sustainable forest management and fire prevention, it is imperative to develop a fire prediction system using algorithms that use meteorological factors and various exploratory features to model the probability of wildfire. Various modeling techniques, such as machine/deep learning and frequent/ Bayesian regression models, have been used in recent literature (Jain et al., 2020) [3]. In particular, deep learning (DL) has been proposed as a methodology suitable for modeling complex variable interactions that are ubiquitous in Earth system problems, particularly in wildfires, by Reichstein et al. (2019) [4] and Camps-Valls et al. (2021) [5]. Indeed, a limited but growing body of work has used DL for short-term fire hazard prediction. Le et al. (2021) [6] use a multilayer perception to predict wildfire hazard, while Zhang et al. (2019, 2021) [7], Bergado et al. (2021) [8], and Bjånes et al. (2021) [9] use convolutional neuronal nets (Krizhevsky et al., 2012) [10] to model wildfire vulnerability.

This study presents a methodology for predicting the likelihood of forest fires based on artificial intelligence, in three phases:

Firstly, employed two explainable artificial intelligence (XAI) tools, namely Local Interpretable Model-Agnostic Explanations (LIME) [11] and Shapley Values (SHAP) [12], to quantify variables' attribution, interpret models' predictions, and compare the results yielded by eight machine-learning models. The objective was to identify, examine, and select the most significant factors contributing to forest wildfires in the Khenchela Province from a large set of features in the dataset. This was achieved by employing CatBoost, XGBoost, SVM, Random Forest, AdaBoost, Logistic Regression, LightGBM, and Decision Trees.

Second, we leveraged diverse deep learning (DL) techniques, including long short-term memory (LSTM), recurrent neural network (RNN), and Prophet, to analyze historical data and predict meteorological factors for the forthcoming day in areas of Khenchela that are prone to fire.

Finally, a Bayesian model was developed to predict the probability of wildfire occurrence based on a set of meteorological factors, including forecasting and moisture codes such as FPMC, DMC, DC, and indices, namely the Initial Spread Index (ISI), build-up index (BUI), and normalized difference vegetation index (NDVI). These factors were identified as potential influences on the spread of fires.

The operation of machine learning algorithms depends on a number of factors, including weather conditions, a set of environmental predictors, and indices. Consequently,

the utilization of prior information and its prediction on the subsequent day via Bayesian multiple logistic regression modelling facilitates the generation of precise and dependable projections concerning the probability of wildfire occurrence. The modeling results indicate that the three proposed phases are able to develop forest fire prediction based on available climate data, selected moisture codes, and easily measurable indices. Furthermore, the qualitative use of Deep Learning lends confidence to taking preventive actions and hints at the mechanisms that drive it.

2 Materials and Methods

2.1 Data Preprocessing

The data used in this study provide detailed information on the occurrence of wildfires in three massifs in this region: Ouled-Yagoub, Beni Oudjana, and Beni Imloul. Data were collected by the Conservation of Forests of the Province of Khenchela (CFPK) in the form of monthly reports over a period of 33 years (1985–2017). The dataset contains 396 records and 35 features, and includes various information such as human factors, relief, combustibility index, forest stand, vegetation moisture, slope, aspect, elevation, distance from settlement, distance from agricultural area, distance from road, number of wildfire hot spots, date, spatial location, type of vegetation affected, the six components of the FWI system, the total area burned, and several meteorological observations (e.g., wind speed, temperature, precipitation), consisting of numerical values and a predicted fire/non-fire class.

3 Results and Discussion

This section presents the experimental results, detailing the identification of key wildfire predictors using explainable AI, deep-learning forecasting, and a Bayesian model for wildfire probability prediction. Model performance is evaluated using multiple metrics: accuracy, precision, recall, F1-score, AUC-ROC, mean absolute error (MAE), root mean square error (RMSE), and mean squared error (MSE).

3.1 Feature Selection Using SHAP and LIME with Machine Learning.

Feature selection across various machine learning (ML) models using Explainable AI (xAI) approaches such as SHAP and LIME is a good starting point for structured analysis. The following methodological steps are recommended by this technique, which considers significant aspects of the feature selection process:

Data Preparation: After removing columns with more than 5% missing values from the data, the 396 data entries and all 35 columns were suitably addressed using the median imputation technique to replace the missing values, and were converted to float and then normalized across all attributes.

Feature Selection Using SHAP: To train, avoid leakage, and evaluate the model performance of each of the eight ML models (XGBoost, Random Forest, AdaBoost,

CatBoost, LightGBM, Decision Tree, Support Vector Machine, and Logistic Regression), the dataset was divided into training and test sets (commonly an 80:20 split). Using Libraries ‘shap’ in Python provide straightforward ways to computed the SHAP values, and visualized it. Subsequently, These SHAP values were aggregated across all models to consistently identify the most influential features across different models. A common approach is to determine the mean absolute SHAP value for each feature across all the models.

Feature Selection Using LIME: Using ‘lime’ library allows to generate local explanations for individual predictions. This can help understand how specific features influence particular predictions. To ensure feature relevance, the results were aggregated to identify which features frequently appear in the top influential features during local explanations. This can provide a complementary perspective to the global insights gathered from the SHAP.

Comparison and Analysis: Feature selection was performed using SHAP and LIME, ranking features based on the top 5% of their importance scores. Table 1.a-1. d summarizes the top 12 features identified across the eight machine-learning models. ‘temp_max’ consistently ranked highest, indicating its crucial role in wildfire prediction. While some features, such as ‘evapotranspiration’, ‘precipitation’, and ‘dmc’, showed less consistent global importance, they exhibited substantial influence on specific subsets of the data, suggesting context-dependent effects. Nine features (‘temp_max’, ‘evapotranspiration’, ‘precipitation’, ‘dc’, ‘dmc’, ‘wind_max’, ‘fwi’, ‘humidity_avg’, ‘bui’) consistently ranked within the top five across all models, despite variations in their overall importance scores. This highlights the complex interplay of the factors influencing wildfire occurrence.

Interestingly, ‘ffmc’, ‘ndvi’, and ‘isi’ were identified as significant predictors only by six of the eight models. This suggests that some features exhibit context-dependent relevance, whereas others have broader applicability for predicting wildfire occurrence. This variability underscores the importance of careful model selection and comprehensive feature analysis in predictive modeling. The results highlight the complex interplay between features and outcomes, emphasizing the intricacies of wildfire prediction and the value of XAI tools such as SHAP and LIME in providing insights into these relationships. For instance, ‘temp_max’ strongly influences evaporation and transpiration rates, whereas ‘evapotranspiration’ reflects the complex interplay between temperature and hydrological processes.

This analysis revealed several key relationships. ‘Precipitation’ directly impacts soil moisture, influencing evaporation and transpiration rates. ‘Humidity_avg’ acts as a mediator of evaporation; lower humidity increases evapotranspiration, while higher humidity reduces it. ‘Wind_max’ accelerates evaporation by removing moisture from surfaces, affecting both evapotranspiration rates and the vegetation microclimate. Finally, the inclusion of ‘dc’, ‘bui’, ‘ndvi’, and ‘isi’ in some models demonstrates the model’s capacity to capture the complex interplay of vegetative health, fire risk, and moisture levels, highlighting its sensitivity to diverse environmental conditions.

Model performance was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC, comparing the results with 35 features to a reduced set of 12 key features.

CatBoost achieved the highest accuracy with 35 features (89.84%) but showed a significant improvement to 95.93% accuracy with only 12 features, representing a 6.09% overall performance gain.

The XGBoost accuracy increased from 88.53% (35 features) to 95.11% (12 features), which is an improvement of 6.58%. LightGBM showed a similar trend, improving from 87.21% to 94.26%, a gain of 7.05%. The random Forest performance improved from 88.53% to 88.80% to approximately 95.14%, representing a gain of approximately 6.61%.

The logistic Regression accuracy improved from 88.53% to 89.43%, to 93.82%–94%, and a gain of 4.53%–5.29%. AdaBoost showed a substantial improvement, increasing from approximately 86% to 92.94%–95.67%, a gain of 6.94%–9.67%. SVM accuracy increased from 88.53% to 90.65%–93.31%, a gain of 2.12%–4.78%. Finally, the Decision Tree accuracy showed the most dramatic improvement, ranging from 59.25%–78% with 35 features to 79.01%–88.41% with 12 features and a gain of up to 19.01%. This highlights the significant benefit of feature reduction in the decision-tree model.

The ROC-AUC scores further support the positive impact of feature reduction. Logistic Regression, Random Forest, and CatBoost achieved the highest ROC-AUC scores with 12 features (99.99%, 99.68%, and 99.56%, respectively), showing improvements of 2.61%, 2.1%, and 4.13%, respectively, over their performance with 35 features. AdaBoost also showed an improvement (99.02% vs. 96.63%, a 2.39% gain), similar to LightGBM (98.64% vs. 94.5%), XGBoost (97.45% vs. 92.2%), and Decision Trees (90.15% vs. 80.19%). Even SVM, which initially had a lower performance, showed a substantial increase (from 72.7% to 44.5%). These results consistently demonstrate the effectiveness of feature reduction in enhancing the model performance, particularly in terms of both accuracy and ROC-AUC. Furthermore, this work can be merged with many techniques (see [13–18]).

4 Conclusions

This study demonstrates the feasibility of predicting wildfire occurrences in Khenchela, Algeria using meteorological data and established fire indices. The vulnerability of the region, particularly during May–September, necessitates accurate predictions for effective mitigation. Our analysis highlighted the importance of meteorological variables (temperature, humidity, wind speed, and rainfall) and the Canadian Forest Fire Weather Index (FWI) system (FFMC, DMC, DC, ISI, BUI, FWI) in assessing fire risk.

To identify the most influential predictors, we employed SHAP and LIME analyses across eight ML models (CatBoost, XGBoost, SVM, RF, AdaBoost, LR, LightGBM, and DT). Table 1 summarizes the top features selected by each method and the corresponding model performance metrics (accuracy, precision, Recall, F1-score, AUC). The results show consistent identification of specific meteorological and FWI components as key predictors across models, with CatBoost showing the highest accuracy for a specific feature subset. This analysis highlights the importance of integrating these key predictors into predictive models to improve wildfire risk assessments.

Table 1. a: Top features derived from both SHAP and LIME analyses for each model and theirs Performance Metrics.

Features	Models	Feature Selection Method			Cited by	Performances						
		Values	Models	Class in model		Accuracy	Precision	Recall	F1 Score	Roc_Auc		
temp_max	XGBoost	2.488394	XGBoost	1	All models with both methods (1 st relevant)	0.898421	0.898421	0.898421	0.838421	0.954284		
	CatBoost	1.1113406	CatBoost	1		0.885263	0.882969	0.885263	0.83937	0.921967		
	LR	0.229901	LGBM	1		0.872105	0.872105	0.872105	0.872105	0.945005		
	LGBM	0.181200	AdaBoost	1		0.885263	0.888036	0.885263	0.864778	0.973826		
	AdaBoost	0.169545	RF	1		0.885263	0.894899	0.885263	0.888603	0.973675		
	DT	0.161491	DT	1		0.858947	0.862143	0.858947	0.860333	0.966286		
	RF	0.144567	LR	3		0.882865	0.882865	0.873109	0.873109	0.901930		
	SVM	0.032119	SVM	12		0.78	0.5925	0.78	0.67887	0.801930		
	evapotranspiration	XGBoost	0.914146	DT		2	6 models with SHap + 6 models using LIME (6 th relevant)	0.898421	0.898421	0.898421	0.838421	0.954284
CatBoost		0.662003	RF	2	0.885263	0.882969		0.885263	0.83937	0.921967		
DT		0.153232	XGBoost	3	0.872105	0.872105		0.872105	0.872105	0.945005		
LGBM		0.052294	CatBoost	3	0.88036	0.888036		0.885263	0.864778	0.973826		
RF		0.051875	LGBM	3	0.885263	0.894899		0.885263	0.888603	0.973675		
SVM		0.036016	SVM	5	0.858947	0.862143		0.858947	0.860333	0.966286		
/		/	LR	11	0.882865	0.882865		0.873109	0.873109	0.901930		
/		/	AdaBoost	12	0.78	0.5925		0.78	0.67887	0.801930		
precipitation		XGBoost	0.763173	XGBoost	2	3 models using SHap + 6 models using LIME (2 nd relevant)		0.956331	0.956331	0.956331	0.956331	0.909594
	CatBoost	0.063120	LGBM	2	0.951081		0.949984	0.951081	0.951081	0.974552		
	AdaBoost	0.057493	AdaBoost	2	0.942802		0.942802	0.942802	0.942802	0.984374		
	LR	0.042734	CatBoost	5	0.951431		0.952818	0.951431	0.952818	0.996774		
	DT	0.039169	RF	9	0.938181		0.93449	0.938181	0.940001	0.996774		
	LGBM	0.034915	DT	11	0.938181		0.93449	0.938181	0.940001	0.996774		
	SVM	0.033763	/	/	0.938181		0.93449	0.938181	0.940001	0.996774		
	XGBoost	0.730655	XGBoost	5	6 models with SHap + 6 models using LIME (4 th relevant)		0.956331	0.956331	0.956331	0.956331	0.909594	
	CatBoost	0.383928	LR	5			0.951081	0.949984	0.951081	0.951081	0.974552	
	LR	0.129503	LGBM	6			0.942802	0.942802	0.942802	0.942802	0.984374	
AdaBoost	0.116068	AdaBoost	6	0.951431		0.952818	0.951431	0.952818	0.996774			
dmc	DT	0.052025	CatBoost	7	0.938181	0.93449	0.938181	0.940001	0.996774			
	LGBM	0.044889	RF	7	0.938181	0.93449	0.938181	0.940001	0.996774			
	/	/	SVM	7	0.938181	0.93449	0.938181	0.940001	0.996774			
	/	/	DT	8	0.938181	0.93449	0.938181	0.940001	0.996774			

(continued)

Table 1. (continued)

Table 1.d: Feature Selection Experiment Results.									
Features	Feature Selection Method				Cited by	Feature Selection Method			
	Models	Values	Models	Values		Models	Values	Models	Class in model
dc	XGBoost	0.077814	RF	0.077814	Shap + 7 models using LIME (3 rd relevant)	XGBoost	0.272056	RF	5
	DT	0.054403	LR	0.054403		DT	0.149400	DT	5
	RF	0.041229	Ada	0.041229		Ada	0.051463	Ada	7
	SVM	0.056688	Boost	0.056688		XGBoost	0.047102	XGBoost	6
	Ada	0.032167	XGBoost	0.032167		LR	0.043153	Boost	9
fwi	Boost	0.031030	LGBM	0.031030	Shap + 8 models using LIME (7 th relevant)	LGBM	0.043153	LR	10
	LGBM	0.031030	DT	0.031030		DT	/	LGBM	10
	/	/	SVM	/		SVM	/	LR	11
	XGBoost	0.501010	SVM	0.501010		Ada	0.175261	SVM	11
	LR	0.195866	LR	0.195866		Boost	0.077969	Ada	5
ndvi	Ada	0.08912	XGBoost	0.08912	Shap + 7 models using LIME (3 rd relevant)	LR	0.064380	SVM	6
	DT	0.090178	Ada	0.090178		XGBoost	0.040483	LGBM	8
	Boost	0.051413	Boost	0.051413		LR	0.030865	LR	8
	/	/	LGBM	/		DT	/	Ada	11
	/	/	/	/		RF	/	Boost	12
humidity_avg	XGBoost	0.485627	RF	0.485627	Shap + 7 models using LIME (12 th relevant)	XGBoost	0.063661	Ada	13
	LR	0.177389	SVM	0.177389		DT	0.053054	LR	9
	DT	0.032599	XGBoost	0.032599		LR	0.049986	RF	8
	/	/	Ada	0.99		Ada	/	Ada	14
	/	/	LGBM	/		Boost	/	Boost	4
humidity_avg	/	/	DT	/	Shap + 8 models using LIME (9 th relevant)	DT	/	DT	4
	XGBoost	0.401528	Ada	0.401528		XGBoost	0.232136	Ada	1
	DT	0.073416	SVM	0.073416		LR	0.135462	Boost	3
	SVM	0.046520	RF	0.046520		SVM	0.125009	SVM	4
	/	/	DT	/		LGBM	0.091250	LGBM	6
humidity_avg	/	/	LGBM	72.80	Shap + 4 models using LIME (9 th relevant)	Ada	0.068614	LR	8
	/	/	XGBoost	/		XGBoost	/	XGBoost	8
	/	/	Ada	/		DT	/	Ada	10
	/	/	Boost	/		Boost	/	Boost	10
	/	/	LR	/		DT	/	DT	11

References

1. Algeria Fire 2023; DREF Operation No. MDRDZ009, Relief web. <https://reliefweb.int/report/algeria-fire-2023-dref-operation-no-mdrdz0009>
2. Office national des travaux forestiers inventaire forestier de la forêt de Beni Imloul dans le massif des Aurès – Algérie. Rapport de la Direction du grand projet Aurès Batna. Projet de mise en valeur agro-sylvo-industriel des Aurès, G.T.Z n° 114, ONTF. (1977)
3. Jain, P., Coogan, S.C., Subramanian, S.G., Crowley, M., Taylor, S., Flannigan, M.D.: Review of machine learning applications in wildfire science and management. *Environ. Rev.* **28**(4), 478–505 (2020)
4. Reichstein, M., et al.: Deep learning and process understanding for data-driven Earth system science. *Nature* **566**(7743), 195–204 (2019)
5. Camps-Valls, G., Tuia, D., Zhu, X.X., Reichstein, M. (eds.) *Deep Learning for Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science, and Geosciences*. John Wiley & Sons (2021)
6. Van Le, H., et al.: A new approach to deep neural computing for spatial prediction of wildfire danger in tropical climate areas. *Eco. Inform.* **63**, 101300 (2021)
7. Zhang, G., Wang, M., Liu, K.: Deep neural networks for global wildfire susceptibility modelling. *Ecol. Ind.* **127**, 107735 (2021)
8. Bergado, J.R., Persello, C., Reinke, K., Stein, A.: Prediction of wildfire burns from large geodata using deep learning. *Saf. Sci.* **140**, 105276 (2021)
9. Bjânes, A., De La Fuente, R., Mena, P.: Deep learning ensemble model for wildfire-susceptibility mapping. *Eco. Inform.* **65**, 101397 (2021)
10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Image classification using deep convolutional neural networks. In: *Advances in neural information processing systems*: 25 (2012)
11. Liu, Y., Liu, Z., Luo, X., Zhao, H.: Diagnosis of Parkinson's based on SHAP-value feature selection. *Biocybernet. Biomed. Eng.* **42**(3), 856–869 (2022)
12. Marçílio, W.E., Eler, D.M.: From explanations of feature selection, we assessed SHAP values as a feature selection mechanism. In: 2020, the 33rd SIBGRAPI Conference on Graphics, Patterns, and Images (SIBGRAPI), pp. 340–347. IEEE (2020)
13. Bounemour, A., Chemachema, M., Essounbouli, N.: Indirect adaptive fuzzy fault-tolerant tracking control for MIMO nonlinear systems with actuator and sensor failures. *ISA Trans.* **1**(79), 45–61 (2018)
14. Bounemour, A., Chemachema, M.: Adaptive fuzzy fault-tolerant control using Nussbaum-type function with state-dependent actuator failures. *Neural Comput. Appl.* **33**, 191–208 (2021)
15. Bounemour, A., Chemachema, M.: Adaptive fuzzy fault-tolerant control for a class of nonlinear systems under actuator faults: application to an inverted pendulum. *Int. J. Robot. Control Syst.* **1**(2), 102–115 (2021)
16. Ziani, S., Djebabla, A., Filali, S.: The mixed H_2/H_∞ initial controller for Enhancing the Youla Parameterization via LMI Optimization. In: 2008 2nd International Conference on Signals, Circuits and Systems, pp. 1–6. IEEE (2008)
17. Ziani, S., Filali, S.: An initial corrector using different approaches for Youla parametrization via LMI optimization. In: *InAIP Conference Proceedings 2008 Jun 12*, vol. 1019, no. 1, pp. 61–66. American Institute of Physics (2008)
18. Ziani, S., Filali, S.: Improving the Envelope of the LMI Optimization using the Youla Parametrization. *Stud. Inform. Univ.* **9**(2), 152–172 (2011)